INTELETICA 3(6)(2026), 15-26
Revista de Inteligencia Artificial, Ética y Sociedad

INTELETICA

https://inteletica.iberamia.org/

Inteligencia artificial conversacional y psicología clínica como complemento o sustitución profesional
Revisión narrativa crítica de evidencia internacional e implicaciones para Ecuador

Conversational artificial intelligence and clinical psychology as support or professional replacement
A critical narrative review of international evidence and implications for Ecuador

Wilmer Geovanny Martínez Rincón, Jamilex Jimabel Lino Castillo

Abstract Conversational artificial intelligence (AI) is increasingly used for emotional support, psychoeducation, and self-guided mental health interventions. This critical narrative review examines clinical benefits, limitations, safety concerns, and professional implications, and discusses their relevance to Ecuador. The synthesis draws on a previously assembled bibliography and a purposive update supported by two documented PubMed searches. Executed on 29 September 2026 (UTC), the searches covered publication dates from 1 January 2021 to 28 September 2026 and retrieved 416 unique records. These records were not all screened. Relevant full-text passages were checked for the 28 scientific publications discussed, alongside five official documents. There was no independent duplicate human review, new meta-analysis, or formal review-level risk-of-bias assessment. Purpose-built agents showed short-term symptom benefits, but heterogeneous effects and low certainty in recent syntheses limit generalization. Benefits from interventions combining chatbots with human support cannot be attributed to the chatbot alone. Perceived working alliance does not establish equivalence with human psychotherapy. Studies of general-purpose models describe limitations in clinical inquiry, crisis responses, and handling of complex presentations. The evidence discussed does not establish safe replacement of clinical psychologists. Supervised, bounded uses warrant evaluation. The proposed implications for Ecuador are interpretive and require local clinical and cultural validation, data protection, human referral pathways, and prospective safety monitoring. The non-systematic selection limits the review's comprehensiveness.

Resumen La inteligencia artificial (IA) conversacional se utiliza para apoyo emocional, psicoeducación e intervenciones autoguiadas. Esta revisión narrativa crítica examina beneficios clínicos, limitaciones, riesgos e implicaciones profesionales, y analiza su pertinencia para Ecuador. La síntesis parte de una bibliografía previamente reunida y una actualización intencional apoyada en dos búsquedas documentadas en PubMed. Ejecutadas el 29 de septiembre de 2026 (UTC), abarcaron publicaciones del 1 de enero de 2021 al 28 de septiembre de 2026 y recuperaron 416 registros únicos, cuyo cribado no se completó. Se cotejaron pasajes de los textos completos de las 28 publicaciones científicas discutidas y se consultaron cinco documentos oficiales. No hubo revisión humana duplicada independiente, nuevo metaanálisis ni evaluación formal propia del riesgo de sesgo. Los agentes especializados mostraron beneficios sintomáticos a corto plazo, aunque la heterogeneidad y la baja certeza descrita en síntesis recientes limitan su generalización. Los efectos de intervenciones que combinan chatbots con apoyo humano no pueden atribuirse únicamente al sistema. La alianza percibida tampoco establece equivalencia con psicoterapia. Los estudios de modelos generales describen dificultades de indagación, respuesta ante crisis y manejo de situaciones complejas. La evidencia discutida no establece una sustitución segura del psicólogo clínico. Los usos complementarios requieren evaluación específica. Las implicaciones para Ecuador son interpretativas y necesitan validación clínica y cultural, protección de datos, derivación humana y seguimiento de seguridad. La selección no sistemática limita la exhaustividad de esta revisión.

ISSN: 3020-7444
© Los Autores. Open Access, bajo Licencia Creative Commons (CC BY-NC).

Palabras clave: inteligencia artificial conversacional, chatbots de salud mental, psicología clínica, salud mental digital, ética de la inteligencia artificial, Ecuador

Keywords: conversational artificial intelligence, mental health chatbots, clinical psychology, digital mental health, artificial intelligence ethics, Ecuador

1 Introducción

Los agentes conversacionales pueden basarse en reglas, modelos de lenguaje generativos o combinaciones de ambos. Los primeros utilizan respuestas y recorridos delimitados. Los modelos generativos producen texto adaptado al contexto, con mayor variabilidad de salida. La revisión tecnológica de Garzón-Quiroz et al. [9] describe diferencias entre sistemas basados en reglas y procesamiento de lenguaje natural en distintos sectores. Aporta contexto técnico, pero sus resultados no demuestran eficacia terapéutica. En salud mental, la arquitectura, la finalidad y las condiciones de uso de cada producto deben describirse antes de interpretar sus resultados.

Algunos agentes diseñados para salud mental ofrecen psicoeducación y ejercicios estructurados. Las síntesis disponibles informan reducciones de síntomas depresivos y, en algunas comparaciones, ansiedad o malestar, principalmente a corto plazo [11, 18]. Las revisiones recientes presentan efectos heterogéneos y limitaciones de certeza [10, 27]. Los resultados de chatbots generativos o híbridos tampoco permiten asumir una eficacia común a todos los productos [8, 33]. Además, las revisiones pueden compartir ensayos primarios, por lo que sus conclusiones no constituyen necesariamente confirmaciones independientes.

Conviene distinguir una aplicación diseñada y evaluada para una tarea clínica de un modelo general utilizado como apoyo emocional. La fluidez del diálogo y la alianza percibida no acreditan por sí solas competencia terapéutica [3, 7, 29]. Las evaluaciones de respuestas simuladas identifican dificultades de indagación, estigma y manejo de crisis en versiones concretas de modelos generales [20, 22, 26]. La revisión de Bucher et al. [5] señala una evaluación clínica todavía limitada de los modelos de lenguaje y la necesidad de estudiar su integración en contextos de atención. Las propuestas de desarrollo responsable recomiendan participación clínica, evaluación y vigilancia [28].

Esta diversidad de evidencia exige separar los efectos de una intervención, las valoraciones de sus usuarios, las pruebas de seguridad y las percepciones profesionales. Los estudios sobre actitudes o experiencia de psicoterapeutas [17, 30] no prueban eficacia ni representan automáticamente otros países. Del mismo modo, una puntuación favorable de alianza no establece equivalencia con psicoterapia, y la ausencia de eventos graves en un ensayo breve no garantiza seguridad general.

En Ecuador, el debate combina la necesidad de ampliar el acceso con la protección de derechos. La bibliografía examinada no permite atribuir una opinión homogénea a los psicólogos clínicos ecuatorianos, y una revisión documental no sustituye una consulta nacional. El objetivo es analizar críticamente evidencia internacional publicada desde 2021, con una actualización bibliográfica hasta el 28 de septiembre de 2026, sobre beneficios, límites y riesgos de la IA conversacional, y discutir sus implicaciones para Ecuador. La pregunta es qué sustento existe para usos complementarios o para la sustitución de funciones clínicas y qué condiciones deben evaluarse antes de trasladar resultados al contexto ecuatoriano.

2 Método

2.1 Diseño y marco de la pregunta

Se elaboró una revisión narrativa crítica con búsqueda bibliográfica documentada y selección intencional de fuentes. La bibliografía de partida combinaba ensayos, estudios observacionales, comparaciones de respuestas, síntesis y aportes conceptuales. No se dispuso de un historial verificable de su selección original. Por ello, no se atribuyen a ese proceso recuentos de cribado, un flujo de selección ni cumplimiento de PRISMA-ScR. El diseño de esta versión es narrativo.

El análisis se organizó en tres dimensiones: personas usuarias y profesionales de psicoterapia, agentes conversacionales para apoyo o intervención en salud mental, e implicaciones interpretativas para Ecuador. La selección intencional permite contrastar diseños y argumentos, pero no estimar la distribución de toda la evidencia disponible ni demostrar ausencia de estudios sobre un tema.

2.2 Estrategia de búsqueda

Se ejecutaron dos consultas nuevas en PubMed mediante NCBI E-utilities el 29 de septiembre de 2026, a las 04:11 UTC, equivalentes al 28 de septiembre a las 23:11 en Ecuador continental. El intervalo de publicación fue del 1 de enero de 2021 al 28 de septiembre de 2026. Q1 recuperó 195 registros y Q2 recuperó 238. Se identificaron 17 coincidencias por PMID y se conservaron 416 registros únicos. Estas cifras corresponden a recuperación bibliográfica, no a publicaciones cribadas o incluidas. El Anexo A reproduce las ecuaciones literales. Se conservaron las respuestas de la API, las traducciones de las consultas y los registros bibliográficos.

Como fuente adicional se auditó la bibliografía ya citada en el manuscrito. De sus 27 publicaciones científicas, 20 aparecieron en las dos consultas y siete no fueron capturadas. Para estas últimas se efectuaron búsquedas por DOI que localizaron cuatro registros en PubMed. Esta comprobación bibliográfica no constituye una búsqueda sistemática adicional en portales editoriales. Los cinco documentos oficiales de la bibliografía se utilizaron exclusivamente para contextualización ética y normativa.

2.3 Selección dirigida de fuentes

Se dio prioridad interpretativa a síntesis con métodos explícitos, ensayos con resultados clínicos, investigaciones sobre alianza y comunicación, evaluaciones empíricas de seguridad y estudios de percepciones profesionales. Las fuentes narrativas o conceptuales se conservaron como contexto y no se trataron como demostraciones experimentales. Los protocolos sin resultados y las aplicaciones exclusivamente diagnósticas no se utilizaron para sostener eficacia terapéutica.

Se conservaron 25 publicaciones de la bibliografía científica inicial y se añadieron dos metaanálisis de 2026 y el ensayo STARS [4, 10, 27]. Esta actualización se eligió por su aporte a la incertidumbre de los efectos y a la distinción entre intervención autónoma y guiada. Se retiraron dos referencias prescindibles por limitaciones de acceso o de comprobación de los datos. El conjunto discutido comprende 28 publicaciones científicas y cinco documentos oficiales. Se prepararon otras fichas exploratorias que no se trataron como fuentes incluidas ni como decisiones de elegibilidad.

La búsqueda, la organización de registros, la extracción dirigida, el cotejo documental y la edición se realizaron con asistencia de ChatGPT. No se realizó una revisión humana duplicada independiente ni un procedimiento de consenso entre revisores. Las señales automáticas por título no se contabilizaron como exclusiones. No se calculó concordancia interevaluador ni se reconstruyeron retrospectivamente decisiones del proceso original. La asistencia de IA no equivale a un segundo revisor ni sustituye la responsabilidad científica de la autoría.

2.4 Extracción y síntesis

La extracción registró referencia, diseño, población, intervención, comparador, resultados y limitaciones. En las 28 publicaciones científicas se cotejaron los pasajes pertinentes de los textos completos, con atención a tamaños muestrales, estimaciones, seguimiento y condiciones de evaluación. La comprobación fue de las afirmaciones utilizadas y no una evaluación integral del riesgo de sesgo. Para el contexto oficial se consultaron el texto de la ley de protección de datos, las dos guías de la OMS y los registros oficiales de publicación de la ley y la política de salud mental. La matriz de auditoría conserva la fuente y la corrección o delimitación correspondiente.

La unidad de exposición es la publicación. Se separaron hallazgos clínicos, evaluaciones de seguridad, percepciones profesionales y contexto conceptual o normativo. No se realizó un nuevo metaanálisis ni una evaluación formal propia del riesgo de sesgo o de la certeza. Estas valoraciones se atribuyen a las fuentes que las realizaron. No se sumaron participantes ni efectos de revisiones que pueden compartir ensayos. Las propuestas para Ecuador se presentan en la discusión como interpretaciones. La planificación de la búsqueda nueva no acredita un protocolo prerregistrado del trabajo original.

2.5 Consideraciones éticas

La revisión utilizó publicaciones y documentos públicos, sin reclutar participantes ni analizar historias clínicas o datos personales individuales. El análisis normativo tiene finalidad académica. Las propuestas para Ecuador derivan de la interpretación de la literatura y no describen prácticas u opiniones representativas de profesionales ecuatorianos.

3 Síntesis de la evidencia

3.1 Características generales de la evidencia

La bibliografía discutida combina síntesis, ensayos, estudios observacionales y cualitativos, comparaciones de respuestas y aportes conceptuales. El ensayo argentino de Tess [15] aporta un contexto latinoamericano. La revisión tecnológica de Garzón-Quiroz et al. [9], con afiliaciones ecuatorianas, contribuye al contexto técnico, pero no constituye validación clínica en Ecuador. Las fuentes aquí citadas no aportan ensayos de validación de chatbots terapéuticos en población ecuatoriana. La selección narrativa no permite afirmar que tales investigaciones sean inexistentes.

Los productos examinados fueron heterogéneos. Se discutieron agentes basados en reglas, sistemas generativos e intervenciones híbridas. Las condiciones también variaron entre uso autoguiado, monitoreo por investigadores y apoyo humano programado. Los modelos generales aparecieron en pruebas de comunicación y seguridad con escenarios simulados. Esta diversidad impide atribuir a la IA conversacional una eficacia única o extrapolar entre productos sin evaluación específica.

3.2 Eficacia clínica y alcance de los beneficios

He et al. [11] sintetizaron 32 ensayos de agentes conversacionales, con 6089 participantes, e informaron efectos pequeños al finalizar la intervención sobre depresión y ansiedad generalizada (g = 0,29 en ambos desenlaces). En seguimiento persistió un efecto pequeño sobre depresión (g = 0,16, IC 95 % 0,06 a 0,26), mientras que la mayoría de los demás desenlaces no mostró efectos significativos. Esta revisión abarcó agentes diversos, no únicamente chatbots de texto. Li et al. [18] revisaron 35 estudios y metaanalizaron 15 ensayos. Encontraron efectos favorables sobre depresión (g = 0,64) y malestar (g = 0,70), pero no resultados significativos para ansiedad o bienestar general. La heterogeneidad fue elevada.

Zhang et al. [33] metaanalizaron 14 ensayos de intervenciones generativas o híbridas y estimaron un efecto global de 0,30 (IC 95 % 0,004 a 0,59). El intervalo de predicción, de −0,85 a 1,67, refleja incertidumbre considerable sobre resultados en nuevos contextos. Nueve ensayos presentaron algunas preocupaciones y cinco alto riesgo de sesgo según la evaluación de esa revisión. Du et al. [8] encontraron una reducción pequeña de depresión con agentes basados en reglas (g = 0,266). La estimación para modelos de lenguaje fue imprecisa y no significativa. Las comparaciones entre tipos de chatbot tampoco fueron significativas, por lo que la diferencia entre un resultado significativo y otro no significativo no demuestra superioridad entre arquitecturas.

Los ensayos muestran la importancia del comparador y del seguimiento. En Tess, 181 estudiantes fueron asignados al azar y 73 aportaron datos a las ocho semanas. Se observó mejoría intragrupo de ansiedad, pero no diferencias significativas entre grupos [15]. XiaoE superó a un libro electrónico y a un chatbot general después de una semana. La diferencia al mes fue menor y su significación frente al libro no se mantuvo en todos los análisis de sensibilidad [12]. En Fido mejoraron ambos grupos, sin superioridad demostrada frente a autoayuda [14]. En el ensayo de Wysa se asignaron 79 personas y se analizaron 68. Hubo mejoría de depresión y ansiedad frente a ausencia de intervención, con las limitaciones de un control pasivo y de excluir participantes del análisis [19].

Heinz et al. [13] asignaron 210 participantes a Therabot o lista de espera. La intervención de cuatro semanas mostró beneficios para síntomas depresivos, ansiedad generalizada y riesgo de trastornos alimentarios, con evaluación adicional a las ocho semanas. El acceso al chatbot continuó durante el seguimiento. El estudio empleó supervisión humana de las respuestas y no incluyó un comparador de psicoterapia profesional.

Sohn et al. [27] sintetizaron 39 estudios y observaron efectos favorables sobre depresión (g = 0,31, IC 95 % 0,17 a 0,46) y ansiedad (g = 0,28, IC 95 % 0,05 a 0,51). Treinta y cinco estudios presentaron alto riesgo de sesgo. Gong et al. [10], en 29 ensayos de chatbots orientados a terapia cognitivo-conductual, estimaron efectos al finalizar la intervención de g = −0,55 para depresión y g = −0,26 para ansiedad. En seguimiento persistió un efecto pequeño para depresión (g = −0,32, IC 95 % −0,55 a −0,09), mientras que el de ansiedad no fue significativo (g = −0,19, IC 95 % −0,43 a 0,04). La certeza fue de muy baja a baja. Los signos responden a convenciones de codificación distintas y no deben compararse sin armonizarlas.

STARS, en Jordania, asignó al azar a 344 participantes de 18 a 21 años [4]. A los tres meses mostró beneficios frente a atención habitual mejorada sobre ansiedad (tamaño de efecto 0,70, IC 95 % 0,41 a 0,98) y depresión (0,61, IC 95 % 0,31 a 0,90). Combinó un chatbot basado en reglas con cinco llamadas semanales de 15 minutos de ayudantes formados y supervisados. Excluyó riesgo suicida inminente. Los resultados respaldan el paquete guiado estudiado y no permiten atribuir sus efectos al chatbot por separado ni establecer seguridad en crisis.

Tabla 1. Evidencia representativa sobre eficacia y alianza de agentes conversacionales

Estudio	Diseño y muestra	Hallazgo principal	Límite decisivo
He et al. [11]	32 ensayos, n = 6089	Depresión y ansiedad g = 0,29 al finalizar	Agentes diversos. La mayoría de los efectos en seguimiento no fue significativa
Li et al. [18]	35 estudios, 15 ensayos metaanalizados	Depresión g = 0,64. Malestar g = 0,70	Ansiedad y bienestar no significativos. Alta heterogeneidad
Zhang et al. [33]	26 estudios, 14 ensayos generativos o híbridos	Efecto global 0,30	Ningún ensayo con bajo riesgo de sesgo. Predicción muy amplia
Du et al. [8]	Metaanálisis, 15 estudios	Efecto pequeño para reglas sobre depresión	Sin diferencia significativa entre tipos de chatbot
Gong et al. [10]	29 ensayos de orientación cognitivo-conductual	Al finalizar: depresión g = −0,55 y ansiedad g = −0,26	Certeza muy baja a baja. Ansiedad no significativa en seguimiento
Sohn et al. [27]	39 estudios	Depresión g = 0,31. Ansiedad g = 0,28	35 estudios con alto riesgo de sesgo
Klos et al. [15]	181 asignados, 73 con datos a ocho semanas	Mejoría intragrupo de ansiedad con Tess	Sin diferencia significativa entre grupos. Pérdidas elevadas
He et al. [12]	Ensayo de tres brazos, n = 148	XiaoE superó a ambos controles a una semana	Efecto menor al mes y sensible al análisis frente al libro
Karkosz et al. [14]	81 asignados, 74 evaluados tras intervención	Mejoría en chatbot y autoayuda	Sin superioridad demostrada. No prueba equivalencia
MacNeill et al. [19]	79 asignados, 68 analizados	Beneficios de Wysa frente a ausencia de intervención	Control pasivo y análisis con exclusiones
Heinz et al. [13]	Ensayo, n = 210	Beneficios de Therabot frente a lista de espera	Supervisión humana. Cuatro semanas y seguimiento hasta ocho
Bryant et al. [4]	STARS, n = 344, Jordania	A tres meses: ansiedad 0,70 y depresión 0,61	Chatbot con llamadas de apoyo. Excluyó riesgo suicida inminente
Darcy et al. [7] y Beatty et al. [3]	Estudios observacionales	Alianza percibida con agentes digitales	Sin grupo concurrente de psicoterapia ni prueba de equivalencia

Los resultados disponibles se concentran en síntomas y periodos de seguimiento breves. La evidencia sintetizada no permite establecer una estimación única para todos los agentes ni determinar su desempeño en recaídas, funcionamiento cotidiano, eventos adversos infrecuentes y poblaciones de alta complejidad. La Tabla 1 reúne publicaciones representativas y no constituye una lista de ensayos independientes, debido al posible solapamiento entre revisiones.

3.3 Alianza, contención y experiencia emocional

Darcy et al. [7] analizaron retrospectivamente respuestas de 36 070 usuarios de Woebot en un diseño transversal. Las comparaciones con psicoterapia procedían de estudios publicados, sin un grupo humano concurrente. Beatty et al. [3] estudiaron Wysa con 1205 valoraciones iniciales y 226 repetidas tras al menos tres días, además de comentarios cualitativos de usuarios. Ambos trabajos describen alianza percibida. Sus diseños y posibles sesgos de selección no establecen equivalencia clínica con psicoterapia. En ambos hubo autores vinculados a las empresas de los productos evaluados.

La revisión narrativa de Tong et al. [29] describe la alianza digital como un concepto en desarrollo y señala la necesidad de delimitar sus componentes. Los puntajes de vínculo y satisfacción no permiten, por sí solos, establecer equivalencia con una relación terapéutica humana. La reciprocidad y la responsabilidad son dimensiones conceptuales que no quedan resueltas mediante escalas de alianza percibida.

El análisis conceptual de Nwosu et al. [23], sobre terapias digitales en general, señala problemas de abandono y de medición de la participación. No permite asignar una tasa única de deserción a los chatbots. Las revisiones éticas describen posibles riesgos de sobreconfianza y revelación de información sensible [31, 32]. La frecuencia de estos riesgos no puede estimarse con esta selección narrativa.

3.4 Límites clínicos y técnicos

Las publicaciones examinadas describen limitaciones en la indagación clínica, la interpretación contextual y la respuesta ante situaciones complejas [22, 26]. Las revisiones identifican también incertidumbre sobre continuidad, cambios de versión, privacidad y atribución de responsabilidades [5, 31]. Estos resultados proceden de diseños distintos y no equivalen a una evaluación integral de competencia profesional.

En la comparación de Scholich et al. [26], los chatbots generales ofrecieron más afirmaciones y sugerencias, solicitaron menos elaboración y mostraron dificultades en escenarios de crisis. El estudio examinó respuestas conversacionales y no resultados de un tratamiento longitudinal. Por tanto, permite identificar diferencias comunicativas, pero no estimar por sí solo la efectividad comparativa de una psicoterapia completa.

McBain et al. [20] evaluaron tres modelos mediante la calificación de respuestas escritas ante ideación suicida. Dos obtuvieron puntuaciones comparables o superiores a referencias publicadas de profesionales, mientras que el tercero se aproximó a un grupo sin formación. Todos tendieron a sobrevalorar la adecuación de las respuestas respecto al criterio experto. Entre el 11 % y el 36 % de sus calificaciones se consideraron atípicas, según el modelo. Se evaluaron juicios sobre respuestas hipotéticas, no intervenciones en crisis reales.

3.5 Riesgos éticos, emocionales y de seguridad

Las fuentes describen riesgos de información incorrecta, respuesta inadecuada a crisis, estigma, sobreconfianza y exposición de datos sensibles, además de sesgo y dificultades de atribución de responsabilidad [5, 16, 22, 31, 32]. Debe distinguirse entre fallas observadas en pruebas, riesgos propuestos en análisis éticos y daños clínicos efectivamente documentados. Esta revisión no estima su incidencia ni permite concluir que se produzcan con igual frecuencia en todos los sistemas.

Moore et al. [22] combinaron una revisión con experimentos mediante escenarios simulados de interacción terapéutica. Observaron estigma y respuestas que reforzaban interpretaciones delirantes en los modelos evaluados. La expresión naturalista utilizada por el estudio alude al formato conversacional de las pruebas, no a seguimiento clínico de pacientes. Sus resultados dependen de los escenarios, instrucciones y versiones examinados.

Campellone et al. [6] asignaron a 160 participantes a una intervención generativa acotada o a otra basada en reglas durante dos semanas. Las valoraciones de experiencia fueron descriptivamente próximas, pero el estudio no tuvo potencia para establecer equivalencia entre grupos. Los filtros cumplieron los criterios predefinidos en la revisión del ensayo, aunque se identificaron cuatro respuestas no deseadas. No se informaron eventos adversos graves ni relacionados con el dispositivo. Este resultado no garantiza seguridad fuera de las condiciones y del periodo estudiados.

La Tabla 2 reúne las publicaciones representativas sobre límites, seguridad y práctica profesional.

Tabla 2. Evidencia representativa sobre límites, seguridad y práctica profesional

Estudio	Aporte	Resultado y alcance
Scholich et al. [26]	17 terapeutas y tres chatbots, escenarios simulados	Más sugerencias y menos indagación. Dificultades en crisis. No evalúa un tratamiento completo
McBain et al. [20]	Calificación de respuestas hipotéticas ante ideación suicida	Sobrevaloración de adecuación y variación entre modelos. No prueba seguridad clínica
Moore et al. [22]	Revisión y experimentos conversacionales simulados	Estigma y refuerzo de interpretaciones delirantes en versiones y escenarios concretos
Campellone et al. [6]	Generación frente a reglas, n = 160, dos semanas	Filtros cumplieron criterios del ensayo, con cuatro respuestas no deseadas. No prueba equivalencia
Wagner y Schwind [30]	181 casos completos, 179 analizados, Alemania	Actitudes asociadas a utilidad y afinidad tecnológica. No evalúa formación ni eficacia
Kuang et al. [17]	Entrevistas a 18 psicoterapeutas, Estados Unidos	Confianza para ideas y borradores revisables, con reservas ante usos sensibles
Wang et al. [31] y Bucher et al. [5]	Revisiones sistemáticas	Riesgos de privacidad y seguridad y limitada evaluación clínica. No estiman una tasa común de daño

3.6 Percepciones de psicoterapeutas

Wagner y Schwind [30] estudiaron actitudes hacia aplicaciones de IA y aprendizaje automático en psicoterapeutas de Alemania. Describieron 181 casos completos y utilizaron 179 en los análisis estadísticos finales. Las actitudes se asociaron con utilidad percibida y afinidad tecnológica, sin demostrar un efecto causal de formación. Kuang et al. [17] entrevistaron a 18 psicoterapeutas estadounidenses. Identificaron mayor confianza para ideas o borradores revisables y reservas ante documentación sensible, decisiones clínicas y sustitución de la relación humana. Estos estudios no prueban seguridad clínica ni representan a profesionales ecuatorianos.

4 Discusión

4.1 Interpretación principal

La literatura revisada muestra beneficios sintomáticos a corto plazo para determinados productos y condiciones de uso, junto con limitaciones de comunicación y seguridad. Sin evaluación formal del riesgo de sesgo y ante la heterogeneidad de los diseños, no puede establecerse equivalencia general con psicoterapia profesional. La complementariedad supervisada se plantea como una orientación prudente para evaluar, no como un modelo cuya superioridad ya esté demostrada.

Therabot [13] ejemplifica la necesidad de interpretar cada producto en sus condiciones de evaluación. En 210 participantes, produjo mayores reducciones de síntomas que una lista de espera a las cuatro y ocho semanas. Todos sus mensajes fueron revisados después de enviarse por clínicos e investigadores, con posibilidad de contactar a los participantes para corregir respuestas o atender riesgos. Se excluyeron suicidabilidad activa, manía y psicosis. Los beneficios de este sistema ajustado por expertos y monitoreado no demuestran eficacia o seguridad de cualquier chatbot comercial ni equivalencia con psicoterapia.

Los tamaños de efecto también requieren contexto. Un efecto pequeño puede ser valioso si la intervención es barata, accesible y llega a personas sin atención. Puede ser insuficiente si reemplaza un tratamiento eficaz, demora una derivación o expone datos sensibles. La decisión debe considerar beneficio incremental, oportunidad perdida y riesgo, no solo significación estadística. En sistemas con brechas de acceso, la IA puede ampliar una primera capa

de apoyo, pero no debe convertirse en una atención de menor calidad reservada para poblaciones con menos recursos.

4.2 Transformación del rol profesional

La automatización podría modificar tareas de documentación, preparación de materiales o seguimiento rutinario. También podría generar trabajo adicional de verificación, supervisión y respuesta a incidentes. Estas posibilidades no deben confundirse con resultados laborales comprobados: la revisión no permite estimar cuántos puestos cambiarán, qué tareas desaparecerán ni en qué plazo. Los efectos dependerán del producto, del servicio y de la regulación.

Es necesario distinguir la automatización de una tarea delimitada de la sustitución de una profesión. Entregar material psicoeducativo o proponer un borrador no supone asumir evaluación, formulación, decisiones terapéuticas, manejo de crisis y responsabilidad. La revisión no demuestra el reemplazo seguro de ese conjunto integrado de funciones. Tampoco permite asegurar que toda tarea rutinaria pueda automatizarse sin riesgos.

El concepto de "obsolescencia" confunde tres niveles. En el nivel técnico, una herramienta puede producir texto empático. En el nivel clínico, debe demostrar beneficio, no inferioridad o superioridad, seguridad y generalización. En el nivel profesional, debe asumir responsabilidades, integrarse al sistema y responder ante daño. La evidencia cubre parcialmente el primero, de manera limitada el segundo y no resuelve el tercero. Por eso, una conversación convincente no equivale a una profesión reemplazada.

Como implicación para la formación, se propone incluir competencias para distinguir arquitectura y finalidad, valorar evidencia, reconocer información inventada y complacencia, revisar condiciones de datos y documentar límites. La participación de profesionales en diseño, evaluación y vigilancia es coherente con las propuestas de desarrollo responsable y gobernanza [24, 25, 28]. Esta revisión no evaluó un programa formativo ni su efecto sobre resultados clínicos.

4.3 Contextualización y propuesta para Ecuador

La OMS propone que la IA en salud proteja autonomía, seguridad, transparencia, equidad, rendición de cuentas y sostenibilidad [25]. Su guía para modelos multimodales añade recomendaciones para gobiernos, desarrolladores y proveedores, con énfasis en supervisión, evaluación del riesgo y participación de las partes afectadas [24]. Estos principios son aplicables a la conversación clínica, donde los datos son íntimos y una respuesta inadecuada puede modificar conductas.

En Ecuador, los artículos 30 y 31 de la Ley Orgánica de Protección de Datos Personales regulan el tratamiento de datos relativos a la salud. Establecen confidencialidad, medidas técnicas y organizativas de seguridad y, cuando sea posible, anonimización o seudonimización, junto con otras condiciones que deben examinarse según el caso [1]. Trasladar relatos clínicos a servicios externos exige considerar quién trata y conserva los datos. Como salvaguarda propuesta, no basta con retirar el nombre si el relato permite identificar a la persona por otros detalles.

Se consideran también documentos nacionales de referencia la Ley Orgánica de Salud Mental, publicada en 2024, y el acuerdo que aprueba la Política Nacional de Salud Mental 2025–2030 [2, 21]. Sus registros oficiales acreditan su publicación y no una validación clínica de psicoterapia autónoma mediante IA. Esta revisión no es un inventario exhaustivo de la normativa ecuatoriana. Una implementación concreta requiere examinar las obligaciones aplicables al servicio, al tratamiento de datos y a la responsabilidad profesional.

Como propuesta de esta revisión, se plantea investigar un modelo complementario de atención escalonada en Ecuador. La Tabla 3 organiza posibles usos, responsabilidades y salvaguardas. Su contenido no procede de un ensayo ecuatoriano ni acredita eficacia o seguridad. Antes de implementarlo deben evaluarse selección de usuarios, exclusiones, supervisión, derivación y efectos adversos. La propuesta reserva las decisiones de alta complejidad y la atención de crisis a la responsabilidad profesional directa.

Tabla 3. Propuesta de funciones y salvaguardas para Ecuador que requiere validación local

Función	Papel posible de la IA	Responsabilidad del psicólogo clínico	Salvaguarda mínima
Psicoeducación y habilidades	Entregar contenido y ejercicios estructurados	Seleccionar, adaptar y verificar comprensión	Contenido validado. Identidad de IA visible. Consentimiento
Tamizaje y monitoreo	Administrar escalas y mostrar tendencias	Interpretar resultados. Descartar diagnósticos diferenciales	Instrumentos validados. Puntos de corte locales. Derivación
Práctica entre sesiones	Recordar tareas y guiar ejercicios de bajo riesgo	Integrar la práctica al plan y revisar efectos	Opción de salida. Registro auditable. Límites de uso
Documentación	Proponer borradores o resúmenes desidentificados	Revisar, corregir y firmar	No usar chatbots públicos. Minimización y contrato de datos
Diagnóstico y plan terapéutico	Apoyo informativo, nunca decisión autónoma	Evaluar, formular y decidir	Validación clínica. Trazabilidad. Responsabilidad humana
Crisis, suicidio, psicosis o violencia	Detectar señales y activar transferencia	Evaluar e intervenir de forma directa	Derivación humana inmediata. Sin manejo autónomo de la crisis
Contención y alianza	Apoyo limitado y explícitamente artificial	Sostener vínculo, límites, reparación y responsabilidad	No antropomorfizar. Prevenir dependencia. Supervisar efectos

4.4 Recomendaciones de implementación e investigación

Antes de una adopción clínica en Ecuador se requieren pilotos prospectivos con español local, diversidad regional, accesibilidad y evaluación cultural. Deben incluir comparadores activos, seguimiento suficiente, registro de eventos adversos, abandono, funcionamiento y calidad de vida, además de síntomas. Los protocolos de crisis han de probarse con recursos ecuatorianos y no limitarse a líneas de ayuda de otros países.

Cada producto debería disponer de una ficha clínica: versión del modelo, finalidad, población excluida, evidencia, datos recolectados, lugar de almacenamiento, tiempo de conservación, responsables, procedimiento de actualización, sesgos conocidos, criterios de escalamiento y contacto humano. Un cambio sustancial de modelo o guardas debe tratarse como una nueva versión que requiere reevaluación. La auditoría necesita muestras en español, variaciones dialectales y escenarios de ambigüedad, no únicamente instrucciones ideales.

La investigación debería comparar, cuando exista justificación ética, atención habitual y atención apoyada por IA supervisada, con criterios explícitos de seguridad y beneficio incremental. La evaluación de intervenciones autónomas requiere una justificación específica según el riesgo y no debe asumirse apropiada para toda población. Son relevantes los estudios de efectividad, no inferioridad debidamente diseñada, costo-efectividad, dependencia emocional y carga profesional. La participación de usuarios y psicólogos ecuatorianos resulta necesaria para valorar aceptabilidad y pertinencia local.

4.5 Limitaciones de la revisión

La selección fue intencional y no exhaustiva. PubMed fue el único índice consultado mediante estrategias documentadas. No se completó el cribado de sus 416 registros y la actualización dirigida puede omitir estudios pertinentes. No se consultaron sistemáticamente PsycINFO, Scopus, Web of Science, LILACS o SciELO. La bibliografía de partida carece de un historial verificable de selección y no permite reconstruir su cobertura original. Estas limitaciones impiden presentar los trabajos citados como el conjunto completo de evidencia elegible.

No hubo revisión humana duplicada independiente ni evaluación formal propia del riesgo de sesgo o de la certeza. El cotejo asistido por IA de los pasajes utilizados mejora la trazabilidad, pero no elimina errores de interpretación ni el sesgo de selección. La heterogeneidad y el solapamiento de las síntesis limitan comparaciones. Los cambios de versión restringen la generalización temporal. Las propuestas para Ecuador son interpretativas y requieren validación local.

5 Conclusiones

Los trabajos discutidos sugieren beneficios a corto plazo de algunos agentes diseñados para salud mental, condicionados por el producto, la población, el comparador y el apoyo humano asociado. Las revisiones recientes también señalan heterogeneidad y limitaciones de certeza. Esta síntesis narrativa no establece una sustitución segura del psicólogo clínico ni equivalencia general con psicoterapia profesional. Tampoco permite predecir definitivamente las capacidades futuras de la tecnología.

Para Ecuador se propone investigar usos complementarios con supervisión profesional, protección de datos y rutas humanas de derivación. La distribución de funciones presentada requiere evaluación clínica y cultural local, seguimiento de eventos adversos y responsabilidades definidas. No constituye un modelo validado ni una autorización normativa. La revisión tampoco permite pronosticar la obsolescencia de la profesión o garantizar la seguridad de la automatización de tareas específicas.

Referencias

[1] Asamblea Nacional del Ecuador. Ley Orgánica de Protección de Datos Personales. Registro Oficial, Quinto Suplemento No. 459, 26 de mayo de 2021. https://repositorio.consejodecomunicacion.gob.ec/handle/CONSEJO_REP/246

[2] Asamblea Nacional del Ecuador. Ley Orgánica de Salud Mental. Registro Oficial, Suplemento No. 471, 5 de enero de 2024. https://www.registroficial.gob.ec/suplemento-al-registro-oficial-no-471/

[3] Beatty, C., Malik, T., Meheli, S., y Sinha, C. Evaluating the therapeutic alliance with a free-text CBT conversational agent (Wysa): A mixed-methods study. *Frontiers in Digital Health*, 4, 847991, 2022. doi: 10.3389/fdgth.2022.847991

[4] Bryant, R. A., de Graaff, A. M., Habashneh, R., Fanatseh, S., Keyan, D., Akhtar, A., Abualhaija, A., Faroun, M., Aqel, I. S., Dardas, L., Afar, H., Servili, C., Hadzi-Pavlovic, D., van Ommeren, M., y Carswell, K. A guided chatbot-based psychological intervention for psychologically distressed older adolescents and young adults: a randomised clinical trial in Jordan. *npj Digital Medicine*, 9(1), 57, 2026. doi: 10.1038/s41746-025-02142-8

[5] Bucher, A., Egger, S., Vashkite, I., Wu, W., y Schwabe, G. "It's not only attention we need": Systematic review of large language models in mental health care. *JMIR Mental Health*, 12, e78410, 2025. doi: 10.2196/78410

[6] Campellone, T. R., Flom, M., Montgomery, R. M., Bullard, L., Pirner, M. C., Pavez, A., Morales, M., Harper, D., Oddy, C., O'Connor, T., Daniels, J., Eaneff, S., Forman-Hoffman, V. L., Sackett, C., y Darcy, A. Safety and user experience of a generative artificial intelligence digital mental health intervention: Exploratory randomized controlled trial. *Journal of Medical Internet Research*, 27, e67365, 2025. doi: 10.2196/67365

[7] Darcy, A., Daniels, J., Salinger, D., Wicks, P., y Robinson, A. Evidence of human-level bonds established with a digital conversational agent: Cross-sectional, retrospective observational study. *JMIR Formative Research*, 5(5), e27868, 2021. doi: 10.2196/27868

[8] Du, Q., Ren, Y., Meng, Z. L., He, H., y Meng, S. The efficacy of rule-based versus large language model-based chatbots in alleviating symptoms of depression and anxiety: Systematic review and meta-analysis. *Journal of Medical Internet Research*, 27, e78186, 2025. doi: 10.2196/78186

[9] Garzón-Quiroz, M., Del Campo-Saltos, G., y Loor-Ávila, B. Análisis sistemático sobre la eficiencia comunicativa entre chatbots basados en reglas y modelos de lenguaje natural. *Universitas-XXI*, 42, 167-192, 2025. doi: 10.17163/uni.n42.2025.07

[10] Gong, B., Yao, N., Xie, H., Huang, C., Kishimoto, T., Berenbaum, H., y Mu, W. Efficacy, User Engagement, and Acceptability of Cognitive Behavioral Therapy-Oriented Psychological Chatbots for Adults With Depressive and/or Anxiety Symptoms: Systematic Review and Meta-Analysis of Randomized Controlled Trials. *Journal of Medical Internet Research*, 28, e82677, 2026. doi: 10.2196/82677

[11] He, Y., Yang, L., Qian, C., Li, T., Su, Z., Zhang, Q., y Hou, X. Conversational agent interventions for mental health problems: Systematic review and meta-analysis of randomized controlled trials. *Journal of Medical Internet Research*, 25, e43862, 2023. doi: 10.2196/43862

[12] He, Y., Yang, L., Zhu, X., Wu, B., Zhang, S., Qian, C., y Tian, T. Mental health chatbot for young adults with depressive symptoms during the COVID-19 pandemic: Single-blind, three-arm randomized controlled trial. *Journal of Medical Internet Research*, 24(11), e40719, 2022. doi: 10.2196/40719

[13] Heinz, M. V., Mackin, D. M., Trudeau, B. M., Bhattacharya, S., Wang, Y., Banta, H. A., Jewett, A. D., Salzhauer, A. J., Griffin, T. Z., y Jacobson, N. C. Randomized trial of a generative AI chatbot for mental health treatment. *NEJM AI*, 2(4), AIoa2400802, 2025. doi: 10.1056/AIoa2400802

[14] Karkosz, S., Szymański, R., Sanna, K., y Michałowski, J. Effectiveness of a web-based and mobile therapy chatbot on anxiety and depressive symptoms in subclinical young adults: Randomized controlled trial. *JMIR Formative Research*, 8, e47960, 2024. doi: 10.2196/47960

[15] Klos, M. C., Escoredo, M., Joerin, A., Lemos, V. N., Rauws, M., y Bunge, E. L. Artificial intelligence-based chatbot for anxiety and depression in university students: Pilot randomized controlled trial. *JMIR Formative Research*, 5(8), e20678, 2021. doi: 10.2196/20678

[16] Kolding, S., Lundin, R. M., Hansen, L., y Østergaard, S. D. Use of generative artificial intelligence (AI) in psychiatry and mental health care: A systematic review. *Acta Neuropsychiatrica*, 37, e37, 2025. doi: 10.1017/neu.2024.50

[17] Kuang, J., Pope, A. L., y Zhang, Y. Psychotherapists' trust, distrust, and generative AI practices in psychotherapy: Qualitative study. *Journal of Medical Internet Research*, 28, e88932, 2026. doi: 10.2196/88932

[18] Li, H., Zhang, R., Lee, Y.-C., Kraut, R. E., y Mohr, D. C. Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. *npj Digital Medicine*, 6, 236, 2023. doi: 10.1038/s41746-023-00979-5

[19] MacNeill, A. L., Doucet, S., y Luke, A. Effectiveness of a mental health chatbot for people with chronic diseases: Randomized controlled trial. *JMIR Formative Research*, 8, e50025, 2024. doi: 10.2196/50025

[20] McBain, R. K., Cantor, J. H., Zhang, L. A., Baker, O., Zhang, F., Halbisen, A., Kofner, A., Breslau, J., Stein, B., Mehrotra, A., y Yu, H. Competency of large language models in evaluating appropriate responses to suicidal ideation: Comparative study. *Journal of Medical Internet Research*, 27, e67891, 2025. doi: 10.2196/67891

[21] Ministerio de Salud Pública del Ecuador. Política Nacional de Salud Mental 2025–2030. Acuerdo Ministerial 00001-2025. Registro Oficial, Segundo Suplemento No. 188, 19 de diciembre de 2025. https://www.registroficial.gob.ec/segundo-suplemento-no-188/

[22] Moore, J., Grabb, D., Agnew, W., Klyman, K., Chancellor, S., Ong, D. C., y Haber, N. 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). Association for Computing Machinery, 2025. doi: 10.1145/3715275.3732039

[23] Nwosu, A., Boardman, S., Husain, M. M., y Doraiswamy, P. M. Digital therapeutics for mental health: Is attrition the Achilles heel? *Frontiers in Psychiatry*, 13, 900615, 2022. doi: 10.3389/fpsyt.2022.900615

[24] Organización Mundial de la Salud. Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models, 2024. https://www.who.int/publications/i/item/9789240084759

[25] Organización Mundial de la Salud. Ethics and governance of artificial intelligence for health: WHO guidance, 2021. https://www.who.int/publications/i/item/9789240029200

[26] Scholich, T., Barr, M., Wiltsey Stirman, S., y Raj, S. A comparison of responses from human therapists and large language model-based chatbots to assess therapeutic communication: Mixed methods study. *JMIR Mental Health*, 12, e69709, 2025. doi: 10.2196/69709

[27] Sohn, J. S., Ha, B. G., Park, S., Kim, J. J., Lee, E., Oh, H., Lee, S., y Kim, E. Systematic review and meta analysis of chatbots in the management of depressive and anxiety symptoms. *npj Digital Medicine*, 9(1), 377, 2026. doi: 10.1038/s41746-026-02566-w

[28] Stade, E. C., Stirman, S. W., Ungar, L. H., Boland, C. L., Schwartz, H. A., Yaden, D. B., Sedoc, J., DeRubeis, R. J., Willer, R., y Eichstaedt, J. C. Large language models could change the future of behavioral healthcare: A proposal for responsible development and evaluation. *npj Mental Health Research*, 3, 12, 2024. doi: 10.1038/s44184-024-00056-z

[29] Tong, F., Lederman, R., D'Alfonso, S., Berry, K., y Bucci, S. Digital therapeutic alliance with fully automated mental health smartphone apps: A narrative review. *Frontiers in Psychiatry*, 13, 819623, 2022. doi: 10.3389/fpsyt.2022.819623

[30] Wagner, J., y Schwind, A.-S. Investigating psychotherapists' attitudes towards artificial intelligence in psychotherapy. *BMC Psychology*, 13, 719, 2025. doi: 10.1186/s40359-025-03071-7

[31] Wang, X., Zhou, Y., y Zhou, G. The application and ethical implication of generative AI in mental health: Systematic review. *JMIR Mental Health*, 12, e70610, 2025. doi: 10.2196/70610

[32] Xian, X., Chang, A., Xiang, Y.-T., y Liu, M. T. Debate and dilemmas regarding generative AI in mental health care: Scoping review. *Interactive Journal of Medical Research*, 13, e53672, 2024. doi: 10.2196/53672

[33] Zhang, Q., Zhang, R., Xiong, Y., Sui, Y., Tong, C., y Lin, F. H. Generative AI mental health chatbots as therapeutic tools: Systematic review and meta-analysis of their role in reducing mental health issues. *Journal of Medical Internet Research*, 27, e78238, 2025. doi: 10.2196/78238