

Exploring the Influence of Chatbots on Healthcare: Systematic Review

Claudia Boza Bocanegra¹, Diego Peña², Enrique R. Yapuchura³, David Villena-Reyes⁴

^{1,2,3,4}Universidad Tecnológica del Perú, Lima-Perú

claudiaboza.1@gmail.com, diego20001210@gmail.com, c22749@utp.edu.pe, david.villena.reyes@gmail.com

Abstract-This RSL (Systematic Literature Review) explores the impact of chatbots on online healthcare, with a focus on patients who, for various reasons, use online healthcare services. The objectives included evaluating the effectiveness of chatbots and identifying their uses and challenges, Likewise proposing recommendations for their implementation. The search strategy was based on the PICO method or framework of questions that help us structure and focus the bibliographic search. Likewise, the Scopus database was used as a bibliographic source. 58 articles were reviewed, and 15 were relevant for the analysis. The studies highlighted the versatility of chatbots in medical contexts and their ability to improve the quality of care and patient satisfaction, especially in perinatal mental care. Accuracy measurements, such as P@1, R@3, and MRR, demonstrated the effectiveness of chatbots in providing relevant answers. The conclusions highlight the need for specific regulation for LLMs (Large Scale Language Models) and the importance of scalability for the success of chatbots. Future research on specific populations and the practical integration of chatbots into clinical settings was suggested. This RSL provides a comprehensive view of the emerging role of chatbots in online healthcare, identifying trends and key areas for future study.

Keywords- ChatBot, Automation, Artificial Intelligence, Patient-Chatbot Interaction, Healthcare

Digital Object Identifier: (only for full papers, inserted by LACCEI).
ISSN, ISBN: (to be inserted by LACCEI).
DO NOT REMOVE

Explorando la influencia de los chatbots en la atención sanitaria: revisión sistemática

Claudia Boza Bocanegra¹, Diego Peña², Enrique R. Yapuchura³, David Villena-Reyes⁴

^{1,2,3,4}Universidad Tecnológica del Perú, Lima-Perú

claudiaboza.1@gmail.com, diego20001210@gmail.com, c22749@utp.edu.pe, david.villena.reyes@gmail.com

Resumen- Esta RSL (Revisión Sistemática de Literatura) explora el impacto de los chatbots en la atención médica en línea, cuyo enfoque va hacia pacientes que por diversos motivos utilizan servicios de atención médica en línea. Los objetivos incluyeron evaluar la efectividad de los chatbots e identificar sus usos y desafíos; así mismo, proponer recomendaciones para su implementación. La estrategia de búsqueda fue basada en el método o marco de preguntas PICO que nos ayudan a estructurar y centrar la búsqueda bibliográfica. Asimismo, fue utilizado la base de datos Scopus como fuente bibliográfica. Se revisaron 58 artículos, seleccionando 15 relevantes para el análisis. Los estudios destacaron la versatilidad de los chatbots en contextos médicos, su capacidad para mejorar la calidad de la atención y la satisfacción del paciente, especialmente en la atención mental perinatal. Las mediciones de precisión, como P@1, R@3 y MRR, demostraron la eficacia de los chatbots en proporcionar respuestas relevantes. Las conclusiones resaltan la necesidad de regulación específica para los LLMs (Modelos de Lenguaje a Gran Escala) y la importancia de la escalabilidad para el éxito de los chatbots. Se sugirió una investigación futura enfocada en poblaciones específicas y la integración práctica de chatbots en entornos clínicos. Esta RSL ofrece una visión completa del papel emergente de los chatbots en la atención médica en línea, identificando tendencias y áreas clave para futuros estudios.

Palabras clave: ChatBot, Automatización, inteligencia artificial, interacción paciente-chatbot, atención médica

I. INTRODUCCIÓN

En el panorama actual, la revolución tecnológica ha transformado radicalmente la prestación de servicios de salud, siendo los chatbots (CB) un desarrollo sobresaliente en este ámbito. Diseñados como sistemas de inteligencia artificial, los chatbots no sólo ofrecen respuestas rápidas y precisas, sino que también desempeñan un papel crucial en la automatización de tareas repetitivas, mejorando así la eficiencia de la atención médica [1].

Aunque los chatbots ofrecen beneficios notables en el ámbito de la salud, su aceptación se ve obstaculizada por desafíos sustanciales. A pesar de ventajas evidentes como el anonimato, la comodidad y el acceso rápido a información relevante, la falta de familiaridad y confianza entre los usuarios restringe su adopción.

En el sector de la atención médica, persisten desafíos como largas esperas, limitado acceso en áreas remotas y una comunicación ineficiente entre profesionales de la salud y pacientes [2].

La implementación de chatbots en el servicio de salud emerge como una solución prometedora para abordar estos desafíos. La necesidad de realizar una revisión sistemática de literatura (RSL) se fundamenta en la urgencia de recopilar y sintetizar la evidencia disponible sobre el impacto de los chatbots en la atención médica. Esta revisión no solo permitirá una comprensión más profunda del estado actual de la tecnología, sino que también establecerá las bases para futuras investigaciones, contribuyendo significativamente al avance de la atención médica [3].

La estructura de la revisión se articulará alrededor de dos aspectos cruciales: la eficiencia y efectividad de los chatbots en la prestación de servicios de salud, y los desafíos y limitaciones inherentes a su aceptación por parte de los usuarios. Las investigaciones recientes, como la de [4], que destaca la eficacia y rapidez de los chatbots en el asesoramiento de salud, y trabajos previos, que exploran la aplicación de chatbots en la predicción y diagnóstico de enfermedades, serán abordadas para proporcionar una visión completa [2]. El objetivo principal de esta revisión sistemática de literatura es ofrecer una perspectiva integral del uso de chatbots en el servicio de salud. Se busca evaluar la eficacia de estos sistemas, identificar los desafíos existentes y proponer recomendaciones para su implementación efectiva.

El presente informe sigue una estructura que refleja la rigurosidad y profundidad de la investigación abordada. La sección 2, Metodología, constituye el punto de partida, donde se expone detalladamente el enfoque metodológico empleado para abordar el problema de estudio. La sección 3, Resultados, se erige como el núcleo central del documento, presentando de manera organizada y estructurada los hallazgos obtenidos tras el análisis exhaustivo de los trabajos primarios. En la sección 4, Discusión, se profundiza en la reflexión crítica sobre las fuentes y tecnologías seleccionadas. Finalmente, la sección 5, Conclusiones, sintetiza los hallazgos más relevantes y las limitaciones identificadas en el estudio de RSL. Además, se señalan las futuras direcciones que podrían explorarse en

Digital Object Identifier: (only for full papers, inserted by LACCEI).

ISSN, ISBN: (to be inserted by LACCEI).

DO NOT REMOVE

investigaciones posteriores, proporcionando una guía clara para aquellos que deseen profundizar en el tema. Este informe, estructurado de manera coherente, ofrece una contribución significativa al campo y establece las bases para la continuación y expansión de la investigación en la problemática abordada.

II. METODOLOGÍA

A. Descripción de la estrategia de búsqueda sistemática:

El presente estudio fue desarrollado en el formato de una revisión sistemática de literatura, que es una manera de evaluar e interpretar toda la investigación disponible, que sea relevante respecto de una interrogante de investigación particular, en un área temática o fenómeno de interés [5]. Además, la estrategia de búsqueda sistemática desempeña un papel fundamental en la investigación científica, ya que facilita la identificación y recopilación organizada y precisa de la información.

Para llevar a cabo este estudio, se empleó la estrategia conocida como PICO (Población, Intervención, Comparación, Resultado), una herramienta que estructura la investigación. Inicialmente, se formuló la pregunta de investigación: En pacientes que utilizan servicios de salud en línea, ¿Cómo afecta el uso de chatbots en la calidad de la atención y la satisfacción del paciente en comparación a la atención convencional? Por tanto, para extraer los componentes PICO relacionados con esta pregunta, se identificaron los siguientes elementos:

Población (P): Pacientes que utilizan servicios de salud en línea.

Intervención (I): Uso de chatbots en la atención de salud en línea.

Comparación (C): Atención convencional.

Resultado (O): Calidad de la atención y satisfacción del paciente.

Además, se desarrolló preguntas de investigación que responden a la pregunta PICO:
RQ1. ¿Cuáles son los efectos de usar chatbots en la atención médica?

RQ2. ¿Cuáles son los principales usos y aplicaciones de los chatbots en el contexto de la atención médica en línea?

RQ3. ¿Cuáles son las ventajas y desventajas de la atención convencional en comparación con el uso de chatbots?

RQ4. ¿Cómo se mide la calidad de la atención en el contexto de la atención de salud en línea?

Y adicionalmente, se hizo uso de la base de datos bibliográfica Scopus, que abarca diversas disciplinas académicas y técnicas.

TABLA I
PALABRAS CLAVES - SINTAXIS

ITEM	Componentes	Palabras Claves	Sintaxis de ecuación de búsqueda
P	Problema / Población	Usuarios de atención médica	"Users", "medical attention", "patient", "health service"
I	Intervención	Uso de chatbots	"Chatbot use", "use of chatbots", "chatbot"
C	Comparación	Atención convencional	"Attention", "conventional care"
O	Resultados	Calidad de atención y satisfacción del paciente	"Quality", "attention", "attention quality", "satisfaction"

Se presenta la ecuación de búsqueda empleada: ("users" OR "medical attention" OR "patient" OR "health service") AND ("chatbot use" OR "use of chatbots" OR "chatbot") AND ("quality" OR "attention" OR "attention quality" OR "satisfaction") AND ("attention" OR "conventional care")

B. Definición de Criterios de Inclusión y Exclusión:

En relación a los criterios de inclusión y exclusión, se establecieron para garantizar la relevancia de los artículos seleccionados en la investigación. Los criterios de inclusión son los siguientes:

CI 1 Estudios pertenecientes a ciencia de la computación, ingeniería, medicina, ciencia de la decisión, profesionales de la salud, enfermería y psicología.

CI 2 Artículos pertenecientes a los formatos de artículos de conferencia y artículos originales.

CI 3 Estudios que se encuentren en su etapa final de publicación.

CI 4 Estudios provenientes de Estados Unidos, Canadá, España, India, China y Reino Unido y Argentina.

Por otro lado, los criterios de exclusión son:

CE 1 Estudios con la palabra clave de entrevista motivacional.

CE 2 Estudios realizados antes del 2019.

CE 3 Estudios pertenecientes a las áreas de negocios y contabilidad.

C. Descripción del proceso de selección:

- Resultados obtenidos del proceso de búsqueda de literatura científica

En la búsqueda usando las palabras claves mostradas en la pregunta PICO se obtuvo un total de 58 artículos, los cuales mediante el uso de PRISMA se seleccionaron las investigaciones más adecuadas y con mayor relación a nuestro tema abordado.

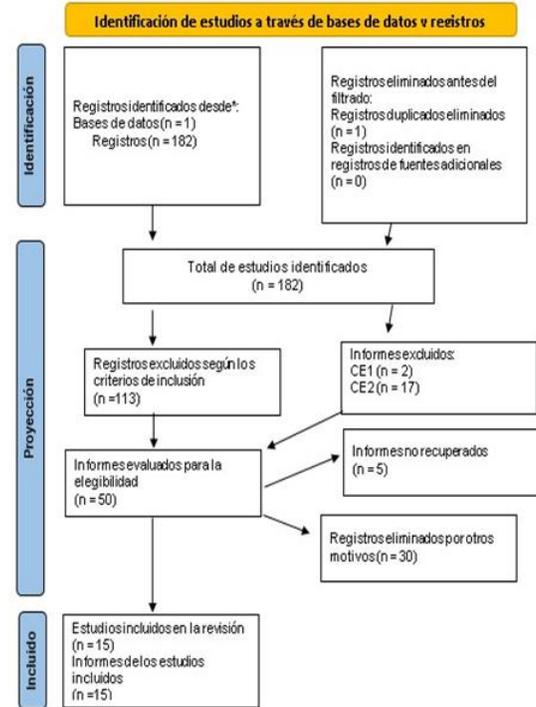
- Descripción de la lógica de selección considerada (PRISMA)

Se desarrolló un proceso de selección por etapas, basado en los lineamientos de la declaración PRISMA. Para lograr detallar los criterios y filtros usados en la literatura referente a nuestro tema de “El uso de chatbots en el servicio de salud”.

- Descripción detallada de los pasos del proceso de selección y sus resultados (PRISMA)

- Ecuación de búsqueda en base a pregunta pico se obtienen 182 artículos.
- Aplicamos filtro automatizado de Scopus. Según los CE1 y CE2 el cual elimina 19 artículos.
- Aplicamos filtro automatizado de Scopus. Según CI 1, CI 2, CI 3 y CI 4 los cuales eliminan 113 artículos.
- Se obtiene un resultado de 50 investigaciones a incluir en la revisión sistemática.
- Se eliminaron 30 artículos por que el contenido de la información no se relaciona con el tema de investigación.
- Se eliminan 5 informes debido a que no se lograron recuperar.
- Finalmente, se obtienen 15 artículos a incluir en la investigación sistemática.

- Diagrama de flujo PRISMA que refleja gráficamente el proceso



III. RESULTADOS

- Síntesis de estudios basados en la pregunta de revisión:

¿Qué definición de [CHATBOT] fue utilizada en el estudio?

En las investigaciones se emplean diversas definiciones clave en el ámbito de los efectos de los chatbots en la atención médica en línea. La referencia [1], se presenta el chatbot como un agente especial que responde al usuario en lenguaje natural, estableciendo así una sólida conexión emocional con el usuario. En este sentido, estos agentes conversacionales en el ámbito de la salud nos permiten automatizar procesos, mejorar la experiencia de los usuarios (pacientes) y, por supuesto, reducir costos. En el contexto específico de la investigación, un asistente virtual de salud mental se define como un sistema diseñado para abordar problemas de salud mental al proporcionar asistencia a los usuarios de manera virtual.

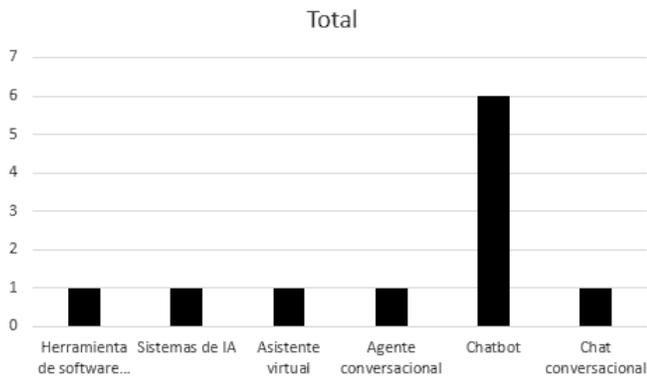


Gráfico 1. Definiciones usadas en los distintos estudios.

¿Qué tipo de [CHATBOT] fue discutido?

El asistente virtual propuesto contendría funciones como chat, evaluación psicológica, módulo de detección de emociones y un sistema de recomendación para mejorar el estado de ánimo del usuario [6]. En otro estudio, se aborda un chatbot terapéutico social inteligente con un enfoque específico en la salud mental de los estudiantes. Este hace uso de técnicas de procesamiento del lenguaje natural y aprendizaje profundo para la detección de emociones y la clasificación del estado mental del usuario. Además, se examina un tipo de agente conversacional denominado Pahola, desarrollado por la Organización Panamericana de la Salud (OPS) como parte de su respuesta a la pandemia de COVID-19.

TABLA II
TABLA COMPARATIVA QUE DESTACA LAS CARACTERÍSTICAS Y FUNCIONALIDADES CLAVE

Chatbot	Funciones	Tecnologías Utilizadas	Contexto
Asistente Virtual	Chat, evaluación psicológica, detección de emociones, sistema de recomendación.	No especificado	Propuesto para mejorar el estado de ánimo del usuario
Chatbot Terapéutico Social	Procesamiento del lenguaje natural, aprendizaje profundo, detección de emociones, clasificación del estado mental.	No especificado	Enfocado en la salud mental de estudiantes.
Pahola (OPS)	Tipo de agente conversacional.	Desarrollado en respuesta a la pandemia de COVID-19	Respuesta de la OPS a la pandemia de COVID-19.

¿A qué modelo teórico relacionado con [CHATBOT] se ajusta lo discutido en el estudio?

El modelo teórico relacionado con los LLMs, según el estudio de [7], se centra en la regulación y aplicación específica en el ámbito de la práctica médica. Este enfoque resalta los riesgos y desafíos particulares que estos modelos presentan en comparación con otras tecnologías de IA dentro del sector de la atención médica. Por otro lado, se propuso un modelo teórico adicional que se orienta hacia el desarrollo de agentes conversacionales respaldados por evidencia científica. Estos agentes tienen como objetivo interactuar con el público de manera accesible y anónima, abordando temas de salud. No obstante, el estudio subraya la necesidad de mejoras continuas en el contenido de estos agentes. Además, se planteó un modelo que implica la aplicación de técnicas de procesamiento del lenguaje natural (NLP) y aprendizaje

profundo. Este modelo tiene como finalidad analizar las interacciones del usuario y clasificar las emociones asociadas.

Importancia Relativa(%)



- Regulación y aplicación en la práctica médica
- Desarrollo de agentes conversacionales
- Aplicación de técnicas de NLP
- Aprendizaje Profundo

Gráfico 2. Importancia relativa de los modelos teóricos.

- Síntesis de estudios cuantitativos:

¿Qué mediciones, métricas, factores son mencionadas en torno a [CHATBOT]?

En el ámbito de los chatbots, se abordan diversas mediciones, métricas y factores [8], que incluyen aspectos como datos de autorrevelación, indicadores de estado mental, comportamiento de llevar un diario, análisis del lenguaje, análisis del uso de palabras, análisis de palabras emocionales, uso de palabras concretas, preguntas de investigación e hipótesis. Paralelamente, se propuso un estudio sobre un chatbot terapéutico destinado a abordar el estrés y la depresión en jóvenes. Entre las mediciones abordadas se encuentran la inteligencia del chatbot, el procesamiento del lenguaje natural (NLP), la terapia con chatbots y el empleo de algoritmos de aprendizaje profundo. Además, se subraya el papel crucial de los sistemas inteligentes, especialmente los chatbots basados en inteligencia artificial (IA), destacando su rol, el desarrollo de un chatbot analítico y los beneficios asociados con los chatbots basados en IA.

¿Qué tipo y tamaño de población, muestra, tipo de caso, etc, se utilizó?

La referencia [9], los autores presentan la propuesta de un chatbot destinado a monitorear y evaluar el estado mental de las mujeres perinatales. Utilizando aprendizaje supervisado, se analizaron las 31 características de 223 muestras para entrenar un modelo capaz de determinar el índice de ansiedad, depresión e hipomanía en estas mujeres. En otro estudio con 30 participantes, se les pidió que mantuvieran un diario en su día a día con el chatbot, que solicitaba el registro de sus estados de ánimo y experiencias durante tres semanas.

TABLA III
TABLA DE LAS DIFERENTES MEDICIONES Y MÉTRICAS DE LOS ESTUDIOS REVISADOS.

Categoría	Mediciones/Métricas/Factores
Datos de autorrevelación	Datos de autorrevelación
Indicadores de estado mental	Indicadores de estado mental
Comportamiento de llevar un diario	Comportamiento de llevar un diario
Análisis del lenguaje	Análisis del uso de palabras
	Análisis de palabras
	Uso de palabras concretas
	Análisis del uso de palabras
	Preguntas de investigación e hipótesis
Medidas de inteligencia del chatbot	Inteligencia del chatbot
Procesamiento del lenguaje natural (NLP)	Procesamiento del lenguaje natural (NLP)
Terapia con chatbots	Terapia con chatbots
Empleo de algoritmos de aprendizaje profundo	Empleo de algoritmos de aprendizaje profundo
Papel crucial de sistemas inteligentes	Rol de sistemas inteligentes, especialmente los chatbots basados en inteligencia artificial (IA)
Desarrollo de un chatbot analítico	Desarrollo de un chatbot analítico
Beneficios asociados con chatbots basados en IA	Beneficios asociados con chatbots basados en IA

Estos participantes compartieron auto-revelaciones detalladas sobre sus experiencias y emociones personales con el chatbot.

Posteriormente, la información de auto-revelación de los participantes se dividió en tres grupos según los cambios en su estado mental antes y después del estudio: mejorado, deteriorado o sin cambios, y se realizaron comparaciones entre los datos de los tres grupos.

Además, el chatbot clasificó el texto en etiquetas emocionales como Felicidad, Alegría, Vergüenza, Enojo, Asco, Tristeza, Culpa y Miedo. Basándose en estas etiquetas emocionales, identificó el estado mental de los usuarios, determinando si estaban estresados o deprimidos mediante el análisis de los datos de chat. Para la detección de emociones, se implementaron tres clasificadores populares de aprendizaje profundo, a saber, Redes Neuronales Convolucionales (CNN), Redes Neuronales Recurrentes (CNN) y Redes de Atención Jerárquica (HAN).

TABLA IV
TABLA DE LOS TIPOS DE ESTUDIOS, EL TAMAÑO DE LA MUESTRA, EL TIPO DE POBLACIÓN, EL MÉTODO DE RECOPIACIÓN DE DATOS Y EL MÉTODO DE ANÁLISIS UTILIZADO EN CADA ESTUDIO.

Estudio	Tipo de estudio	Tamaño	Tipo de población	Método de recopilación de datos	Método de análisis
Supervised Machine Learning Chatbots for Perinatal Mental Healthcare	Aprendizaje Supervisado	223	Mujeres perinatales	Análisis de 31 características	Análisis de índices de ansiedad, depresión e hipomanía
Estudio de diario con chatbot	Estudio de Caso	30	Participantes que mantuvieron un diario personal a diario	Auto-revelaciones detalladas sobre estado de ánimo	Comparaciones entre grupos mejorados, deteriorados o sin cambios
Detección de emociones con chatbot	Estudio Experimental	No se especifica	Usuarios del chatbot	Clasificación de texto en etiquetas emocionales	Implementación de clasificadores de aprendizaje profundo (CNN, RNN, HAN)

Estudios y tamaño de las muestras

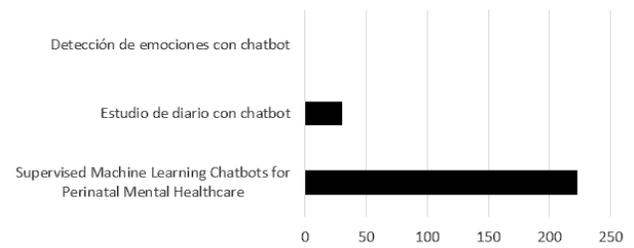


Gráfico 3. Estudios y su tamaño de muestras realizadas en los estudios.

¿Qué métodos/ técnicas fueron utilizados/ analizados/ discutidos en el estudio?

El estudio implementó avanzadas técnicas de aprendizaje automático, haciendo uso de redes neuronales recurrentes y mecanismos de atención para crear un chatbot de inteligencia artificial capaz de abordar preguntas relacionadas con COVID-19. Se llevó a cabo un análisis exhaustivo de las estrategias de atención implementadas, resaltando el rendimiento notable del mecanismo de atención de Punto en contextos específicos. Paralelamente, otro estudio abordó diversas metodologías, tales como el aprendizaje supervisado, análisis de características, y el desarrollo de un modelo de chatbot orientado a evaluar y monitorear la salud mental perinatal. En otra vertiente, se combinaron métodos de diseño de chatbot, integración de procesamiento de lenguaje natural y análisis de texto para evaluar de manera integral el impacto del chatbot en el estado mental de los participantes.

TABLA V
TABLA DE RESUMEN DE LOS MÉTODOS Y TÉCNICAS UTILIZADOS EN CADA ESTUDIO.

Estudio	Método/Técnicas utilizadas	Enfoque/Objetivo
Estudio sobre Chatbot de COVID-19	Redes neuronales recurrentes, mecanismos de atención	Crear un chatbot de IA para preguntas sobre COVID-19. Análisis de estrategias de atención, destacando el mecanismo de atención de punto.
Supervised Machine Learning Chatbots for Perinatal Mental Healthcare	Aprendizaje supervisado, análisis de características.	Desarrollar un chatbot para evaluar y monitorear la salud mental perinatal. Análisis de características y aprendizaje supervisado.
Estudio de diseño de chatbot para salud mental	Diseño de un chatbot, procesamiento de lenguaje natural, análisis de texto	Evaluar integralmente el impacto del chatbot en el estado mental de los participantes. Combinación de métodos de diseño, procesamiento de lenguaje natural y análisis de texto.

¿Qué mediciones de precisión/ rendimiento/ eficacia/ etc., fueron utilizadas?

El estudio introdujo el conjunto de datos CovidQA y evaluó diversas técnicas de pregunta-respuesta (QA) centradas en COVID-19. Para medir la efectividad de estos modelos, se aplicaron varias métricas, entre las que se incluyen Precision at rank one (P@1), Mean Reciprocal Rank (MRR) y Recall at rank three (R@3) [9]. La referencia [4], se implementaron diversas métricas para

evaluar el rendimiento del modelo de chatbot en respuesta a las preguntas de los usuarios. A continuación, se detallan las métricas utilizadas: BLEU Score: Esta métrica de rendimiento se empleó para evaluar la calidad de las respuestas generadas por la máquina en comparación con respuestas humanas. Evaluación Humana (V1 y V2): Dos voluntarios, identificados como V1 y V2, participaron en la calificación de las respuestas generadas por el modelo. Utilizaron una escala del 1 al 10 para realizar una verificación humana.

En otro estudio, se resalta la efectividad del modelo SVM y el chatbot en el análisis de la ansiedad y la depresión en mujeres perinatales. Se pone énfasis en la confiabilidad del modelo, destacando su mejora en la precisión con el tiempo. Además, se destaca la capacidad del chatbot para proporcionar apoyo emocional y consejos médicos personalizados.

TABLA VI
TABLA DE RESUMEN DE LAS MEDICIONES DE PRECISIÓN, RENDIMIENTO Y EFICACIA

Estudio	Mediciones de Precisión/Rendimiento/Eficacia	Justificación
Estudio de Tang et al. (2020)	-Precision at rank one (P@1) -Mean Reciprocal Rank (MRR) -Recall at rank three (R@3)	-Evalúa la efectividad de modelos de pregunta-respuesta centrados en COVID-19. -P@1 mide la precisión de la respuesta principal. MRR mide la calidad de las respuestas en el rango -R@3 evalúa la capacidad de recuperar respuestas correctas en los tres.
Estudio sobre Chatbot de COVID-19	-BLEU Score -Evaluación Humana (V1 y V2)	-BLEU Score mide la calidad de las respuestas generadas comparadas con respuestas humanas. -Evaluación humana (V1 y V2) implica la calificación por parte de voluntarios.
Estudio sobre SVM y Chatbot en Salud Mental	-Medición de la efectividad del modelo SVM -Capacidad del chatbot para proporcionar apoyo emocional y consejos médicos personalizados	-Destaca la confiabilidad del modelo SVM y su mejora en precisión con el tiempo. -Subraya la capacidad del chatbot para brindar apoyo emocional y asesoramiento médico personalizado en el contexto de la ansiedad y la depresión de las mujeres perinatales.

¿Cuál fue el valor de estas mediciones?

En relación con los resultados específicos del estudio anterior se presentan los siguientes hallazgos: CovidQA utilizando T5:

- Precision at rank one (P@1): 0.282
- Recall at rank three (R@3): 0.404
- Mean Reciprocal Rank (MRR): 0.415

CovidQA utilizando BERT:

-Precision at rank one (P@1): 0.234
En otro estudio, se detallan los valores asociados a las métricas utilizadas: BLEU Score:
- Valor: El BLEU Score varía de 0 a 1, donde 1 indica una coincidencia perfecta y 0 indica ninguna coincidencia. Evaluación Humana (V1 y V2):
-Valor: Se llevaron a cabo dos pruebas, una con preguntas exactas del conjunto de datos y otra con variaciones en las preguntas.

Además, en una investigación adicional que abordó la precisión, rendimiento y eficacia del modelo y del chatbot en el análisis de la ansiedad y la depresión en mujeres perinatales, se describen las siguientes métricas y sus valores asociados:

Precisión del Modelo para Diagnosticar Ansiedad:
-Valor: La confiabilidad del modelo para juzgar la ansiedad se informó como alta, alcanzando un 86%.
Precisión del Modelo para Diagnosticar Depresión:
- Valor: La confiabilidad del modelo para juzgar la depresión se informó como alta, alcanzando un 89%.
Exactitud de las Recomendaciones del Modelo:
- Valor Después de Tres Semanas: Tras cinco simulaciones, la precisión diagnóstica y la exactitud de las recomendaciones del modelo para la prueba alcanzaron el 93%.

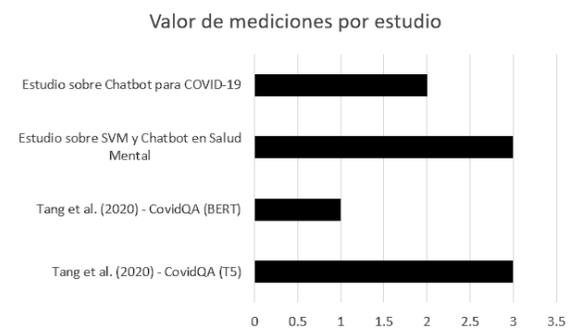


Gráfico 4. Valores de mediciones por estudio.

¿Se propusieron y desarrollaron soluciones/ se llegó a conclusiones? ¿Cuáles?

El estudio destaca la prometedora eficacia de la intervención digital terapéutica W-SUDs, evidenciando mejoras significativas y una favorable aceptación. Se identifican áreas de interés para investigaciones futuras, tales como la evaluación del impacto a largo plazo y la comparación con otras formas de tratamiento.

Asimismo, se concluye que los sistemas de respuesta a preguntas (QA) deberían procesar consultas utilizando tanto la coincidencia de palabras clave como el análisis semántico. El modelo implementa un mecanismo de atención (dot attention mechanism) para abordar preguntas relacionadas con COVID-19. Durante las pruebas con preguntas extraídas directamente de la base de datos, el modelo logró una

precisión del 87%. No obstante, al introducir variaciones naturales en las preguntas, la precisión disminuyó al 54%.

TABLA VII
TABLA DE RESUMEN DE LAS CONCLUSIONES Y
PROPUESTAS/DESARROLLOS FUTUROS

Estudio	Propuestas o Desarrollos Futuros	Conclusiones
Intervención digital terapéutica W-SUDs	Investigaciones futuras: Evaluación del impacto a largo plazo.	Eficacia prometedora de la intervención digital terapéutica W-SUDs.
	Comparación con otras formas de tratamiento.	Mejoras significativas y favorable aceptación. Áreas de interés para investigaciones futuras.
Estudio de sistemas de respuestas a preguntas (QA)	Implementación de mecanismo de atención (dot attention) para abordar preguntas relacionadas con COVID-19.	Recomendación de procesar consultas utilizando coincidencia de palabras clave y análisis semántico.
	Evaluación de precisión con variantes naturales en las preguntas.	Pruebas con preguntas de la base de datos: Precisión del 87%. Disminución de precisión al 54% al introducir variaciones.

¿Qué solución fue más exitosa?

W-SUDs evidenció mejoras significativas en diversas métricas relacionadas con el uso de sustancias y la salud mental. Paralelamente, un estudio señala que CO-search, un motor de búsqueda semántica basado en SBERT con la combinación de TF-IDF y BM25, se posiciona como la opción preferida en la actualidad. Esto sugiere que, en este contexto específico, la eficacia conjunta de SBERT junto con técnicas de coincidencia de palabras clave como TF-IDF y BM25 ha demostrado ser exitosa para el procesamiento de consultas en lenguaje natural.

TABLA VIII
TABLA DE LA EFICACIA DE LAS SOLUCIONES MENCIONADAS Y LAS TÉCNICAS ESPECÍFICAS

Solución	Resultados y Eficacia	Técnicas Utilizadas
W-SUDs	Mejoras significativas en métricas relacionadas.	Intervención digital terapéutica.
	Evidencia de mejoras en uso de sustancias y salud mental.	
CO-search	Posicionado como la opción preferida.	Motor de búsqueda semántica basado en SBERT
		Combinación de TF-IDF y BM25.
		Uso conjunto de SBERT con técnicas de coincidencia de palabras clave.

- Síntesis de estudios cualitativos:

¿El estudio apoya o discute algún modelo/ técnica/ solución? En caso afirmativo, ¿Qué modelo/ técnica/ solución y por qué?

El análisis resalta la necesidad de un enfoque regulatorio específico para los Modelos de Lenguaje a Gran Escala (LLMs) en contraste con las tecnologías médicas basadas en Inteligencia Artificial (IA) ya sujetas a regulación [10].
Microservicios para Escalabilidad: Se enfatiza la importancia de la escalabilidad en el desarrollo de chatbots destinados a pacientes crónicos.
HL7 FHIR para Intercambio de Datos: Este estándar se considera crucial para garantizar la interoperabilidad y la

integración efectiva con registros de salud electrónicos y personales.

AIML para Modelado de Conversación: La elección de AIML (Artificial Intelligence Markup Language) como estándar para el modelado de conversaciones se justifica por sus capacidades avanzadas de procesamiento de lenguaje natural. Esto permite la obtención de datos de salud valiosos a partir de las interacciones entre el usuario y el bot.

Adicionalmente, otro estudio aborda la implementación de un chatbot como solución tecnológica para superar las barreras en el acceso a la atención de salud mental. Este enfoque innovador en la introducción de un asistente virtual para proporcionar información y servicios en el ámbito de la salud mental destaca la aplicación práctica de la tecnología en el campo de la atención médica.

TABLA IX
TABLA COMPARATIVA DE LOS MODELOS, TÉCNICAS Y SOLUCIONES RESPALDADOS POR CADA ESTUDIO

Estudio	Método/Técnica/Solución	Justificación
Estudio de Regulación de LLMs	Enfoque regulatorio específico para modelos de lenguaje a gran escala (LLMs) en contraste con tecnologías médicas basadas en IA ya sujetas a regulación.	Reconoce la distinción clara entre LLMs y otras tecnologías médicas, destacando la necesidad de una categoría regulatoria específica.
Estudio de Roca S. et al. (2020)	-Microservicios para escalabilidad. -HL7 FHIR para intercambio de datos. -AIML para modelado de conversación.	-Escalabilidad es crucial en el desarrollo de chatbots para pacientes crónicos. -HL7 FHIR es esencial para garantizar interoperabilidad e integración con registros de salud. -AIML es elegido por sus capacidades avanzadas de procesamiento de lenguaje natural.

¿Hay alguna afirmación significativa hecha por el estudio? El estudio destaca de manera significativa la necesidad imperante de establecer una nueva categoría regulatoria específica para los Modelos de Lenguaje a Gran Escala (LLMs), reconociendo su distinción clara respecto a las tecnologías médicas basadas en IA que ya están sujetas a regulación. De manera paralela, otro análisis resalta la crucial importancia de la escalabilidad, la estandarización de datos y el modelado de conversación para garantizar el éxito de los chatbots en entornos de salud digital. Este enfoque proporciona una arquitectura adaptable a las cambiantes necesidades de los pacientes crónicos, proponiendo además que la automatización en la traducción de recursos FHIR a archivos AIML mejora significativamente la recopilación de información médica y personal. En este sentido, se plantea como otra afirmación de gran relevancia la hipótesis de que el chatbot MIRA será efectivo en la conexión de clientes con información y servicios de salud mental apropiados según sus necesidades individuales. Finalmente, en el estudio [16] se revela que en el uso de métricas tales como, BLEU y ROUGE son una desviación del objetivo final de la recomendación de diálogos, debido a que directamente no miden la capacidad de que el sistema entienda y responda adecuadamente a las preferencias del usuario.

TABLA X
TABLA DE RESUMEN DE LAS AFIRMACIONES SIGNIFICATIVAS
POR ESTUDIO

Estudio	Afirmación Significativa
Regulación de LLMs	Necesidad de una nueva categoría regulatoria para LLMs. Distinguirlos de las tecnologías médicas basadas en IA sujetas a regulación.
Análisis sobre chatbots en Salud Digital	Importancia de la escalabilidad estandarizada de datos y modelado de conversación para el éxito de los chatbots en salud digital. Propuesta de arquitectura adaptable a las necesidades de pacientes crónicos. Automatización en la traducción de recursos FHIR a archivos AIML mejora la recopilación de información médica y personal
Evaluación del chatbot MIRA	Hipótesis de que el chatbot MIRA será efectivo en conectar clientes con información y servicios de salud mental según sus necesidades individuales.
Survey on Conversational Recommendation Algorithms	El objetivo final de los sistemas de recomendación de diálogos (CRS) es obtener información sobre las preferencias del usuario durante el proceso de interacción, lo que implica que los enunciados generados deben contener palabras válidas sobre las preferencias del usuario.

IV. DISCUSIÓN

Los resultados de la revisión sistemática de literatura proporcionan una visión profunda sobre la integración de chatbots en la atención médica, con un enfoque específico en pacientes que utilizan servicios de salud en línea. La aplicación de la estrategia PICO ha proporcionado un marco estructurado para abordar la pregunta central de investigación: ¿Cómo afectan los chatbots a la atención médica?

Aplicaciones variadas de los chatbots en la atención médica:

El análisis exhaustivo de los estudios revisados revela una amplia variedad de aplicaciones para los chatbots en el ámbito de la atención médica. Desde el monitoreo del estado mental hasta la respuesta a preguntas relacionadas con COVID-19, estos agentes conversacionales están siendo empleados de manera versátil para abordar diversas necesidades de los pacientes y profesionales de la salud. Esta diversidad de aplicaciones subraya el potencial de los chatbots para mejorar la accesibilidad y eficiencia de los servicios de atención médica, al mismo tiempo que ofrecen soluciones innovadoras para desafíos específicos dentro del campo.

Utilización de técnicas avanzadas:

La adopción de técnicas avanzadas, como el procesamiento del lenguaje natural (NLP) y el aprendizaje profundo, emerge como un componente fundamental para potenciar la capacidad de los chatbots en la comprensión y respuesta a las necesidades de los usuarios. Estos enfoques tecnológicos permiten una interacción más natural y efectiva entre los usuarios y los chatbots, lo que resulta en una experiencia más satisfactoria y personalizada. Además, el empleo de estas técnicas avanzadas posibilita la adaptación de los chatbots a contextos específicos, como el monitoreo del estado mental o la respuesta a preguntas médicas urgentes relacionadas con la pandemia de COVID-19.

Importancia de la medición del rendimiento:

La evaluación del rendimiento de los chatbots emerge como un aspecto crucial para comprender su efectividad y eficacia en la práctica clínica. Los estudios revisados utilizan diversas

métricas, como el BLEU Score, Precision at rank one (P@1), Mean Reciprocal Rank (MRR) y evaluación humana, para medir diferentes aspectos del rendimiento de los chatbots. Estas métricas proporcionan información valiosa sobre la calidad de las respuestas generadas por los chatbots, su capacidad para comprender y satisfacer las necesidades de los usuarios, y su impacto en la experiencia general del usuario. Además, la evaluación del rendimiento permite identificar áreas de mejora y optimización para futuras implementaciones de chatbots en la atención médica.

V. CONCLUSIONES

En respuesta al objetivo fundamental de esta investigación, se ha explorado exhaustivamente el impacto de los chatbots en la atención médica, centrándose en su implementación en servicios de salud en línea. Los resultados de la revisión sistemática de literatura (RSL) revelan una amalgama de hallazgos notables que dan forma al panorama actual de los chatbots en el ámbito de la salud.

En primer lugar, la necesidad apremiante de una categoría regulatoria específica para los Modelos de Lenguaje a Gran Escala (LLMs) se destaca como un descubrimiento crucial. Esta distinción resalta la singularidad y complejidad de los LLMs en comparación con otras tecnologías médicas basadas en Inteligencia Artificial (IA), subrayando la importancia de una regulación adaptada a estas herramientas poderosas. Además, la atención especializada a la escalabilidad, la estandarización de datos y el modelado de conversación se identifican como elementos esenciales para el éxito de los chatbots, especialmente en entornos de salud digital.

La contribución de esta RSL a la literatura existente radica en su capacidad para sintetizar y contextualizar los avances recientes en el uso de chatbots en la atención médica. Al destacar las tendencias emergentes, como la necesidad regulatoria específica y la importancia de la escalabilidad, la RSL proporciona una guía valiosa para investigadores y profesionales de la salud. Sin embargo, es fundamental reconocer las limitaciones inherentes a esta revisión. La rápida evolución del campo de la inteligencia artificial puede haber dejado fuera desarrollos más recientes, y la variabilidad en los enfoques metodológicos de los estudios revisados puede introducir sesgos potenciales.

Como vía para futuras investigaciones, se recomienda profundizar en la efectividad y aceptación de los chatbots en poblaciones específicas, como pacientes crónicos o aquellos con necesidades particulares de salud mental. Además, explorar en mayor medida la interacción entre chatbots y sistemas de salud existentes, como el uso de HL7 FHIR para el intercambio de datos, puede proporcionar conocimientos más detallados sobre la integración práctica de estas

tecnologías en entornos clínicos. Para concluir, se recomienda la aplicación o uso de técnicas avanzadas como Machine Learning, PLN, a su vez tener en cuenta modelos de aprendizaje profundo con arquitecturas escalables y adaptables; sin olvidar la seguridad y privacidad de la data de los usuarios en todo momento. Ya que, estos puntos trabajando de la mano con los hallazgos de esta revisión sistemática, proporcionarían una guía hacia el desarrollo e implementación efectiva de los futuros chatbots específicamente en el área de la atención médica.

VI. REFERENCIAS

[1] F. N. Patel, R. Thakore, I. Nandwani, and S. Bharti, "Combating Depression in Students using an Intelligent ChatBot: A Cognitive Behavioral Therapy", 2019 IEEE 16th India Council International Conference (INDICON), pp. 1–4, Dec. 01, 2019. Available: <https://doi.org/10.1109/INDICON47234.2019.9030346>

[2] T. Nadarzynski, O. Miles, A. Cowie, and D. Ridge, "Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study", *Digital Health*, vol. 5, Jan. 01, 2019. Available: <https://doi.org/10.1177/2055207619871808>

[3] X. Wang, R. Luo, Y. Liu, P. Chen, Y. Tao, and Y. He, "Revealing the complexity of users' intention to adopt healthcare chatbots: A mixed-method analysis of antecedent condition configurations", *Inf. Process. Manag.*, vol. 60, p. 103444, Sep. 01, 2023. Available: <https://doi.org/10.1016/j.ipm.2023.103444>

[4] W. X. Hui, N. Aneja, S. Aneja, and A. G. Naim, "Conversational chat system using attention mechanism for COVID-19 inquiries", *International Journal of Intelligent Networks*, vol. 4, pp. 140–144, Jun. 17, 2023. Available: <https://doi.org/10.1016/j.ijin.2023.05.003>

[5] R.-min Han, A. Todd, S. Wardak, S. Partridge, and R. Raeside, "Feasibility and Acceptability of Chatbots for Nutrition and Physical Activity Health Promotion Among Adolescents: Systematic Scoping Review With Adolescent Consultation", *JMIR Human Factors*, vol. 10, Oct. 04, 2022. Available: <https://doi.org/10.2196/43227>

[6] S. Iyer, D. Shetty, P. Badgujar, A. Nambiar, and K. Jewani, "A Proposal for Virtual Mental Health Assistant", 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), vol. 1, pp. 953–957, Mar. 19, 2021. Available: <https://doi.org/10.1109/ICACCS51430.2021.9441990>

[7] B. Meskó and E. Topol, "The imperative for regulatory oversight of large language models (or generative AI) in healthcare", *NPJ Digital Medicine*, vol. 6, Jul. 06, 2023. Available: <https://doi.org/10.1038/s41746-023-00873-0>

[8] M. Kawasaki, N. Yamashita, Y.-C. Lee, and K. Nohara, "Assessing Users' Mental Status from their Journaling Behavior through Chatbots", *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, Oct. 19, 2020. Available: <https://doi.org/10.1145/3383652.3423870>

[9] R. Wang, J. Wang, Y. Liao, and J. Wang, "Supervised Machine Learning Chatbots for Perinatal Mental Healthcare", 2020 International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI), pp. 378–383, Dec. 01, 2020. Available: <https://doi.org/10.1109/ICHCI51889.2020.00086>

[10] S. Roca, J. Sancho, J. García, and Álvaro A. Iglesias, "Microservice chatbot architecture for chronic patient support", *Journal of biomedical informatics*, p. 103305, Feb. 01, 2020. Available: <https://doi.org/10.1016/j.jbi.2019.103305>

[11] C. de Cock, M. Milne-Ives, M. V. van Velthoven, A. Alturkistani, C. Lam, and E. Meinert, "Effectiveness of Conversational Agents (Virtual Assistants) in Health Care: Protocol for a Systematic Review", *JMIR Research Protocols*, vol. 9, Nov. 06, 2019. Available: <https://doi.org/10.2196/preprints.16934>

[12] L. Ni, C. Lu, N. Liu, and J. Liu, "MANDY: Towards a Smart Primary Care Chatbot Application". pp. 38–52, Nov. 17, 2017. Available: https://doi.org/10.1007/978-981-10-6989-5_4

[13] A. Perdana and I. A. Mokhtar, "Leveraging digital technologies for information technology-enabled healthcare transformation at SingHealth", *Journal of Information Technology Teaching Cases*, vol. 13, pp. 97–103, May 18, 2022. Available: <https://doi.org/10.1177/20438869221091582>

[14] B. Ali and V. Ravi, "Developing Dialog Manager in Chatbots via Hybrid Deep Learning Architectures". pp. 301–310. Available: https://doi.org/10.1007/978-981-15-5679-1_28

[15] T. Lee et al., "Intelligent Career Advisers in Your Pocket? A Need Assessment Study of Chatbots for Student Career Advising". Available: <https://www.semanticscholar.org/paper/cc563fc914607c1b2249593420f3f6a7270032b0>

[16] Zhao, M.-Y., Huang, X.-W., Sang, J.-T., Yu, J. "Survey on Conversational Recommendation Algorithms". Available: <https://www.jos.org.cn/josen/article/pdf/6521>

[17] Shedhi B., Shetty V., Chadaga R., Bhat R., Preethi B., Kinki K. "Implementation of Chatbot that Predicts an Illness Dynamically using Machine Learning Techniques" Available: <https://doi.org/10.5829/ije.2024.37.02b.08>