UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Tecnura

https://revistas.udistrital.edu.co/ojs/index.php/Tecnura/issue/view/1274

DOI: https://doi.org/10.14483/22487638.22808

REVISIÓN

Explorando el uso de inteligencia artificial generativa para el desarrollo de chatbots para portales web universitarios: un mapeo sistemático

Exploring the use of generative artificial intelligence for the development of chatbots for university web portals: A systematic mapping

Arnold Steeven Catamuscay Pérez [1], Cristian Eduardo Núñez Valencia [2] y Hugo Armando Ordoñez Erazo [3]

Fecha de Recepción: 24 de octubre de 2024 **Fecha de Aceptación**: 15 de marzo de 2025

Cómo citar: Catamuscay Pérez A. S., Núñez Valencia C. E., Ordoñez Erazo H. A. Explorando el uso de inteligencia artificial generativa para el desarrollo de chatbots para portales web universitarios: un mapeo sistemático, *Tecnura*, 29(83), 144-183. https://doi.org/10.14483/22487638.22808

Resumen

Contexto: los chatbots con inteligencia artificial generativa (GAI, por su sigla en inglés) han evolucionado significativamente, impulsados por avances sobre grandes modelos de lenguaje (LLM, por su sigla en inglés). Estos sistemas ofrecen interacciones más naturales y adaptativas, a la vez que transforman diversos sectores y plantean nuevos desafíos tecnológicos y éticos. **Objetivo:** identificar las principales tendencias, oportunidades y desafíos en el desarrollo de chatbots con GAI en los últimos años.

Metodología: se realizó un mapeo sistemático adaptado, por medio del cual se analizó el uso de GAI en chatbots. Se definieron tres preguntas de investigación y se hizo una búsqueda exhaustiva en las bases Web of Science, Scopus y ScienceDirect. Los estudios fueron clasificados para responder a las preguntas de investigación.

Resultados: los sectores de educación y salud son los más investigados, en los que se destaca el uso de LLM como GPT-4 (*generative pre-trained transformer*), para personalización del aprendizaje y apoyo en salud mental, por ejemplo. También se identificaron aplicaciones en tecnología, comercio e industria. Los modelos de OpenAI son los predominantes, aunque existen alternativas especializadas. Los principales desafíos incluyen .ªlucinaciones", necesidad de supervisión humana, sesgos y altos costos computacionales.

Conclusiones: la flexibilidad y rendimiento de modelos como GPT-4 los posicionan como opciones prominentes

[1]Estudiante de la Facultad de Ingeniería y Telecomunicaciones de la Universidad del Cauca. Correo electrónico: ascatamuscay@unicauca.edu.co

[2]Estudiante de la Facultad de Ingeniería y Telecomunicaciones de la Universidad del Cauca. Correo electrónico: cnunez@unicauca.edu.co

[3]Docente de la Facultad de Ingeniería y Telecomunicaciones de la Universidad del Cauca. Correo electrónico: hugoordonez@unicauca.edu.co

para implementaciones de chatbots. Los desafíos identificados son cruciales para guiar un desarrollo efectivo, para así considerar oportunidades y limitaciones actuales.

Palabras clave: inteligencia artificial generativa, chatbots, mapeo sistemático.

Abstract

Context: Generative artificial intelligence (GAI) chatbots have evolved significantly, driven by advances in large language models (LLM). These systems offer more natural and adaptive interactions, transforming various industries and posing new technological and ethical challenges.
Objective: Identify the main trends, opportunities and challenges in the development of chatbots with GAI in recent years.
Methodology: An adapted systematic mapping was conducted, analyzing the use of GAI in chatbots. Three research questions were defined and an exhaustive search was carried out in Web of Science, Scopus, and ScienceDirect databases. The studies were classified to answer the research questions.
Results: The education and health sectors are the most researched, highlighting the use of LLM such as GPT-4 for learning personalization and mental health support. Applications in technology, commerce, and industry were also identified. OpenAI models are dominant, although specialized alternatives exist. The main challenges include "hallucinations,"the need for human supervision, biases, and high computational costs.
Conclusions: The flexibility and performance of 144-183models like GPT-4 position them as prominent options for chatbot implementations. The identified challenges are crucial for guiding effective development, considering current opportunities and limitations.

Keywords: Generative artificial intelligence, Chatbots, Systematic map.

Introducción

La inteligencia artificial generativa (GAI, por su sigla en inglés) es hoy una tecnología utilizada en aplicaciones significativas para diversos campos, entre los cuales se encuentra el educativo (Ilagan y Ilagan, 2024); (Javaid, Haleem, Singh, *et al.*, 2023). Las capacidades de esta tecnología GAI permiten, por ejemplo, mantener conversaciones coherentes y contextuales, responder a preguntas de forma específica y generar texto relacionado a una entrada, o *input*, por lo cual, han surgido nuevas formas para mejorar la interacción en los portales educativos. Aunque obtener información sobre aspectos generales, como los requisitos de un curso, los procesos de inscripción, la obtención de expedientes académicos, la orientación profesional o los horarios, no es directamente esencial para el aprendizaje, sí influye en la calidad de la experiencia universitaria (Seeman y O'Hara, 2006). Además, cuando los comportamientos, el entorno y los procesos institucionales son eficientes, es probable que se refleje una mejora en los resultados académicos de los estudiantes (Prebble *et al.*, 2004).

Por tanto, la implementación de chatbots para portales web universitarios supone un avance significativo para el entorno de la Educación Superior. Un chatbot basado en la tecnología GAI ofrecería una herramienta de soporte a estudiantes, profesores y personal administrativo; mejoraría la accesibilidad de la información educativa de una universidad y optimizaría la gestión de los recursos educativos. En la actualidad, cuando la tecnología evoluciona rápidamente, las instituciones educativas enfrentan el desafío de mantenerse al día con las innovaciones que pueden mejorar significativamente el apoyo educativo. Sin embargo, los grandes modelos de lenguaje (LLM, por su sigla en inglés), con los que funcionan chatbots como ChatGPT, son reservados con su funcionamiento interno, lo cual dificulta comprender el proceso por el cual estos modelos llegan a sus conclusiones de una forma más detallada (Meskó y Topol, 2023); (Wölfel *et al.*, 2023). Además, estos chatbots requieren procesar datos personales, lo que representaría riesgos en cuanto a la privacidad y seguridad de una institución, ya que los datos serían susceptibles a ataques malintencionados (Escalante *et al.*, 2023).

Un mapeo sistemático de la literatura es una herramienta fundamental para identificar, analizar y clasificar el conocimiento existente sobre un área de estudio, lo cual facilita reconocer tendencias, tecnologías emergentes y vacíos en la investigación. En el campo de la GAI aplicada a los chatbots, este enfoque metodológico resulta pertinente para reconocer las mejores prácticas y las lecciones aprendidas de implementaciones previas. De este modo, el presente estudio proporciona un fundamento para la toma de decisiones estratégicas en la implementación del chatbot para el portal web de la Universidad del Cauca.

El resto del documento se estructura de la siguiente manera: primero se exponen los antecedentes relacionados con el estudio. Luego, se explica la metodología de empleada. Posteriormente, se exponen los resultados del mapeo sistemático y se discuten dichos resultados. Por último, se ofrecen las conclusiones del estudio.

Antecedentes

En los últimos años, el desarrollo y aplicación de chatbots impulsados por GAI han experimentado un crecimiento notable, lo cual supone un contexto más amplio de la evolución de la inteligencia artificial. La historia de la GAI puede dividirse en varias etapas, desde los primeros sistemas expertos en la segunda mitad del siglo XX, hasta la llegada de las redes neuronales profundas y el procesamiento en la nube; una evolución que ha llevado a avances significativos, particularmente con la introducción de grandes modelos de lenguaje y la computación en la nube (Wang *et al.*, 2023).

Los orígenes de los chatbots generativos se remontan a la década del cincuenta, con la propuesta del test de Turing, como punto de partida para evaluar la capacidad de las máquinas para simular la inteligencia del ser humano. A lo largo de las siguientes décadas surgieron sistemas como ELIZA, en los años 60, que despertaron el interés por el procesamiento de lenguaje natural, más tarde en los años 1970 con PARRY y en los 90 con ALICE, este último aprovechó el lenguaje marcado AIML para mejorar su capacidad de respuesta. En la década de 2010, los asistentes virtuales como Siri, Google Assistant y Alexa llevaron los chatbots al ámbito comercial; así se consolidó su uso en diversas aplicaciones. Estos asistentes inicialmente se basaban en modelos de recuperación de información que seleccionaban respuestas predefinidas a partir de bases de datos, pero, desde 2019, los modelos generativos (*generative pre-trained transformer*), como GPT-2, GPT-3 y GPT-4 desarrollados por OpenAI, comenzaron a dominar el campo, destacados por su capacidad para generar respuestas más naturales y adaptativas, con lo cual se mejoró la calidad de la interacción con los usuarios (Wölfel *et al.*, 2023); (Prasad *et al.*, 2024).

En este sentido, los transformadores han marcado un antes y un después en el desarrollo de modelos de lenguaje generativos y en la IA en general, puesto que, antes de su aparición se utilizaban arquitecturas como las redes neuronales recurrentes (RNN, por su sigla en inglés) y las memorias a largo y corto plazo (LSTM, por su sigla en inglés) para tareas de procesamiento de lenguaje natural (NLP, por su sigla en inglés), pero estas presentaban limitaciones al manejar dependencias a largo plazo y secuencias largas (Wang *et al.*, 2023; Bengesi *et al.*, 2024). Los transformadores, introducidos por Vaswani *et al.* (2023), lograron superar estos problemas a través del mecanismo de *self-attention*, o atención propia, que le permite al modelo prestar atención a las partes más relevantes de una secuencia, sin importar su longitud (Bengesi *et al.*, 2024). La arquitectura de un transformador usualmente se basa en dos componentes principales: el codificador y el decodificador. El primero procesa la información de manera bidireccional, incluso el contexto completo del texto de entrada; mientras que el decodificador genera las respuestas palabra por palabra, en un proceso conocido como *decodificación autorregresiva*. Lo anterior ha permitido que los modelos de lenguaje generativos sean más eficientes y precisos en tareas como la traducción automática, la generación de texto y la respuesta a preguntas (Wang *et al.*, 2023).

La aparición de los grandes modelos de LLM ha sido fundamental para el desarrollo de chatbots generativos, y para establecer interacciones más fluidas y humanas. Estos modelos se entrenan con grandes cantidades de datos no estructurados, lo que les permite generar texto coherente y comprender tareas lingüísticas complejas. Hay diversas variantes de LLM, incluidas las arquitecturas de solo codificador, como BERT, que capturan información contextual bidireccionalmente, y las de solo decodificador, como la familia GPT, que sobresalen en la generación de secuencias coherentes y contextualmente relevantes (Cascella *et al.*, 2024). Además,

uno de los avances más notables en la aplicación de LLM a los chatbots es el ajuste fino (*fine-tuning*) de los modelos, lo que mejora su capacidad para manejar la conversación y generar respuestas apropiadas (Wölfel *et al.*, 2023; Cascella *et al.*, 2024); esta técnica, junto con el aprendizaje por refuerzo basado en retroalimentación humana (RLHF, por su sigla en inglés) ha sido clave para el éxito de modelos como ChatGPT (Cascella *et al.*, 2024). Tales mejoras permiten que los chatbots comprendan las consultas de los usuarios, mantengan y proporcionen respuestas más precisas y personalizadas.

Debido a la necesidad de grandes recursos computacionales para su entrenamiento y despliegue, la mayoría de estos modelos se entrenan y ejecutan en servidores en la nube, lo que permite a los usuarios enviar solicitudes y recibir contenido generado sin tener que procesar los modelos localmente (Wang *et al.*, 2023).

Metodología

El objetivo de este estudio es identificar las principales tendencias, oportunidades y desafíos en la creación de chatbots con GAI en los últimos años, de manera que se pueda obtener una idea de cuáles son las aplicaciones que se han diseñado recientemente, y cómo esto sería de utilidad para la implementación e integración de chatbots personalizados para portales universitarios.

Este estudio se fundamenta en una metodología de proceso de mapeo sistemático descrita en la conferencia titulada *Systematic mapping studies in software engineering* (Petersen *et al.*, 2008). Teniendo en cuenta dicho enfoque, se ha logrado realizar, estructurar y llevar a cabo el mapeo sistemático de una manera eficiente y rigurosa, adaptándolo al área de investigación sobre el uso de la GAI en la construcción de chatbots. A continuación, se detallan los pasos y los procesos seguidos:

Definición de preguntas de investigación

Esta etapa inicial implica la formulación de preguntas de investigación que guiarán el alcance y ayudarán a cumplir con el objetivo planteado. Se busca, entonces, ofrecer una visión general sobre el uso de la tecnología de GAI en chatbots, que abarque aspectos como la cantidad y tipos de investigaciones, y las tendencias de publicación a lo largo del tiempo.

En ese sentido, para recopilar la información relevante correspondiente al mapeo sistemático, se plantearon las siguientes preguntas de investigación:

- *RQ1. ¿Cuáles son los temas más investigados en el desarrollo y aplicación de chatbots elaborados con GAI?* Para identificar las áreas de enfoque predominantes en la literatura científica y determinar cómo se ha estudiado el uso de GAI a lo largo de la elaboración de chatbots, es esencial comprender los temas más investigados. Además, este conocimiento puede orientar implementaciones de chatbots para portales universitarios en el contexto nacional, lo que garantiza que se integren las prácticas y enfoques más actuales y pertinentes. Esto situará investigaciones en el contexto actual.

- *RQ2. ¿Cuáles son los principales LLM utilizados en la producción y aplicación de chatbots con inteligencia artificial generativa?* El núcleo de los chatbots basados en GAI son los LLM. Para elegir las tecnologías más avanzadas y efectivas para un contexto universitario, es fundamental identificar los modelos más empleados. Con el propósito de facilitar la toma de decisiones sobre la selección del LLM más adecuado para las necesidades particulares de portales web universitarios, esta información también puede brindar una comprensión de las habilidades y limitaciones de los modelos actuales.

- *RQ3. ¿Cuáles son los principales desafíos y limitaciones identificados en aplicaciones de chatbots realizadas con inteligencia artificial generativa?* Reconocer los desafíos y limitaciones es importante para anticipar problemas potenciales en implementaciones de chatbots con GAI. Al entender las dificultades que otros estudios han presentado, es posible diseñar estrategias para mitigarlas y mejorar las opciones de éxito en la implementación de este tipo de chatbots.

Búsqueda

Se empleó una cadena de búsqueda estructurada según los criterios de población, intervención, comparación y resultado (PICO, por su sigla en inglés) recomendados por Kitchenham y Charters (2007) (tabla 1). Se utilizaron diferentes bases de datos científicas, en las cuales se realizaron búsquedas exhaustivas para reconocer estudios primarios relevantes para el mapeo sistemático. Con esta metodología se buscó una cobertura amplia y no restrictiva del campo de estudio, sin sesgos y con garantía de una representación más amplia del uso de la tecnología GAI en chatbots.

Las palabras clave escogidas son: chatbot, *generative artificial intelligence, generative model.* Además, con el fin de cubrir la mayor cantidad de literatura relacionada, se agregaron sinónimos asociados a esas palabras clave:

- Para limitar el alcance a aplicaciones de chatbots, naturalmente son usadas las palabras chatbot, chatbots.

Tabla 1. Criterios PICO y elección realizada para plantear la cadena de búsqueda

Criterio	Elección
Población (rol de ingeniería de software, tipo de ingeniero, área de aplicación o un grupo de la industria).	Artículos sobre chatbots que emplean inteligencia artificial generativa (GAI).
Intervención (metodología, herramienta, tecnología, procedimiento que aborda una cuestión en concreto).	Inteligencia artificial generativa (GAI).
Comparación (metodología, herramienta, tecnología, procedimiento contrastado con la intervención).	No aplicable (dado que no se está comparando diferentes intervenciones).
Resultados.	Tendencias en investigación, principales modelos de lenguaje utilizados, desafíos y limitaciones identificados.

- La tecnología central del presente estudio es la inteligencia artificial generativa (GAI), así que el otro conjunto de sinónimos que complementan la cadena de búsqueda es: *generative AI*, *generative artificial intelligence* (GAI), *generative model*.
- Y con el objetivo de tener un enfoque en los avances y desarrollos recientes, se utilizan también *innovation*, *advance*, *development*.

En la tabla 2 se presenta la cadena de búsqueda resultante, con la cual se busca maximizar la cobertura de la literatura relevante.

Tabla 2. Cadena de búsqueda empleada

Cadena de búsqueda
(“chatbot” OR “chatbots”) AND ("generative AI” OR ”generative artificial intelligence” OR ”GAI” OR ”generative model”) AND (innovation OR advance OR development)

Las fuentes o recursos usados para la búsqueda fueron las bases de datos científicas Web of Science, Scopus y ScienceDirect, accesibles mediante la división de bibliotecas de la Universidad del Cauca. En la tabla 3 puede observarse el número de artículos encontrados en cada base de datos en la búsqueda inicial.

Selección de artículos para inclusión y exclusión

Con el propósito de filtrar los estudios que no sean adecuados para el mapeo, se definieron los criterios específicos de inclusión y exclusión (tabla 4). Estos se fundamentaron en las preguntas de investigación, para asegurar que solo se incluyan artículos que aborden de manera significativa el tema central del estudio (Petersen *et al.*, 2008).

Tabla 3. Cantidad de artículos encontrados en cada base de datos

Base de datos	Cantidad de resultados
Web of Science	111
Scopus	181
ScienceDirect	1005

Tabla 4. Criterios para la inclusión o exclusión

Criterio	Incluidos	Excluidos
Temática	Artículos que traten sobre chatbots desarrollados con tecnologías de inteligencia artificial generativa.	Artículos que no abordan la aplicación de inteligencia artificial generativa (GAI) en chatbots. Artículos que traten sobre chatbots sin mencionar el uso de tecnologías de inteligencia artificial generativa (GAI).
Periodo de tiempo	Artículos publicados en los últimos cinco años para asegurar la relevancia y actualidad de la información.	Artículos con más de cinco años.
Idioma	Artículos en inglés	Artículos en un idioma distinto al inglés.
Tipo de acceso	Artículos con acceso al texto completo para facilitar una evaluación detallada.	Artículos que requieren una suscripción de paga.
Duplicación de artículos	Artículos únicos y no repetidos.	Artículos que repiten el mismo contenido.

Adicionalmente, con el objetivo de evaluar la calidad de los estudios primarios, se formularon las siguientes preguntas:

- ¿Está claramente definido el objetivo del estudio, siendo relevante para la temática del mapeo?
- ¿Los resultados y discusiones del estudio aportan información útil para la implementación futura de un chatbot en el contexto del portal web de una universidad?
- ¿El estudio es relevante para responder a las preguntas de investigación del mapeo sistemático?

- ¿El estudio aborda la implementación práctica de la tecnología o proporciona ejemplos concretos de uso?

Esquema de clasificación

Una vez seleccionados los estudios a incluir en el mapeo, se busca entonces desarrollar un esquema de clasificación coherente, por lo que, mediante la lectura de los resúmenes de los artículos seleccionados, se identifican palabras clave y conceptos que reflejan el área de contribución de cada artículo, y determinan el contexto de la investigación. Posteriormente, estas palabras clave se combinan para generar un conjunto de categorías representativas del campo de estudio. Por otra parte, en caso de que los resúmenes no permitan la extracción de palabras clave, también se analizan las secciones "Introducción" o "Conclusión", con el fin de obtener palabras clave significativas. Este proceso sistemático busca que el esquema de clasificación sea exhaustivo y representativo.

Extracción de datos y mapeo de estudios

Todos los artículos relevantes se clasificaron dentro de un esquema. Además, los datos se documentaron en una tabla de Excel con el objetivo de facilitar el registro de cada categoría para la clasificación de cada artículo. Finalmente, se calcularon las frecuencias de los artículos en cada categoría; así, se obtuvo un análisis detallado de las áreas más y menos investigadas, los LLM que más se mencionan, y se logró el reconocimiento de posibles desafíos a tratar en la implementación de un chatbot con GAI.

Resultados

Luego de realizar la búsqueda de las diferentes bases de datos, la cantidad de estudios que se tuvieron en cuenta para el mapeo son: 14 de Scopus, 20 de ScienceDirect y 11 de Web of Science, para un total de 45 que se sometieron al mapeo sistemático (figura **1**).

A pesar de tener un rango de consulta de cinco años, los estudios seleccionados fueron publicados entre 2023 y la actualidad (septiembre de 2024), debido, tal vez, a que el auge de los chatbots con GAI se dio a finales del año 2022 con ChatGPT, y por tanto, los publicados después de 2023 serían más útiles. Según la figura **2**, en 2023 hubo un total de 19 publicaciones relevantes para este mapeo, y hasta septiembre de 2024 se ubicaron otros 26 artículos para añadir al mapeo.

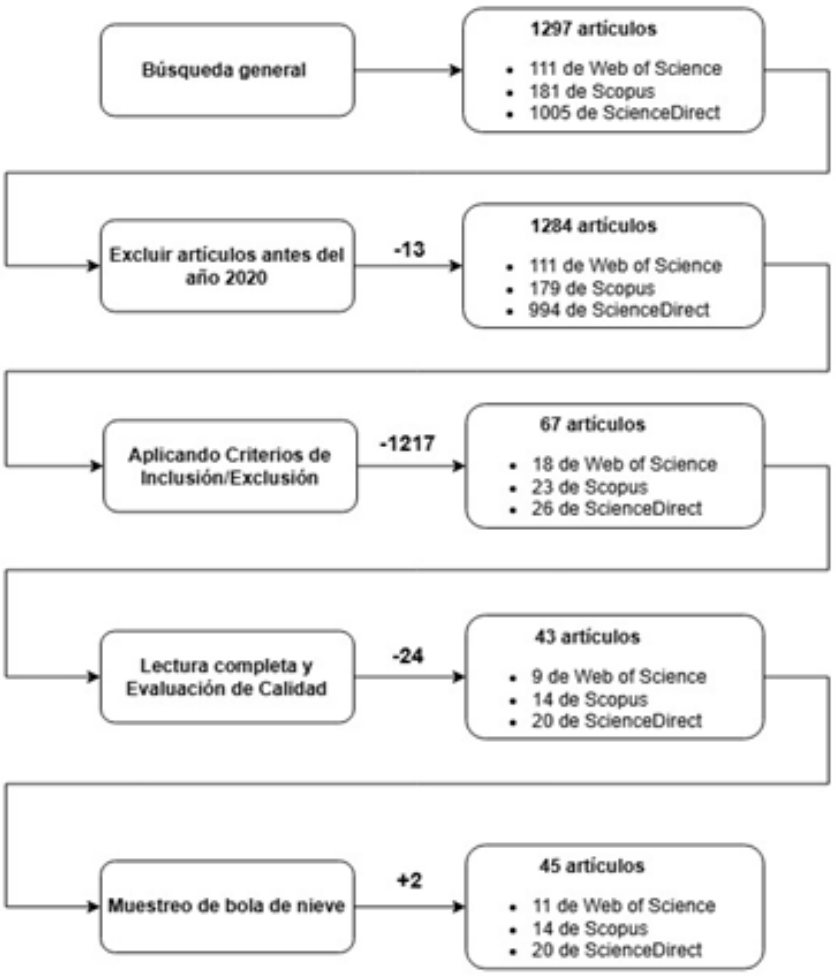

Figura 1. Pasos de la selección de artículos

Temas más investigados (RQ1)

Las principales áreas de investigación abarcan los campos de educación, salud, industria, comercio, gobierno y tecnología (figura **3**), con un claro dominio por las áreas de educación y salud, con 13 y 16 artículos respectivamente; seguidos por 7 estudios relacionados con tecnología, y entre las áreas con menos artículos están: comercio con 4, industria con 3 y gobierno con 2 artículos. Las descripciones de estas áreas se detallan en la tabla 5.

Acerca del área de educación, se encontraron numerosos estudios que discuten la aplicación de chatbots con GAI de una forma más amplia (figuras **4** y **5**). Por ejemplo el artículo de Wölfel *et al.* (2023) se destaca por comparar chatbots educativos basados en conocimiento frente a aquellos que usan de GAI, estos utilizan materiales didácticos como fuente y se percibe que los chatbots basados en conocimiento facilitan un mayor control sobre las respuestas que brindan, aun así carecen de la flexibilidad de los chatbots generativos, como ChatGPT, ya que

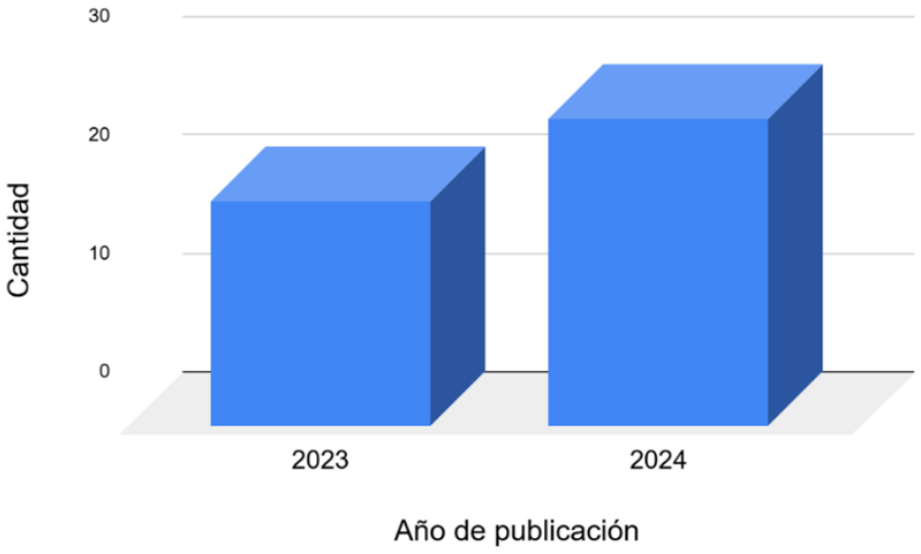

Figura 2. Distribución de publicaciones por año

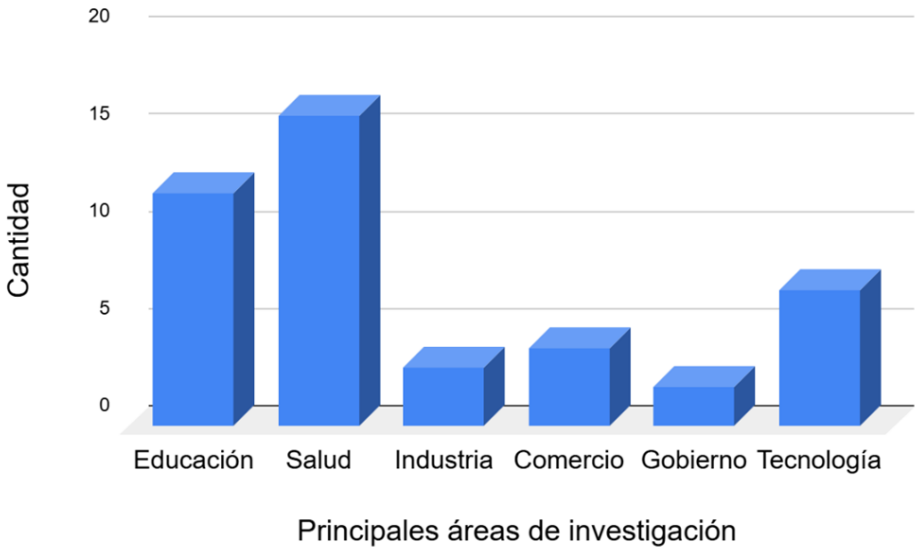

Figura 3. Principales áreas de investigación

estos sí pueden ofrecer una mejor interacción educativa y presentar nuevas posibilidades en el ámbito pedagógico.

En otro trabajo (Drelick *et al.*, 2024), se propone un enfoque innovador que combina simulaciones clínicas con la fabricación de chatbots educativos, se ofrece una guía estructurada en cuatro fases: conceptualización, diseño de protocolos, diseño técnico, y pruebas y revisiones;

Tabla 5. Áreas de investigación definidas

Área	Descripción
Educación	Trabajos dirigidos a mejorar o transformar procesos educativos mediante tecnologías innovadoras como chatbots educativos o sistemas de tutoría digital apoyados por IA generativa.
Salud	Investigaciones enfocadas en aplicaciones de chatbots con inteligencia artificial generativa en el ámbito de la salud, como sistemas de diagnóstico asistido por GAI, medicina personalizada, o intervenciones para el cambio de comportamiento.
Industria	Trabajos que exploran aplicaciones de chatbots generativos en contextos industriales, como optimización de procesos, control de calidad, o sistemas de producción automatizados.
Comercio	Investigaciones que estudian el impacto de chatbots generativos en el comercio, como sistemas de recomendación personalizados o chatbots para servicio al cliente en e-commerce.
Gobierno	Trabajos que examinan desde la perspectiva de gobierno, el uso de la IA generativa, en conjunto con chatbots, incluyendo temas como la gestión de datos públicos, sistemas de seguridad digital, o análisis de políticas públicas basados en datos.
Tecnología	Innovaciones en procesamiento del lenguaje natural y grandes modelos de lenguaje (LLM), internet de las cosas (IoT) y análisis forense digital. Avances importantes para la mejora de chatbots generativos.

este enfoque menciona la importancia de integrar prácticas clínicas en la educación para mejorar la experiencia de aprendizaje. También se presentó un trabajo en el sector universitario que detalla la realización de un prototipo de agente virtual de apoyo universitario basado en un modelo de lenguaje generativo (Ilagan y Ilagan, 2024); el agente conversacional está diseñado para responder preguntas sobre políticas y normativas universitarias, y así mejorar la experiencia de servicio para estudiantes, profesores y el personal administrativo.

Por otro lado, en Zhu *et al.* (2024) se explora cómo ChatGPT está siendo utilizado con el fin de generar arte, asistir el proceso de escritura creativa y facilitar la colaboración entre artistas. Del mismo modo, se trabaja en Javaid, Haleem, Singh *et al.* (2023), quienes comentan las diversas aplicaciones de ChatGPT en la educación, y su implicación en la personalización del aprendizaje y la automatización de tareas como la calificación de exámenes y asignaciones.

Además de las áreas mencionadas, también hay documentos que se enfocan en sectores particulares, los cuales contemplan diversas temáticas. Popovici (2023), por ejemplo, se enfoca en la educación en informática o ciencias de la computación, y analiza el uso de ChatGPT en un curso de programación funcional de la Universidad Politécnica de Bucarest; así, evidencia que ChatGPT puede ser una herramienta efectiva para la revisión de código, lo que sugiere su integración en la mejora de las habilidades de programación de los estudiantes. En relación con

Figura 4. Resumen de casos prácticos de chatbots con GAI en educación

el aprendizaje del idioma inglés, Escalante *et al.* (2023) evaluaron la retroalimentación generada por ChatGPT en comparación con la de tutores humanos, y no mostraron diferencias significativas en los resultados de aprendizaje; además observaron una preferencia dividida entre la retroalimentación generada por IA y la humana, lo que sugiere un enfoque híbrido que combina ambas formas de retroalimentación.

La educación en química orgánica también se abordó en Yik y Dood (2024), donde ChatGPT se evaluó como una herramienta para explicar mecanismos de reacción, y se resaltó el *prompt engineering* para mejorar la calidad de las explicaciones generadas. Para la enseñanza de geometría, un artículo evaluó el desempeño de varios chatbots en la resolución de problemas geométricos, pero con la particularidad de que era en español; destacó, también, que los modelos de lenguaje actuales tienen dificultades para manejar conceptos geométricos complejos, lo que resalta la necesidad de enfoques metodológicos más informados para integrar estas tecnologías en la educación (Parra *et al.*, 2024). A propósito, un chatbot personalizado, basado en un modelo de GPT, fue evaluado en el área de diseño instruccional, con el objetivo de crear materiales educativos para la alfabetización informacional, donde se detectaron beneficios relevantes en términos de eficiencia de tiempo y costos, aunque los materiales generados requieren una revisión y optimización cuidadosas antes de su implementación en programas de instrucción (Madunić y Sovulj, 2024).

En el contexto de las humanidades, se realizó un análisis comparativo de seis chatbots, incluyendo ChatGPT, en la redacción científica (Lozić y Štular, 2023). Por otra parte, el artículo (Yager, 2023), en el área de ciencias físicas, presentó un chatbot adaptado a temas científicos específicos mediante incrustaciones de texto para proporcionar información contextual relevante; dicho enfoque demostró ser útil para los científicos en la agilización de sus esfuerzos de investigación, situación que sugiere que los modelos de lenguaje pueden adaptarse efectivamente para su uso en dominios específicos. Finalmente, en Indonesia, se investigó la percepción de los expertos del área de edafología sobre el uso de ChatGPT; los resultados mostraron una puntuación alta en la calidad de las respuestas generadas por ChatGPT-4 y se percibió que estos chatbots pueden servir como herramientas de asistencia, pero no pueden reemplazar el conocimiento experto (Cahyana *et al.*, 2024).

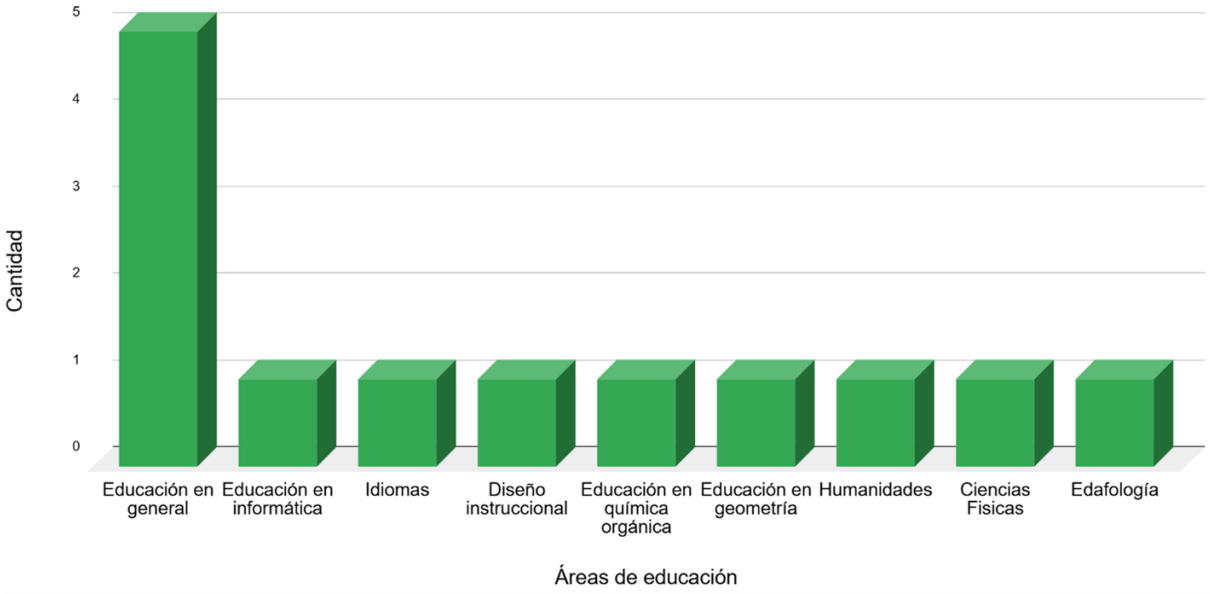

Figura 5. Áreas de investigación en educación

En el campo de la salud se encuentra la mayor cantidad de artículos que incluye este mapeo; se tratan diversos enfoques sobre el uso de LLM y su implementación en este ámbito, específicamente en la elaboración de chatbots y sistemas de interacción (figura **7**). A continuación, se expone un resumen de los hallazgos agrupados según los artículos analizados (figura **6**).

En el artículo de Cascella *et al.* (2024) se presenta una cronología de los LLM, desde diciembre de 2022 hasta diciembre de 2023; a su vez, se resalta cómo a lo largo de 2023 ha habido un aumento en la aparición de estos modelos de lenguaje grandes, lo que ha facilitado su uso en aplicaciones como chatbots y asistentes virtuales personalizados para el sector salud, lo cual brinda una mejora en la gestión de enfermedades crónicas y la interacción con pacientes. Asimismo, se ha trabajado en la adaptación de modelos de lenguaje específicos para el área mé-

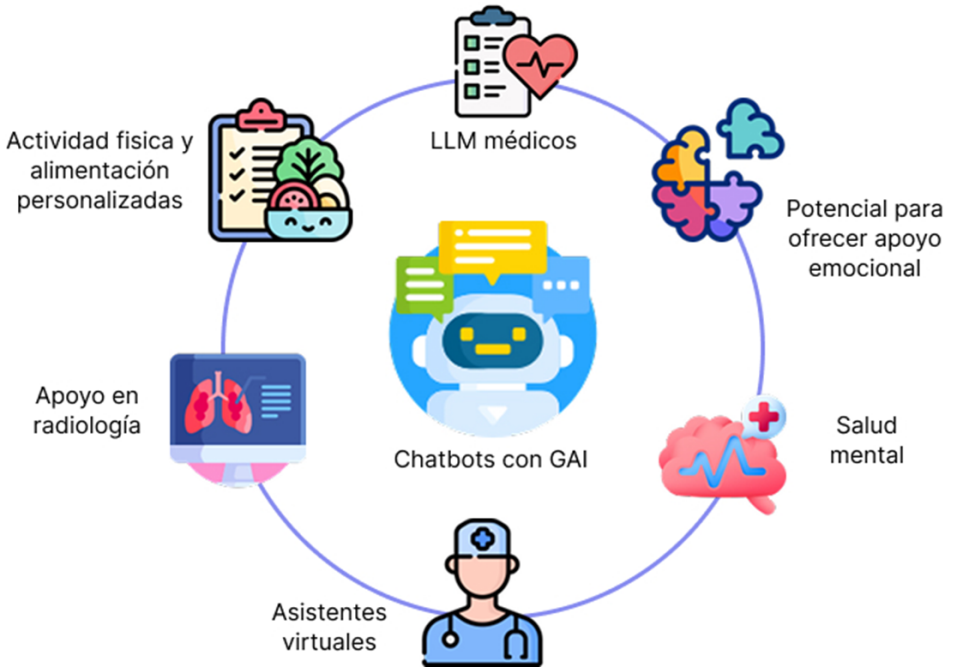

Figura 6. Resumen de casos prácticos de chatbots con GAI en salud

dica, como el caso de "ChatDoctor" (Li, Y. *et al.*, 2023), un modelo ajustado sobre LLaMA que recurre a conocimientos del dominio médico. Concretamente, este modelo fue entrenado con diálogos reales entre médicos y pacientes, lo cual optimizó significativamente su capacidad para entender las necesidades de los pacientes y ofrecer asesoramiento preciso. La incorporación de mecanismos de recuperación de información en tiempo real de fuentes confiables también ha potenciado su precisión, con un avance importante en la capacidad de los LLM para ofrecer respuestas informadas en contextos de alto riesgo como la medicina. Por otro lado, Meskó y Topol (2023) subrayan la necesidad de una supervisión regulatoria para el uso de LLM como GPT-4 en el cuidado de la salud, dado que, a diferencia de las tecnologías médicas convencionales, estos modelos no están regulados de la misma manera, lo que plantea ciertos riesgos de seguridad y privacidad para los pacientes. También es importante mencionar que el uso de ChatGPT en servicios de salud se presenta como una perspectiva innovadora con aplicaciones significativas en el apoyo al cuidado del paciente, la investigación y la planificación de tratamientos (Javaid, Haleem y Singh, 2023).

Dentro de las áreas con más cantidad de artículos se encuentra la salud mental, contexto en el que se concentra el potencial de los chatbots con GAI para ofrecer apoyo emocional, terapia y asistencia en la detección de problemas de salud mental. Un enfoque destacado es el uso de chatbots generativos en el tratamiento y monitoreo de condiciones como la depresión. Por ejemplo, un artículo evaluó la seguridad de chatbots basados en GPT-3.5 en la identificación de

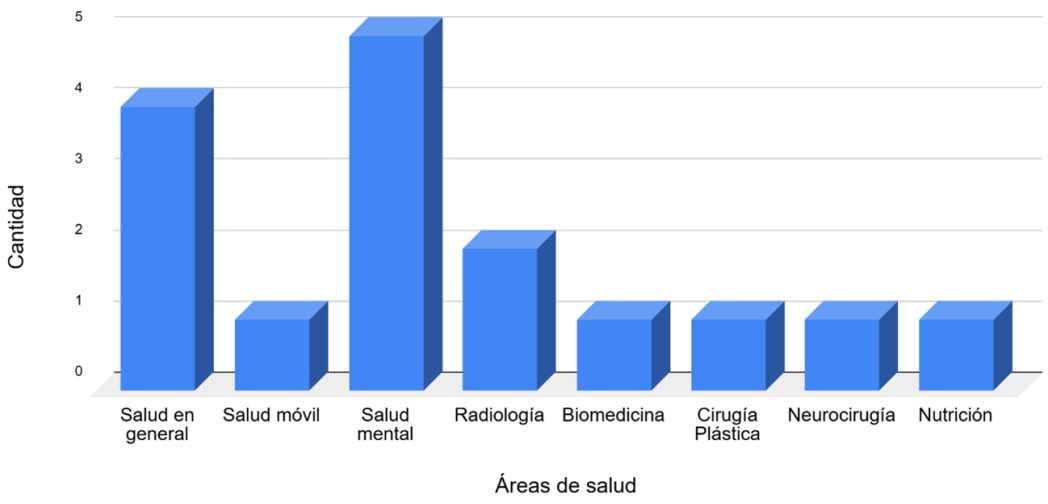

Figura 7. Áreas de investigación en salud

riesgos de suicidio y depresión severa, mediante dos simulaciones de actitudes de un paciente con depresión; dicha evaluación se llevó a cabo sobre 25 chatbots de FlowGPT.com (Heston, 2023).

Otro artículo tuvo como objetivo diseñar, desarrollar y medir la eficacia de un sistema de chatbot con el fin de aumentar la disposición a dejar de fumar mediante reflexiones generativas basadas en entrevistas motivacionales (EM), enfoque que aumentó la confianza y la importancia percibida para dejar de fumar (Brown *et al.*, 2023). Igualmente, en Alotaibi y Alshahre (2024) se investigó el rol de los chatbots en mitigar la soledad y mejorar la conectividad social de personas aisladas; según los hallazgos preliminares, estos agentes conversacionales proporcionan apoyo significativo y compañía personalizada, y ayudan al bienestar general y a la conexión social. En otro artículo (Chowdhury *et al.*, 2024), se utilizó un LLM ajustado, DepGPT, para la detección temprana de depresión con base en conjuntos de datos de redes sociales (Reddit y X), en el idioma bengalí, modelo que superó a otros tan o más avanzados, y logró una precisión casi perfecta y un excelente desempeño en la clasificación de textos depresivos. Otro artículo propuso un chatbot de aprendizaje por refuerzo a partir de retroalimentación humana (RLHF, por su sigla en inglés) como apoyo en las terapias de salud mental; este agente utiliza estrategias adaptativas y personalizadas para interpretar mejor las respuestas emocionales de los usuarios, lo cual es crucial para la provisión de intervenciones terapéuticas a medida (Abubakar *et al.*, 2024).

En cuanto a radiología, encontramos a Kim *et al.* (2024), quienes ofrecen una visión integral sobre la aplicación de modelos generativos en el sector médico, en la creación de datos

sintéticos para tratar problemas de privacidad de los pacientes, y en modelos multimodales en la medicina, como LLaVA-Med. También, el artículo (Pan *et al.*, 2024) detalla la evolución y los avances recientes de los grandes modelos de lenguaje en radiología, incluyendo su aplicación en la generación de reportes radiológicos y la educación en radiología; aquí también se mencionan, primero, las ventajas de los modelos multimodales, como un hito en la práctica radiológica, ya que proporcionan nuevas maneras de visualizar y analizar imágenes médicas, y segundo, los retos en cuanto a su implementación.

Desde la cirugía plástica, el artículo de Labouchère y Raffoul (2024) evalúa el uso de ChatGPT y Bard. Compara cómo ambos modelos tienen la capacidad para asistir a cirujanos plásticos, estudiantes de medicina y pacientes. Según los resultados, ambos son útiles para adelantar revisiones científicas y ofrecer material educativo, pero carecen de profundidad en temas específicos. Adicionalmente, se resalta su valor en la ayuda a los pacientes a comprender mejor los riesgos y cuidados pre- y posoperatorios.

En lo que respecta a la neurocirugía, se encontraron estudios como el de Bečulić *et al.* (2024), en el que se realiza una revisión sistemática sobre el uso de ChatGPT en este campo. Se evidencian tanto los beneficios como las limitaciones, incluyendo riesgos asociados a sesgos algorítmicos, y la necesidad de validación de contenido generado por esta herramienta. A pesar de estas limitaciones, ChatGPT muestra un gran potencial para asistir en la planificación quirúrgica, el procesamiento de datos y la creación de planes de tratamiento personalizados. En el área de la biomedicina, el escrito de Li, C. *et al.* (2023) propone un asistente conversacional multimodal para el análisis de imágenes médicas, lo que combina texto e imágenes. Se utiliza un modelo entrenado con datos de figuras y descripciones de PubMed, capaz de responder preguntas abiertas sobre imágenes biomédicas, lo cual mejora el análisis visual de datos médicos, por lo que este modelo se posiciona como una herramienta poderosa para el futuro de la investigación biomédica.

En cuanto a las intervenciones en actividad física, Vandelanotte *et al.* (2023) describen una plataforma móvil que utiliza GAI para promover la actividad física, mediante contenido personalizado en tiempo real. Esta plataforma utiliza procesamiento de lenguaje natural y aprendizaje por refuerzo para ofrecer recomendaciones basadas en datos como GPS y condiciones climáticas. Y con respecto a la nutrición, Yang *et al.* (2024) presentan a ChatDiet, un chatbot para ofrecer recomendaciones alimentarias personalizadas, el cual integra modelos personales y poblacionales lo que proporciona recomendaciones adaptadas a las necesidades nutricionales individuales, con un 92 % de efectividad en pruebas. El valor de esta herramienta se encuentra en su capacidad para explicar las decisiones tomadas, lo cual optimiza la personalización y la interacción en el campo de la nutrición.

También surgen otras áreas de investigación (figuras **8** y **9**), como la industria automotriz, con la creación de chatbots enfocados en mejorar los servicios de atención al cliente y de posventa, mediante la interacción con manuales automotrices (Medeiros *et al.*, 2023). Saka *et al.* (2024) analizan el potencial de los modelos chatbot GPT en la industria de la construcción, donde hay oportunidades para implementarlos en todo el ciclo de vida de un proyecto, con énfasis en la selección y optimización de los materiales. Otra industria es la de manufactura aditiva (impresión 3D), si bien las publicaciones sobre este tema aún son limitadas, Westphal y Seitz (2024) proponen tres casos de uso específicos en los que las herramientas de GAI optimizan los procesos, lo cual provoca que estos sean más rápidos, creativos y rastreables digitalmente.

Figura 8. Resumen de casos prácticos de chatbots con GAI en otras áreas

En Dubravova *et al.* (2024) se explora la posibilidad de integrar modelos de chatbot como GPT en el sector de la seguridad pública, con énfasis en el análisis de datos y la asistencia en la producción de documentos para que estas herramientas alivien la carga administrativa de los agentes de policía, y así se enfoque en tareas clave de seguridad.

Desde una perspectiva más legal, el uso de diversos modelos GAI para resolver tareas legales en el contexto jurídico alemán, ChatGPT-4 logró este objetivo de forma realista; además, aprobó un examen de derecho empresarial en Alemania. No obstante, se recomienda la verificación manual por los profesionales, debido a la variabilidad en casos complejos (Schweitzer y Conrads, 2024). Ya en el área de comercio, se exploran los beneficios potenciales de implementar ChatGPT en las empresas, particularmente para la automatización de procesos como

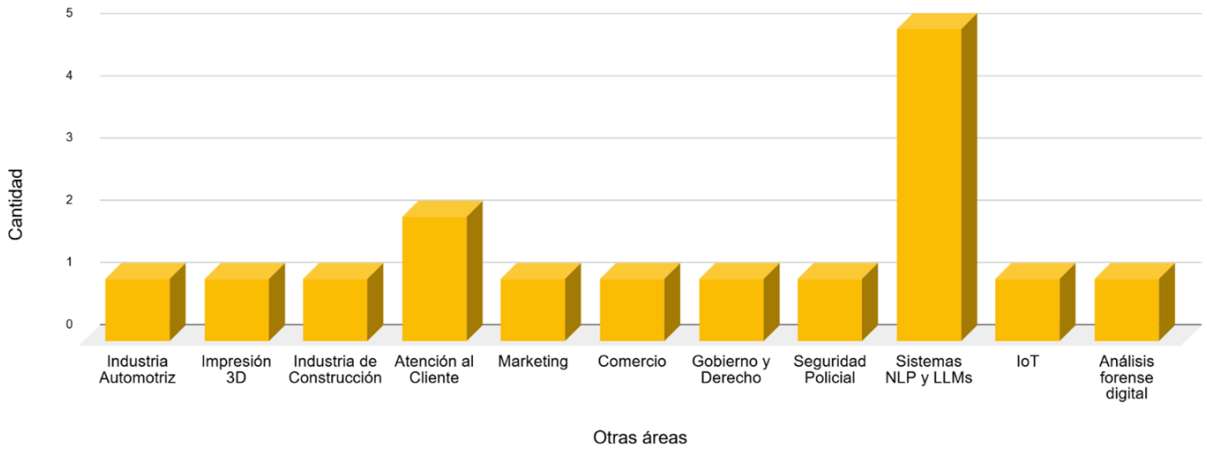

Figura 9. Otras áreas de investigación

el seguimiento de pedidos y facturación. Pero se destaca la necesidad de entrenar y ajustar ChatGPT según los requerimientos de cada empresa, con límites claros y medidas de seguridad que reduzcan riesgos como la generación de noticias falsas o sesgos en las respuestas (Raj *et al.*, 2023). Además, el impacto de esta herramienta en la estrategia de marketing de la industria cosmética en Indonesia se refleja en la capacidad de mejorar la segmentación de mercado, estrategias de precios y la personalización del marketing (Roumeliotis *et al.*, 2024).

En relación con la satisfacción del cliente, Wilendra *et al.* (2024) recalcan cómo estos modelos de GAI tienen el potencial para mejorar y ofrecer un análisis más profundo y preciso del sentimiento del cliente, crucial para la sostenibilidad del *e-commerce.* Además, analizan cómo ChatGPT está transformando el servicio al cliente, mayormente en el sector de la salud, dado que se destaca por su capacidad para superar barreras lingüísticas, mejorar la satisfacción del paciente y ofrecer una asistencia personalizada; sin embargo, el reto consiste en asegurar la exactitud y actualización de la información (Haleem *et al.*, 2024).

En el ámbito tecnológico, en el análisis de 194 estudios sobre ChatGPT y otros LLM, se destacan las innovaciones que han impulsado su rendimiento, como el preentrenamiento a gran escala, la sintonización de instrucciones, y el RLHF (Liu *et al.*, 2023). Chen *et al.* (2024) ofrecen una visión general de la evolución de estos LLM y su transición hacia modelos multimodales (LMM) que integran datos de diferentes modalidades (texto, imágenes, sonido) para mejorar la adaptabilidad de los sistemas de IA.

Existen múltiples modelos LLM para chatbots, los cuales pueden afectar su rendimiento en términos de precisión y tiempo de respuesta; comparar y elegir un buen LLM para una temática

específica impacta directamente en el potencial para la creación de chatbots eficientes y fáciles de usar (Prasad *et al.*, 2024). Con la ayuda de una revisión sistemática de la literatura sobre la tecnología de chatbots, con metodologías de NLP y a través del modelo PRISMA, se examina la evolución, motivación, logros y desafíos en el desarrollo de chatbots en diversas áreas (Suryanto *et al.*, 2023). Por ejemplo, en Scanlon *et al.* (2023) se explora el impacto de ChatGPT en la informática forense, se avalúan casos de uso como la comprensión de artefactos, búsqueda de evidencia, detección de anomalías y respuesta a incidentes. Asimismo, Gill y Kaur (2023) examinan los fundamentos y desafíos de ChatGPT, y exploran sus aplicaciones actuales y su potencial futuro, particularmente en combinación con el internet de las cosas (IoT). Finalmente, Sohail *et al.* (2023) ofrecen una revisión exhaustiva de más de 100 publicaciones indexadas en Scopus sobre ChatGPT, y destacan los desafíos relacionados con los sesgos y la confianza en las respuestas de este modelo, para luego proponer futuras direcciones de investigación para abordar estas limitaciones.

Principales LLM (RQ2)

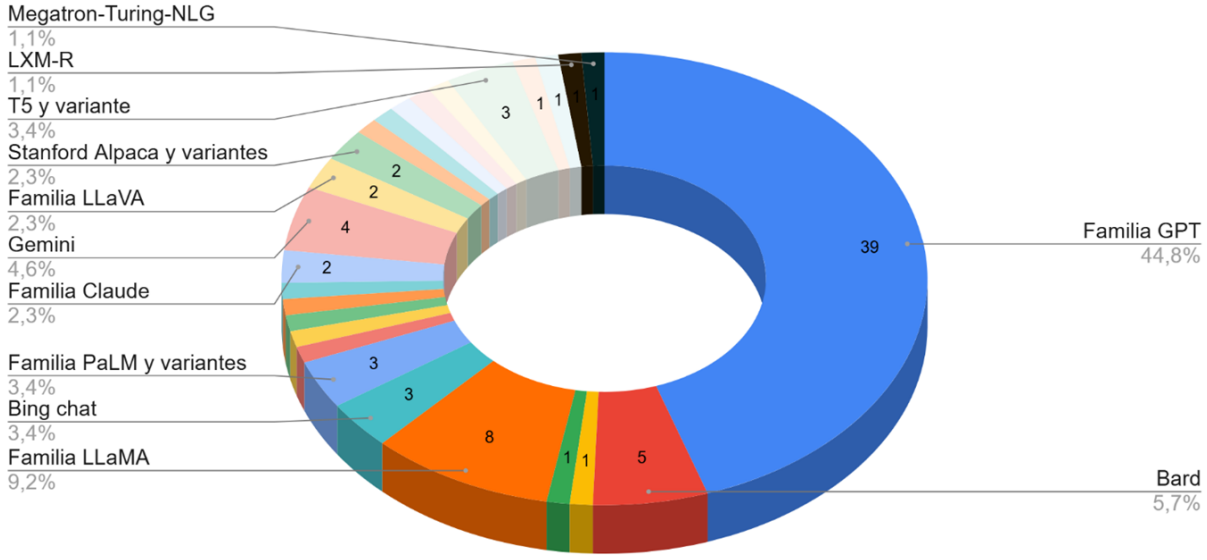

Figura 10. Chatbots y LLM mencionados en los estudios

Para responder a la pregunta de investigación RQ2, se identificaron varios grandes modelos de lenguaje y chatbots, que están siendo utilizados en diferentes contextos. A continuación, se describen los hallazgos principales en los estudios revisados.

La figura **10** ilustra los porcentajes de mención de LLM y un par de chatbots; allí, hay una clara distinción por parte de los modelos de lenguaje GPT (*generative pre-trained transformer*)

desarrollados por OpenAI. Estos últimos son los más mencionados en los artículos, y entre los más destacados se encuentran ChatGPT y sus versiones más recientes (GPT-4 y GPT-3.5) (figura **11**). Como se observa en la tabla 6, los estudios que simplemente usaron ChatGPT o no especificaron el modelo de GPT que utilizaron o evaluaron fueron agrupados en una misma categoría general.

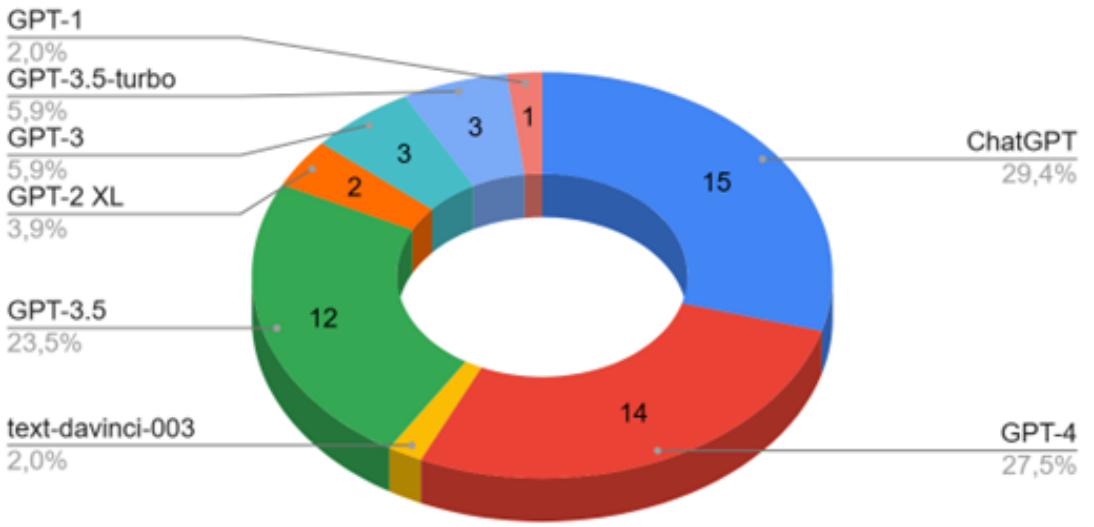

Figura 11. Distribución de ChatGPT y LLM dentro de la familia GPT

En el proceso de lectura de artículos, se determinó que había menciones a LLM que tenían cosas en común, y por tanto se podían agrupar en unas categorías que se denominaron *familias de LLM*, las cuales se detallan en la tabla 7. Dentro de estas familias, la que tiene más influencia en el mapeo, después de GPT, es la familia basada en LLaMA, seguida de cerca por la familia PaLM y MPT-7B. Con un número menor de estudios se tienen las familias de LLM Stanford Alpaca, Mistral 7B, Claude, LLaVA y T5.

Aparte de las familias de LLM mencionadas, se hallaron otra serie de LLM y chatbots nombrados en los estudios, estos pueden observarse en la tabla 8.

Principales desafíos y limitaciones (RQ3)

Diversos estudios han identificado los principales desafíos y limitaciones más comunes en la implementación y uso de chatbots con GAI, los cuales son agrupados en varias categorías para facilitar su explicación (tabla 9).

Uno de los desafíos más reportados es la inexactitud en las respuestas generadas, lo que generalmente se reporta como *alucinaciones* o *afirmaciones incorrectas*. Estos problemas son co-

Tabla 6. Menciones a ChatGPT y la familia de modelos de GPT encontrados

Modelo o chatbot	Estudios
ChatGPT/GPT no especificado	(Javaid, Haleem, Singh, *et al.*, 2023; Zhu *et al.*, 2024; Popovici, 2023; Javaid, Haleem, y Singh, 2023; Brown *et al.*, 2023; Vandelanotte *et al.*, 2023; Saka *et al.*, 2024; Dubravova *et al.*, 2024; Raj *et al.*, 2023; Wilendra *et al.*, 2024; Haleem *et al.*, 2024; Liu *et al.*, 2023; Suryanto *et al.*, 2023; Gill y Kaur, 2023; Sohail *et al.*, 2023)
GTP-4	(Meskó y Topol, 2023; Wölfel *et al.*, 2023; Escalante *et al.*, 2023; Cascella *et al.*, 2024; Parra *et al.*, 2024; Lozić y Štular, 2023; Yager, 2023; Cahyana *et al.*, 2024; Chowdhury *et al.*, 2024; Kim *et al.*, 2024; Bečulić *et al.*, 2024; Li, C.*et al.*, 2023; Schweitzer y Conrads, 2024; Scanlon *et al.*, 2023)
GPT-3.5-turbo	(Madunić y Sovulj, 2024; Yager, 2023; Yang *et al.*, 2024)
GPT-3.5	(Ilagan y Ilagan, 2024; Wölfel *et al.*, 2023; Yik y Dood, 2024; Parra *et al.*, 2024; Lozić y Štular, 2023; Cahyana *et al.*, 2024; Heston, 2023; Chowdhury *et al.*, 2024; Pan *et al.*, 2024; Labouchère y Raffoul, 2024; Schweitzer y Conrads, 2024; Roumeliotis *et al.*, 2024)
GPT-3	(Prasad *et al.*, 2024; Alotaibi y Alshahre, 2024; Pan *et al.*, 2024)
GPT-2 XL	(Brown *et al.*, 2023; Pan *et al.*, 2024)
GPT-1	(Pan *et al.*, 2024)
text-davinci-003	(Medeiros *et al.*, 2023)

munes y perjudiciales en áreas donde se requiere precisión, como la medicina o el derecho. Además, a medida que las conversaciones se extienden, la fiabilidad de estas respuestas disminuye, a la vez que afecta la continuidad del diálogo. De otra manera, los modelos de GAI, al estar entrenados con grandes volúmenes de datos, pueden incorporar sesgos presentes en dichos conjuntos de datos. Estos sesgos pueden llevar a respuestas parciales o desbalanceadas, y a problemas éticos, lo cual impacta la rectitud y afecta negativamente a ciertos grupos de usuarios. Así mismo, los chatbots normalmente enfrentan problemas al interpretar datos no textuales, como imágenes o gráficos, y tienen dificultades para manejar grandes cantidades de contexto en una conversación. Por tanto, la capacidad de los modelos para funcionar de manera acertada en varios idiomas sigue siendo un reto. Del mismo modo, el uso de chatbots que manejan datos personales presenta riesgos de privacidad y seguridad, ya que los datos pueden ser vulnerables a ataques o a un mal uso. Esto es particularmente relevante en áreas como la medicina, donde los historiales médicos deben mantenerse seguros y confidenciales.

Tabla 7. Familias de LLM identificadas

Familia de LLM	Modelos	Estudios
LLaMA	LLaMA	(Cascella *et al.*, 2024; Kim *et al.*, 2024)
	LLaMA-7B	(Li, Y. *et al.*, 2023)
	LLaMA 13B	(Abubakar *et al.*, 2024)
	PMC-LLaMA	(Li, C. *et al.*, 2023)
	LLaMA 2	(Roumeliotis *et al.*, 2024)
	LLaMA 2 7B	(Prasad *et al.*, 2024)
	LLaMA 2 13B	(Prasad *et al.*, 2024)
PaLM	PaLM	(Kim *et al.*, 2024)
	PaLM-2	(Parra *et al.*, 2024)
	Med-PaLM	(Kim *et al.*, 2024)
	Med-PaLM 2	(Cascella *et al.*, 2024)
	Med-PaLM	(Cascella *et al.*, 2024)
MPT-7B	MPT-7B	(Cascella *et al.*, 2024)
	MPT-7B-Instruct	(Cascella *et al.*, 2024)
	MPT-7B-Chat	(Cascella *et al.*, 2024)
	MPT-7B-StoryWriter-65k+	(Cascella *et al.*, 2024)
Mistral 7B	Mistral 7B	(Cascella *et al.*, 2024)
	Mistral 8x7B	(Cascella *et al.*, 2024)
Claude	Claude	(Cascella *et al.*, 2024a)
	Claude 2	(Lozić y Štular, 2023)
LLaVA	LLaVA	(Kim *et al.*, 2024; Li *et al.*, 2023)
	LLaVA-Med	(Kim *et al.*, 2024)
Stanford Alpaca	Stanford Alpaca	(Liu *et al.*, 2023)
	Med-Alpaca	(Li, C. *et al.*, 2023)
	Visual Med-Alpaca	(Li, C. *et al.*, 2023)
T5	T5	(Pan *et al.*, 2024; Suryanto *et al.*, 2023)
	LaMini Flan T5 783M	(Prasad *et al.*, 2024)

Otro reto consiste en que la ejecución de los LLM demanda recursos computacionales significativos, tanto en términos de potencia de procesamiento como de almacenamiento. Esto supone costos altos para su operación y mantenimiento, lo que puede limitar su uso en entornos con restricciones presupuestarias. A parte de eso, los grandes modelos de lenguaje funcionan como ”cajas negras”, lo que significa que es difícil entender cómo llegan a sus conclusiones, esta falta de “explicabilidad” afecta la confianza y la adopción de estas herramientas. Finalmente, a pesar de todas las capacidades de los chatbots generativos, no son capaces de operar de manera autónoma. Se requiere la supervisión constante de expertos para corregir errores y evitar

Tabla 8. Otros LLM y chatbots mencionados

Modelo o chatbot	Estudios
Bard	(Labouchère y Raffoul, 2024; Lozić y Štular, 2023; Pan *et al.*, 2024; Schweitzer y Conrads, 2024; Vandelanotte *et al.*, 2023)
BioMedLM	(Cascella *et al.*, 2024)
Bing chat	(Cascella *et al.*, 2024; Lozić y Štular, 2023; Westphal y Seitz, 2024)
Hippocratic AI	(Cascella *et al.*, 2024)
XGen-7B	(Cascella *et al.*, 2024)
GatorTronGPT	(Cascella *et al.*, 2024)
Gemini	(Cascella *et al.*, 2024; Kim *et al.*, 2024; Schweitzer y Conrads, 2024; Westphal y Seitz, 2024)
Clinical Camel	(Li, C. *et al.*, 2023)
DoctorGLM	(Li, C. *et al.*, 2023)
Huatuo	(Li, C. *et al.*, 2023)
Ernie	(Westphal y Seitz, 2024)
Falcon 7B	(Prasad *et al.*, 2024)
DepGPT	(Chowdhury *et al.*, 2024)
Opera Aria	(Lozić y Štular, 2023)
LXM-R	(Pan *et al.*, 2024)
Megatron-Turing-NLG	(Pan *et al.*, 2024)
Github Copilot	(Popovici, 2023)

consecuencias negativas, sobre todo en sectores como la salud, donde una respuesta incorrecta podría tener serias implicaciones.

Discusión

Los resultados en este mapeo sistemático revelan tendencias y patrones significativos en el desarrollo y la aplicación de chatbots basados en inteligencia artificial generativa (GAI), con un enfoque particular en los grandes modelos de lenguaje (LLM) y los desafíos inherentes a su implementación. A través de una síntesis de la literatura reciente, este estudio proporciona una comprensión contextualizada de cómo la GAI está transformando la interacción con los usuarios en diversos campos.

Tabla 9. Resumen de los principales desafíos o limitaciones encontradas y estudios donde se resaltan

Categoría	Descripción	Estudios
Precisión y fiabilidad	Los chatbots pueden generar respuestas inexactas o alucinaciones.	(Bečulić *et al.*, 2024; Cascella *et al.*, 2024; Chen *et al.*, 2024; Escalante *et al.*, 2023; Haleem *et al.*, 2024; Ilagan y Ilagan, 2024; Javaid, Haleem, y Singh, 2023; Javaid, Haleem, Singh, *et al.*, 2023; Kim *et al.*, 2024; Li, C. *et al.*, 2023; Li, Y. *et al.*, 2023; Lozić y Štular, 2023; Madunić y Sovulj, 2024; Medeiros *et al.*, 2023; Pan *et al.*, 2024; Parra *et al.*, 2024; Popovici, 2023; Prasad *et al.*, 2024; Saka *et al.*, 2024; Scanlon *et al.*, 2023; Schweitzer y Conrads, 2024; Sohail *et al.*, 2023; Suryanto *et al.*, 2023; Vandelanotte *et al.*, 2023; Wölfel *et al.*, 2023; Yager, 2023; Yang *et al.*, 2024; Yik y Dood, 2024)
Sesgos en los datos	Los modelos GAI pueden producir sesgos presentes en los datos de entrenamiento.	(Abubakar *et al.*, 2024; Cascella *et al.*, 2024; Chowdhury *et al.*, 2024; Gill y Kaur, 2023; Ilagan y Ilagan, 2024; Javaid, Haleem, y Singh, 2023; Javaid, Haleem, Singh, *et al.*, 2023; Kim *et al.*, 2024; Li, C. *et al.*, 2023; Liu *et al.*, 2023; Lozić y Štular, 2023; Meskó y Topol, 2023; Pan *et al.*, 2024; Prasad *et al.*, 2024; Saka *et al.*, 2024; Sohail *et al.*, 2023; Suryanto *et al.*, 2023; Westphal y Seitz, 2024; Wilendra *et al.*, 2024; Wölfel *et al.*, 2023; Yager, 2023; Yang *et al.*, 2024; Zhu *et al.*, 2024)
Privacidad y seguridad	Riesgos relacionados con la privacidad de los datos de los usuarios y la seguridad de la información procesada.	(Alotaibi y Alshahre, 2024; Bečulić *et al.*, 2024; Cascella *et al.*, 2024; Dubravova *et al.*, 2024; Escalante *et al.*, 2023; Gill y Kaur, 2023; Haleem *et al.*, 2024; Javaid, Haleem, y Singh, 2023; Javaid, Haleem, Singh, *et al.*, 2023; Li, Y. *et al.*, 2023; Liu *et al.*, 2023; Meskó y Topol, 2023; Pan *et al.*, 2024; Roumeliotis *et al.*, 2024; Saka *et al.*, 2024; Westphal y Seitz, 2024; Wilendra *et al.*, 2024; Yang *et al.*, 2024)
Costos computacionales y monetarios	El uso de modelos generativos de gran escala implica altos costos computacionales y monetarios, lo que limita su implementación masiva.	(Chowdhury *et al.*, 2024; Kim *et al.*, 2024; Madunić y Sovulj, 2024; Medeiros *et al.*, 2023; Prasad *et al.*, 2024; Raj *et al.*, 2023; Saka *et al.*, 2024; Suryanto *et al.*, 2023; Westphal y Seitz, 2024; Wölfel *et al.*, 2023)

Limitaciones técnicas	Dificultades para interpretar, por ejemplo, elementos visuales, gestionar contextos largos y adaptarse a múltiples idiomas.	(Abubakar *et al.*, 2024; Bečulić *et al.*, 2024; Cascella *et al.*, 2024; Chen *et al.*, 2024; Chowdhury *et al.*, 2024; Haleem *et al.*, 2024; Heston, 2023; Javaid, Haleem, y Singh, 2023; Javaid, Haleem, Singh, *et al.*, 2023; Medeiros *et al.*, 2023; Pan *et al.*, 2024; Parra *et al.*, 2024; Prasad *et al.*, 2024; Saka *et al.*, 2024; Westphal y Seitz, 2024; Zhu *et al.*, 2024)
Control y supervisión	Los chatbots requieren supervisión humana constante para evitar errores que pueden ser graves.	(Bečulić *et al.*, 2024; Cahyana *et al.*, 2024; Dubravova *et al.*, 2024; Gill y Kaur, 2023; Javaid, Haleem, Singh, *et al.*, 2023; Kim *et al.*, 2024; Labouchère y Raffoul, 2024; Madunić y Sovulj, 2024; Meskó y Topol, 2023; Popovici, 2023; Schweitzer y Conrads, 2024; Westphal y Seitz, 2024; Wilendra *et al.*, 2024)
Falta de explicabilidad	La çaja negra"de los LLM dificulta entender cómo llegan a sus conclusiones.	(Cascella *et al.*, 2024; Kim *et al.*, 2024; Meskó y Topol, 2023; Saka *et al.*, 2024; Wölfel *et al.*, 2023)

Implicaciones del dominio de educación y salud

El presente estudio revela una clara prevalencia de la investigación en los sectores de educación y salud, lo que subraya la importancia crítica de la interacción personalizada y la gestión de información compleja en estos ámbitos. En el sector educativo, esta inclinación sugiere un vasto potencial para transformar los procesos de aprendizaje, no solo mediante la personalización de contenidos y la automatización de tareas, sino también al ofrecer una interacción más flexible y adaptativa que las herramientas pedagógicas tradicionales. Para instituciones como la Universidad del Cauca, esto supone una oportunidad estratégica para mejorar la experiencia universitaria de estudiantes, profesores y personal administrativo.

Por su parte, el campo de la salud concentra la mayor cantidad de estudios, lo que indica una urgencia y un beneficio significativo en la implementación de LLM para asistencia en diagnóstico, gestión de enfermedades crónicas y mejora de la interacción con los pacientes. La atención a la salud mental destaca el rol de los chatbots con GAI en ofrecer apoyo emocional y detección temprana, evidencia de cómo estas tecnologías pueden cubrir brechas importantes en la atención, siempre bajo una estricta supervisión. La flexibilidad de la GAI permite su adaptación a áreas especializadas, desde la radiología hasta la nutrición y la neurocirugía, lo que abre caminos para la creación de datos sintéticos, la generación de reportes y la personalización de recomendaciones. Esta concentración en áreas de la alta interacción y datos sensibles define la dirección futura de la investigación y las aplicaciones más impactantes de los chatbots generativos.

Dominio de OpenAI y la emergencia de la especialización

Los resultados del mapeo sistemático revelan un claro domino de las aplicaciones desarrolladas por OpenAI, especialmente ChatGPT y sus versiones GPT-3.5 y GPT-4, en el panorama del uso de chatbots con GAI. Esta preponderancia se atribuye a la robustez y versatilidad de estos modelos, que han demostrado un rendimiento superior en una amplia variedad de tareas. La popularidad de ChatGPT se debe a su capacidad para generar texto coherente y natural, lo que lo convierte en una herramienta clave en educación, salud y comercio.

Sin embargo, es igualmente relevante destacar la presencia y especialización de otras familias de LLM, la familia LLaMA (*large language model meta AI*), por ejemplo, se posiciona como la segunda con mayor influencia en el mapeo. Modelos como ChatDoctor, ajustado sobre LLaMA, han demostrado una capacidad significativa para comprender las necesidades de los pacientes y ofrecer asesoramiento médico preciso, incorporando mecanismos de recuperación de información en tiempo real de fuentes confiables. En el panorama de la GAI aplicada a los chatbots, LLaMA ha emergido como una gran opción para el desarrollo de chatbots especializados, particularmente debido a su naturaleza de código abierto, a diferencia de modelos propietarios como los de OpenAI (Li, Y. *et al.*, 2023; Roumeliotis *et al.*, 2024).

Los modelos LLaMA están construidos sobre la arquitectura *transformer,* que utiliza mecanismos de autoatención para capturar relaciones contextuales en secuencias de entrada, sin depender de capas recurrentes o convolucionales (Cascella *et al.*, 2024). La familia LLaMA abarca un rango de tamaños de parámetros, desde 7000 millones hasta 70 000 millones. Por ejemplo, LLaMA-2 fue preentrenado utilizando 2 billones de tokens de datos de fuentes de acceso público (Prasad *et al.*, 2024).

A pesar de su tamaño relativamente más modesto en comparación con otros LLM (como GPT-3 con 175 billones de parámetros), el modelo LLaMA-7B ha demostrado un rendimiento comparable (Cascella *et al.*, 2024). Esta eficiencia se logra mediante la diversificación de los datos de entrenamiento en lugar de un aumento desproporcionado de los parámetros (Li, Y. *et al.*, 2023). Otra variante, LLaMA-2-13B, es capaz de generar texto de alta calidad en una amplia gama de temas, ofrecer respuestas precisas, proporcionar explicaciones detalladas y entablar conversaciones en lenguaje natural. También es competente en la comprensión y respuesta a consultas en varios idiomas, aunque su rendimiento puede variar según el idioma y el dominio (Prasad *et al.*, 2024).

El verdadero potencial de LLaMA en aplicaciones especializadas, como chatbots, sale a relucir a través del refinamiento (*fine-tunning*) (figura **12**), el cual implica refinar el modelo preen-

trenado, por medio del uso de conjuntos de datos más pequeños y específicos para una tarea o dominio (Roumeliotis *et al.*, 2024). En el contexto del modelo ChatDoctor, LLaMA fue adaptado y refinado con un conjunto de datos de 100 000 diálogos, entre pacientes y médicos. Esto mejoró significativamente su capacidad para comprender las necesidades de los pacientes y brindar consejos informados. Un avance clave en ChatDoctor es su mecanismo de recuperación de información autodirigida, que le permite acceder y utilizar información en tiempo real de fuentes en línea como Wikipedia y bases de datos médicas *off-line*. Esto le permite responder preguntas sobre enfermedades o términos médicos relativamente nuevos que no estaban incluidos en los diálogos de entrenamiento, como la viruela del mono (Mpox) o medicamentos recientemente aprobados como Daybue, donde ChatGPT no pudo proporcionar una respuesta satisfactoria (Li, Y. *et al.*, 2023).

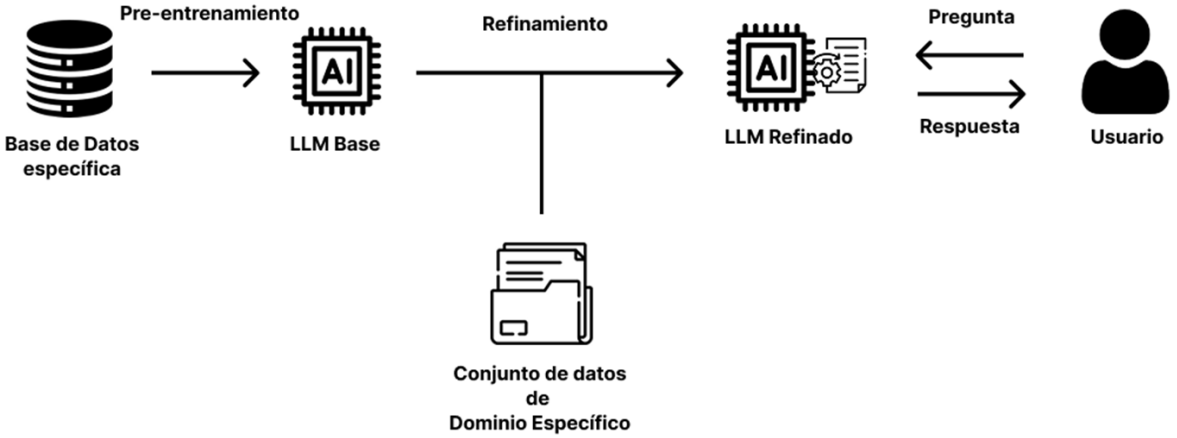

Figura 12. Mejoramiento de un LLM mediante refinamiento (fine-tunning)

El refinamiento es fundamental para lograr una mayor eficiencia de costos y resultados mejorados. En Roumeliotis *et al.* (2024), el modelo base LLaMA-2, por su naturaleza orientada al chat, a menudo producía texto de diálogo, incluso cuando se solicitaba un formato JSON específico; sin embargo, después del refinamiento, se observó una mejora sustancial en la adhesión al formato JSON en todas las respuestas generadas. Esta adaptabilidad es una clara ventaja para integrar chatbots basados en LLaMA en sistemas que requieren un formato de salida estructurado. Además, el refinamiento puede facilitar que los modelos generen resultados equivalentes con indicaciones más concisas, lo que reduce los costos computacionales. La cantidad de datos de entrenamiento también es crucial; una reducción del 50 % en estos resultó en una disminución del 1,45 % en la precisión de respuesta de LLaMA-2, lo que demuestra la importancia de la cantidad y calidad de los datos para el refinamiento.

En este sentido, la integración de LLaMA en chatbots ofrece beneficios. Prasad *et al.* (2024) los describen de la siguiente manera:

- *Experiencia de usuario mejorada.* LLaMa-2-13B se ha destacado por su tiempo de respuesta mejorado y su mayor precisión al responder a las consultas de los usuarios, lo que lo convierte en una opción óptima para chatbots eficientes.

- *Manejo de conversaciones complejas.* Los chatbots basados en LLaMA son capaces de crear respuestas fluidas y coherentes que se asemejan más a los patrones de conversación humana, adaptándose y aprendiendo de las interacciones a lo largo del tiempo, lo que hace que las conversaciones sean más personales y atractivas.

- *Integración de conocimiento externo.* La capacidad de los modelos LLaMA de integrarse con mecanismos de recuperación de información externa, como FAISS para búsqueda de similitud, y de proporcionar referencias de sus fuentes (especialmente a través de interfaces como Chainlit), mejora la credibilidad y la confianza en las respuestas del chatbot.

Por tanto, LLaMA ofrece un equilibrio entre tamaño, rendimiento y flexibilidad, gracias a que es de código abierto y a su capacidad de ser finamente ajustado a dominios específicos. Su éxito en tareas médicas y de comercio electrónico demuestra su potencial para revolucionar la comunicación y la toma de decisiones.

Asimismo, la familia PaLM (*pathways language model*) de Google representa otra vertiente fundamental en la evolución de los LLM, enfocada en la especialización de dominio y en explorar las fronteras de la escalabilidad masiva y las capacidades que emergen de ella. Esta familia es particularmente reconocida en el ámbito de la salud por su variante Med-PaLM, diseñada para abordar la complejidad y la alta exigencia del dominio médico (Cascella *et al.*, 2024; Kim *et al.*, 2024).

El enfoque de Google con PaLM se ha centrado en investigar los límites del rendimiento a través de una escala sin precedentes. En su artículo, Chowdhery *et al.* (2022) presentaron un modelo de 540 000 millones de parámetros, entrenado mediante el sistema de orquestación *Pathways*, una arquitectura de *software* que permite distribuir de manera eficiente el entrenamiento a través de miles de aceleradores. El hallazgo principal de este trabajo fue la demostración empírica de las capacidades emergentes: habilidades complejas que no están presentes en modelos más pequeños, pero que aparecen de manera predecible una vez que el modelo alcanza una escala masiva. Una de las capacidades más notables popularizadas por este trabajo es el razonamiento a través de la técnica de *cadena de pensamiento (chain-of-thought* [CoT]). Al proporcionar al modelo ejemplos de cómo descomponer un problema complejo en pasos intermedios de razonamiento, PaLM fue capaz de resolver tareas de lógica, aritmética y razonamiento simbólico que antes estaban fuera del alcance de los LLM.

El sucesor, PaLM 2, continuó esta línea de trabajo, y se enfocó en mejorar la calidad y diversidad del corpus de entrenamiento para lograr un rendimiento superior con un modelo más eficiente (Anil *et al.*, 2023). El informe técnico de PaLM 2 destaca mejoras significativas en el razonamiento avanzado y en su rendimiento multilingüe. Al entrenarse con un conjunto de datos que abarca una mayor cantidad de idiomas y textos paralelos, PaLM 2 demostró una capacidad muy superior en tareas de traducción, compresión y generación en múltiples lenguas, una característica que ha sido valorada en estudios comparativos (Parra *et al.*, 2024).

La especialización de PaLM en el dominio médico con Med-PaLM y Med-PaLM 2 es un ejemplo claro de su adaptación a contextos específicos. Estas variantes no solo se ajustan con datos médicos, sino que también son reconocidas por su capacidad para crear datos sintéticos, una estrategia que ayuda en la gestión de problemas estrictos de privacidad y confidencialidad de la información de los pacientes (Cascella *et al.*, 2024). Esta diversificación hacia modelos como LLaMA y PaLM evidencia un esfuerzo por adaptar los chatbots a necesidades y contextos que los modelos generalistas no pueden cubrir por completo.

Lo anterior plantea la cuestión crucial de cómo elegir el modelo más adecuado para una aplicación particular, sugiriendo que los desarrolladores buscan modelos que ofrezcan no solo precisión, sino también capacidades adaptativas a dominios específicos.

De los obstáculos a la interpretabilidad

Tras haber identificado los principales obstáculos (figura **13**) en la implementación de chatbots con GAI, resulta fundamental destacar que muchos de estos obstáculos tienen un impacto directo en la confianza y la adopción de esta tecnología. En este sentido, se considera que la falta de explicabilidad emerge como un problema transversal que amplifica riesgos asociados a la precisión, los sesgos y la seguridad, por lo tanto, se plantea no solo como un reto técnico, sino como un requisito esencial para avanzar hacia un uso confiable y responsable de los LLM.

Para fomentar una mayor confianza y uso de los LLM, se requieren avances significativos en la explicabilidad de dichos modelos. La interpretabilidad busca ofrecer explicaciones sobre el funcionamiento del modelo de una manera comprensible para los seres humanos, y para abordar esta necesidad creciente se han propuesto dos enfoques (Cascella *et al.*, 2024):

- *Modelos intrínsecos.* Son aquellos construidos desde su diseño inicial con la transparencia y la interpretabilidad como principios fundamentales. Su arquitectura inherente permite una comprensión más directa de sus decisiones.
- *Modelos post-hoc.* Se aplican después de que el modelo ha sido entrenado para proporcionar explicaciones sobre su comportamiento y predicciones. Estos métodos buscan "abrir" la caja negra sin modificar la estructura original del modelo.

Figura 13. Obstáculos en la implementación de chatbots con GAI

Un aspecto de interés crucial en la interpretabilidad de los modelos se centra en la generación de procesos de configuración altamente controlados y en el análisis de la dinámica de entrenamiento. Con este propósito, por ejemplo, EleutherAI ha desarrollado Pythia, una *suite* que comprende 16 LLM entrenados con datos públicos, diseñada específicamente para analizar el comportamiento y la interpretabilidad de los LLM a través de sus fases de entrenamiento y escalado (Cascella *et al.*, 2024).

Implicaciones prácticas y futuras aplicaciones

Como ya se ha mencionado anteriormente, con este mapeo se busca proporcionar un fundamento para orientar decisiones estratégicas en torno a la implementación de chatbots, particularmente en iniciativas como la del portal web de la Universidad del Cauca. Los hallazgos confirman que la flexibilidad y el rendimiento de modelos como GPT-4 los posicionan como opciones prominentes para la implementación de chatbots. Sin embargo, la variedad de LLM enfocados en aplicaciones especializadas sugiere que la elección del modelo debe estar alineada con las necesidades específicas del proyecto. Para una institución educativa como la Universidad del Cauca, esto implica considerar la robustez de un LLM flexible como GPT-4, para un soporte generalizado a estudiantes, profesores y personal administrativo, pero también explorar modelos especializados si se busca abordar dominios educativos muy específicos.

Los desafíos identificados son cruciales para guiar el desarrollo de chatbots efectivos, teniendo en cuenta oportunidades y limitaciones de las tecnologías actuales. La necesidad de supervisión humana y la mitigación de las "alucinaciones" y los sesgos son consecuencias por mitigar, para así garantizar la precisión y la equidad en las respuestas del chatbot universitario. Asimismo, la consideración de los altos costos computacionales y la búsqueda de modelos eficientes será vital para la sostenibilidad y escalabilidad del chatbot.

Al entender estas dificultades, es posible diseñar estrategias proactivas para atenuarlas y mejorar significativamente las posibilidades de éxito en la implementación del chatbot. En última instancia, este estudio enfatiza que, a pesar de que los LLM son herramientas poderosas, su correcta implementación requiere un enfoque informado y una vigilancia constante, especialmente cuando se manejan interacciones en entornos críticos.

La metodología de mapeo sistemático aplicada en este estudio está diseñada para ser reproducible, es decir, para que otros investigadores puedan obtener resultados similares en cuanto a la identificación, análisis y clasificación de la literatura existente. Sin embargo, es importante mencionar que las aplicaciones y experimentos con chatbots basados en GAI, que son objeto de análisis en este mapeo, presentan limitaciones intrínsecas, lo cual impacta la capacidad de otros investigadores o entidades para reproducir sus resultados en diferentes contextos o escenarios.

Conclusiones

El objetivo de este estudio fue identificar las principales tendencias, modelos LLM, desafíos y limitaciones en el desarrollo de chatbots con inteligencia artificial generativa (GAI), con el fin de orientar su futuro personalizado para el portal web de una universidad. A través de una metodología basada en el enfoque de mapeo sistemático, propuesto en "Systematic mapping studies in software engineering" (Petersen *et al.*, 2008), se ha logrado sintetizar información clave sobre el uso de GAI en chatbots, como respuesta a las preguntas de investigación planteadas. La figura **14** resume los hallazgos principales de este mapeo.

Se descubrió que las áreas de educación y salud dominan la investigación en este campo. Estos sectores representan aproximadamente el 64,4 % (29 de los 45 estudios relevantes analizados), con 13 estudios en educación y 16 en salud. En el ámbito educativo, los chatbots basados en GAI han demostrado su potencial para personalizar contenidos, ofrecer retroalimentación adaptativa y mejorar la experiencia de aprendizaje de los estudiantes. Esto demuestra las oportunidades que GAI puede ofrecer en estos sectores, lo cual es relevante para la universidad en términos de posibles aplicaciones educativas y de soporte estudiantil.

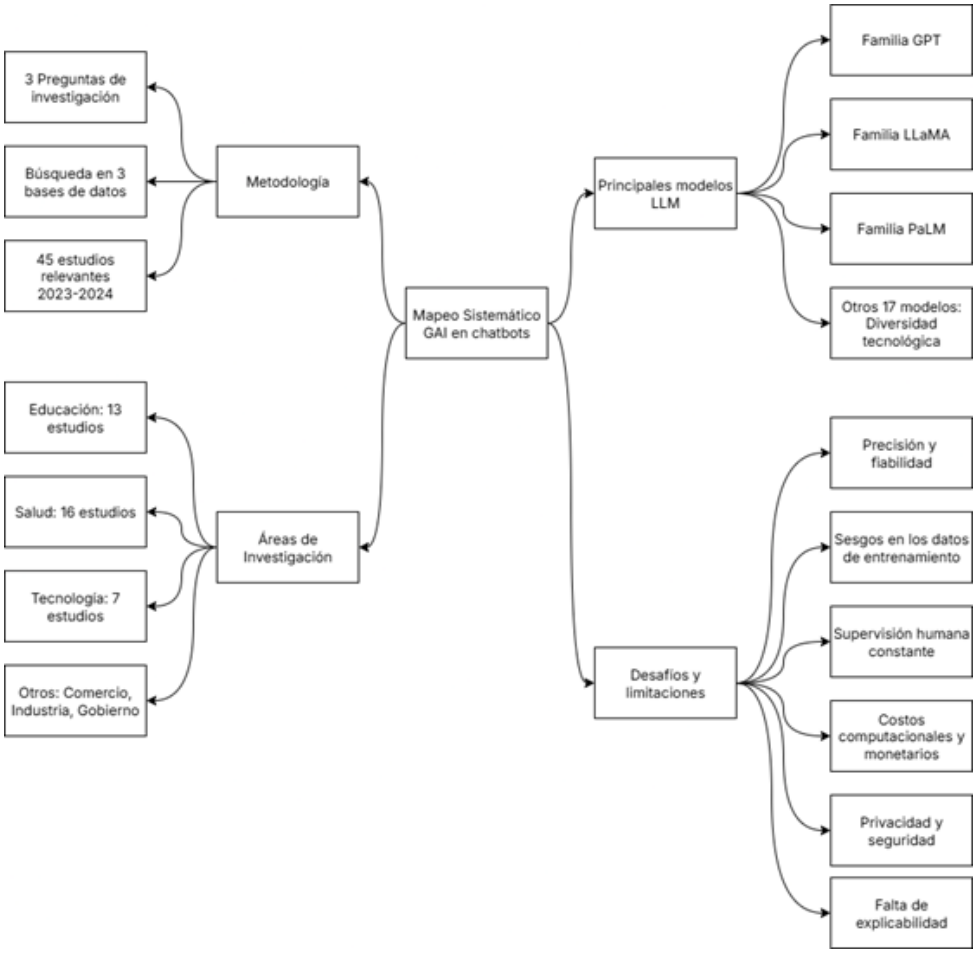

Figura 14. Resumen del artículo

En cuanto a los modelos utilizados, se identificó un dominio de los modelos desarrollados por OpenAI, como GPT-3.5 y GPT-4, que dominan la investigación debido a su versatilidad y rendimiento en diferentes tareas (siendo mencionados en el 79 % de los estudios que especifican el LLM utilizado). Analizando su presencia sobre el total de 45 estudios relevantes, GPT-4 fue mencionado en el 31,1 %; GPT-3.5 en el 26,7 %; y un 33,3 % se refirió a ChatGPT/GPT de forma no especificada. Sin embargo, más allá de la prominencia de GPT, el análisis permitió agrupar otros modelos en familias tecnológicas para evaluar su nivel de influencia en la investigación. La familia LLaMA emerge como la alternativa más influyente, con presencia en 13,3 % de los estudios. Le siguen las familias PaLM y T5, ambas con un 6,7 % de representación, consolidadas como opciones secundarias relevantes.

También, se mencionan las familias Claude, LLaVa y Stanford Alpaca, cada una citada en el 4,4 % de los estudios, lo que indica su adopción en nichos más específicos, mientras que familias como MPT-7B y Mistral fueron identificadas en un 2,2 % cada una. Adicionalmente, se

observó una notable diversidad de otros 17 modelos y chatbots (tabla **8**), como Gemini, Bard y Falcon 7B, que, aunque con menor frecuencia, fueron mencionados en el 26,7 % de los estudios, evidenciando un ecosistema tecnológico amplio y en constante expansión. Esta variedad, especialmente con modelos especializados como Med-PaLM o LLaVA-Med, subraya que la elección del LLM debe estar alineada con las necesidades específicas del proyecto. En este caso, una universidad podría considerar recurrir a un LLM flexible y robusto como GPT-4 para su chatbot, pero también explorar modelos especializados en dominios educativos.

En el uso de chatbots con GAI, se encontraron varios retos clave. El problema de generación de *respuestas inexactas*, o "alucinaciones", o es el obstáculo reportado con mayor frecuencia, con el 62,2 % de los estudios analizados. La necesidad de supervisión humana es crucial, para corregir errores y mitigar sesgos en los datos de entrenamiento, y es mencionada en el 26,7 % de los estudios, mientras que los sesgos inherentes a los datos de entrenamiento fueron identificados como una limitación en el 46,7 %, lo que evidencia que los LLM aún no pueden operar de manera autónoma.

Otros desafíos importantes incluyen los altos costos computacionales y monetarios, señalados en el 22,2 % de los estudios, y la falta de explicabilidad de los modelos (su naturaleza de "caja negra"), un problema resaltado en el 11,1 % lo que afecta la confianza en sus respuestas. Por tanto, estos desafíos deben ser considerados cuidadosamente en el desarrollo de un chatbot para el portal web de una universidad; en particular, garantizar tanto que sea un modelo eficiente en términos de costos, como su supervisión adecuada.

Este trabajo ha proporcionado una visión clara de las áreas más investigadas, los LLM más utilizados y los principales desafíos asociados con los chatbots basados en GAI en los últimos años. Los resultados proporcionados son fundamentales para guiar la selección de tecnologías y estrategias que garanticen el éxito en la implementación de un chatbot educativo, lo cual constituye un aprovechamiento de las oportunidades actuales y anticiparse a posibles dificultades.

Referencias

Abubakar, A. M., Gupta, D., y Parida, S. (2024). A reinforcement learning approach for intelligent conversational chatbot for enhancing mental health therapy. *Procedia Computer Science, 235*, 916-925. https://doi.org/10.1016/j.procs.2024.04.087

Alotaibi, J. O., y Alshahre, A. S. (2024). The role of conversational AI agents in providing support and social care for isolated individuals. *Alexandria Engineering Journal, 108*, 273-284.

https://doi.org/10.1016/j.aej.2024.07.098

Anil, R., Dai, A. M., Firat, O., Johnson, M., Lepikhin, D., Passos, A., Shakeri, S., Taropa, E., Bailey, P., Chen, Z., Chu, E., Clark, J. H., Shafey, L. El, Huang, Y., Meier-Hellstern, K., Mishra, G., Moreira, E., Omernick, M., Robinson, K., … Wu, Y. (2023). PaLM 2 Technical Report. *arXiv:2305.10403*. https://doi.org/10.48550/arXiv.2305.10403

Bečulić, H., Begagić, E., Skomorac, R., Mašović, A., Selimović, E., y Pojskić, M. (2024). ChatGPT's contributions to the evolution of neurosurgical practice and education: a systematic review of benefits, concerns and limitations. *Medicinski Glasnik, 21*(1), 126-131. https://doi.org/10.17392/1661-23

Bengesi, S., El-Sayed, H., Sarker, M. K., Houkpati, Y., Irungu, J., y Oladunni, T. (2024). Advancements in Generative AI: a comprehensive review of GANs, GPT, autoencoders, diffusion model, and transformers. *IEEE Access, 12*, 69812-69837. https://doi.org/10.1109/access.2024.3397775

Brown, A., Kumar, A. T., Melamed, O., Ahmed, I., Wang, Y. H., Deza, A., Morcos, M., Zhu, L., Maslej, M., Minian, N., Sujaya, V., Wolff, J., Doggett, O., Iantorno, M., Ratto, M., Selby, P., y Rose, J. (2023). A motivational interviewing chatbot with generative reflections for increasing readiness to quit smoking: iterative development study. *JMIR Mental Health, 10*, e49132. https://doi.org/10.2196/49132

Cahyana, D., Hadiarto, A., Irawan, N., Hati, D. P., Pratamaningsih, M. M., Karolinoerita, V., Mulyani, A., Sukarman, N., Hikmat, M., Ramadhani, F., Gani, R. A., Yatno, E., Heryanto, R. B., Suratman, N., Gofar, N., y Suriadikusumah, A. (2024). Application of ChatGPT in soil science research and the perceptions of soil scientists in Indonesia. *Artificial Intelligence in Geosciences*, *5*, 100078. https://doi.org/10.1016/j.aiig.2024.100078

Cascella, M., Semeraro, F., Montomoli, J., Bellini, V., Piazza, O., y Bignami, E. (2024a). The breakthrough of large language models release for medical applications: 1-year timeline and perspectives. *Journal of Medical Systems, 48*(1). https://doi.org/10.1007/s10916-024-02045-3

Chen, Z., Xu, L., Zheng, H., Chen, L., Tolba, A., Zhao, L., Yu, K., y Feng, H. (2024). Evolution and prospects of foundation models: From large language models to large multimodal models. Computers, Materials y Continua/Computers, Materials y Continua (Print), 80(2), 1753-1808. https://doi.org/10.32604/cmc.2024.052618

Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H. W., Sutton, C., Gehrmann, S., Schuh, P., Shi, K., Tsvyashchenko, S., Maynez, J., Rao, A.,

Barnes, P., Tay, Y., Shazeer, N., Prabhakaran, V., … Fiedel, N. (2022). PaLM: scaling language modeling with pathways. *arXiv:2204.02311*. https://doi.org/10.48550/arXiv.2204.02311

Chowdhury, A. K., Sujon, Md. S. R., Shafi, Md. S. S., Ahmmad, T., Ahmed, S., Hasib, K. M., y Shah, F. M. (2024). Harnessing large language models over transformer models for detecting Bengali depressive social media text: a comprehensive study. *Natural Language Processing Journal, 7*, 100075. https://doi.org/10.1016/j.nlp.2024.100075

Drelick, A. M., Woodfield, C., y Freedman, J. E. (2024). Educational chatbot development informed by clinical simulations. *Interactive Learning Environments, 33*(3), 2044-2055. https://doi.org/10.1080/10494820.2024.2388782

Dubravova, H., Cap, J., Holubova, K., y Hribnak, L. (2024). Artificial intelligence as an innovative element of support in policing. *Procedia Computer Science, 237*, 237-244. https://doi.org/10.1016/j.procs.2024.05.101

Escalante, J., Pack, A., y Barrett, A. (2023). AI-generated feedback on writing: insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education, 20*(1), 57. https://doi.org/10.1186/s41239-023-00425-2

Gill, S. S., y Kaur, R. (2023). ChatGPT: Vision and challenges. *Internet of Things and Cyber-Physical Systems, 3*, 262-271. https://doi.org/10.1016/j.iotcps.2023.05.004

Haleem, A., Javaid, M., y Singh, R. P. (2024). Exploring the competence of ChatGPT for customer and patient service management. *Intelligent Pharmacy, 2*(3), 392-414. https://doi.org/10.1016/j.ipha.2024.03.002

Heston, T. F. (2023). Safety of large language models in addressing depression. *Cureus*. https://doi.org/10.7759/cureus.50729

Ilagan, J. B., y Ilagan, J. R. (2024). A prototype of a conversational virtual university support agent powered by a large language model that addresses inquiries about policies in the student handbook. *Procedia Computer Science, 239*, 1124-1131. https://doi.org/10.1016/j.procs.2024.06.278

Javaid, M., Haleem, A., y Singh, R. P. (2023). ChatGPT for healthcare services: an emerging stage for an innovative perspective. *BenchCouncil Transactions on Benchmarks Standards and Evaluations, 3*(1), 100105. https://doi.org/10.1016/j.tbench.2023.100105

Javaid, M., Haleem, A., Singh, R. P., Khan, S., y Khan, I. H. (2023). Unlocking the opportunities through ChatGPT Tool towards ameliorating the education system. *BenchCoun-*

cil Transactions on Benchmarks Standards and Evaluations, *3*(2), 100115. https://doi.org/10.1016/j.tbench.2023.100115

Kim, K., Cho, K., Jang, R., Kyung, S., Lee, S., Ham, S., Choi, E., Hong, G.-S., y Kim, N. (2024a). Updated primer on generative artificial intelligence and large language models in medical imaging for medical professionals. *Korean Journal of Radiology*, *25*(3), 224. https://doi.org/10.3348/kjr.2023.0818

Kitchenham, B., y Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering (Issue technical report EBSE 2007-001).* Keele University y Durham University.

Labouchère, A., y Raffoul, W. (2024). ChatGPT and Bard in plastic surgery: hype or hope? *Surgeries*, *5*(1), 37-48. https://doi.org/10.3390/surgeries5010006

Li, C., Wong, C., Zhang, S., Usuyama, N., Liu, H., Yang, J., Naumann, T., Poon, H., y Gao, J. (2023). LLAVA-MeD: training a large language-and-vision assistant for biomedicine in one day. *arXiv:2306.00890.* https://doi.org/10.48550/arxiv.2306.00890

Li, Y., Li, Z., Zhang, K., Dan, R., Jiang, S., y Zhang, Y. (2023). ChatDoctor: a medical chat model fine-tuned on a large language model Meta-AI (LLAMA) using medical domain knowledge. *Cureus.* https://doi.org/10.7759/cureus.40895

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhao, L., Zhu, D., Li, X., Qiang, N., Shen, D., Liu, T., y Ge, B. (2023). Summary of ChatGPT-related research and perspective towards the future of large language models. *Meta-Radiology*, *1*(2), 100017. https://doi.org/10.1016/j.metrad.2023.100017

Lozić, E., y Štular, B. (2023). Fluent but not factual: a comparative analysis of ChatGPT and other AI chatbots' proficiency and originality in scientific writing for humanities. *Future Internet*, *15*(10), 336. https://doi.org/10.3390/fi15100336

Madunić, J., y Sovulj, M. (2024). Application of ChatGPT in information literacy instructional design. *Publications*, *12*(2), 11. https://doi.org/10.3390/publications12020011

Medeiros, T., Medeiros, M., Azevedo, M., Silva, M., Silva, I., y Costa, D. G. (2023). Analysis of language-model-powered chatbots for query resolution in PDF-based automotive manuals. *Vehicles*, *5*(4), 1384-1399. https://doi.org/10.3390/vehicles5040076

Meskó, B., y Topol, E. J. (2023). The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *Npj Digital Medicine*, *6*(1). https://doi.org/10.1038/s41746-023-00873-0

Pan, L., Zhao, Z., Lu, Y., Tang, K., Fu, L., Liang, Q., y Peng, S. (2024). Opportunities and challenges in the application of large artificial intelligence models in radiology. *Meta-Radiology, 2*(2), 100080. https://doi.org/10.1016/j.metrad.2024.100080

Parra, V., Sureda, P., Corica, A., Schiaffino, S., y Godoy, D. (2024). Can generative AI solve geometry problems? Strengths and weaknesses of LLM for geometric reasoning in spanish. *International Journal of Interactive Multimedia and Artificial Intelligence*, (en prensa), 1. https://doi.org/10.9781/ijimai.2024.02.009

Petersen K., Feldt, R., Mujtaba, S., y Mattsson, M. (26-27 de junio de 2008). *Systematic mapping studies in software engineering* [Conferencia]. 12th International Conference on Evaluation and Assessment in Software Engineering (EASE) (EASE).

Popovici, M.-D. (2023). ChatGPT in the classroom. Exploring its potential and limitations in a functional programming course. *International Journal of Human-Computer Interaction, 40*(22), 7743-7754. https://doi.org/10.1080/10447318.2023.2269006

Prasad, S., Gupta, H., y Ghosh, A. (2024). Leveraging the potential of large language models. *Informatica, 48*(8), 1-16. https://doi.org/10.31449/inf.v48i8.5635

Prebble, T., Hargraves, H., Leach, L., Naidoo, K., Suddaby, G., y Zepke, N. (2004). *Impact of student support services and academic development programmes on student outcomes in undergraduate tertiary study.* Ministry of Education Wellington.

Raj, R., Singh, A., Kumar, V., y Verma, P. (2023). Analyzing the potential benefits and use cases of ChatGPT as a tool for improving the efficiency and effectiveness of business operations. *BenchCouncil Transactions on Benchmarks Standards and Evaluations, 3*(3), 100140. https://doi.org/10.1016/j.tbench.2023.100140

Roumeliotis, K. I., Tselikas, N. D., y Nasiopoulos, D. K. (2024). LLM in e-commerce: a comparative analysis of GPT and LLaMA models in product review evaluation. *Natural Language Processing Journal, 6*, 100056. https://doi.org/10.1016/j.nlp.2024.100056

Saka, A., Taiwo, R., Saka, N., Salami, B. A., Ajayi, S., Akande, K., y Kazemi, H. (2024). GPT models in construction industry: opportunities, limitations, and a use case validation. *Developments in the Built Environment, 17*, 100300. https://doi.org/10.1016/j.dibe.2023.100300

Scanlon, M., Breitinger, F., Hargreaves, C., Hilgert, J.-N., y Sheppard, J. (2023). ChatGPT for digital forensic investigation: the good, the bad, and the unknown. *Forensic Science International Digital Investigation, 46*, 301609. https://doi.org/10.1016/j.fsidi.2023.301609

Schweitzer, S., y Conrads, M. (2024). The digital transformation of jurisprudence: an evaluation of ChatGPT-4's applicability to solve cases in business law. *Artificial Intelligence and Law.* https://doi.org/10.1007/s10506-024-09406-w

Seeman, E. D., y O'Hara, M. (2006). Customer relationship management in higher education. *Campus-Wide Information Systems, 23*(1), 24-34. https://doi.org/10.1108/10650740610639714

Sohail, S. S., Farhat, F., Himeur, Y., Nadeem, M., Madsen, D. Ø., Singh, Y., Atalla, S., y Mansoor, W. (2023). Decoding ChatGPT: a taxonomy of existing research, current challenges, and possible future directions. *Journal of King Saud University - Computer and Information Sciences, 35*(8), 101675. https://doi.org/10.1016/j.jksuci.2023.101675

Suryanto, T. L. M., Wibawa, A. P., Hariyono, H., y Nafalski, A. (2023). Evolving conversations: a review of chatbots and implications in natural language processing for cultural heritage ecosystems. *International Journal of Robotics and Control Systems, 3*(4), 955-1006. https://doi.org/10.31763/ijrcs.v3i4.1195

Vandelanotte, C., Trost, S., Hodgetts, D., Imam, T., Rashid, M., To, Q. G., y Maher, C. (2023). Increasing physical activity using a just-in-time adaptive digital assistant supported by machine learning: a novel approach for hyper-personalised mHealth interventions. *Journal of Biomedical Informatics, 144*, 104435. https://doi.org/10.1016/j.jbi.2023.104435

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., y Polosukhin, I. (2023). Attention is all you need. *arXiv:1706.0376.* https://doi.org/10.48550/arXiv.1706.03762

Wang, Y.-C., Xue, J., Wei, C., y Kuo, C. J. (2023). An overview on generative AI at scale with edge-cloud computing. *IEEE Open Journal of the Communications Society, 4*, 2952-2971. https://doi.org/10.1109/ojcoms.2023.3320646

Westphal, E., y Seitz, H. (2024). Generative artificial intelligence: analyzing its future applications in additive manufacturing. *Big Data and Cognitive Computing, 8*(7), 74. https://doi.org/10.3390/bdcc8070074

Wilendra, W., Nadlifatin, R., y Kusumawulan, C. K. (2024). ChatGPT: the AI game-changing revolution in marketing strategy for the indonesian cosmetic industry. *Procedia Computer Science, 234*, 1012-1019. https://doi.org/10.1016/j.procs.2024.03.091

Wölfel, M., Shirzad, M. B., Reich, A., y Anderer, K. (2023). Knowledge-based and generative-AI-driven pedagogical conversational agents: a comparative study of Grice's cooperative principles and trust. *Big Data and Cognitive Computing, 8*(1), 2. https://doi.org/10.3390/bdcc8010002

Yager, K. G. (2023). Domain-specific chatbots for science using embeddings. *Digital Discovery, 2*(6), 1850-1861. https://doi.org/10.1039/d3dd00112a

Yang, Z., Khatibi, E., Nagesh, N., Abbasian, M., Azimi, I., Jain, R., y Rahmani, A. M. (2024). ChatDiet: empowering personalized nutrition-oriented food recommender chatbots through an LLM-augmented framework. *Smart Health, 32*, 100465. https://doi.org/10.1016/j.smhl.2024.100465

Yik, B. J., y Dood, A. J. (2024). ChatGPT convincingly explains organic chemistry reaction mechanisms slightly inaccurately with high levels of explanation sophistication. *Journal of Chemical Education, 101*(5), 1836-1846. https://doi.org/10.1021/acs.jchemed.4c00235

Zhu, S., Wang, Z., Zhuang, Y., Jiang, Y., Guo, M., Zhang, X., y Gao, Z. (2024). Exploring the impact of ChatGPT on art creation and collaboration: benefits, challenges and ethical implications. *Telematics and Informatics Reports, 14*, 100138. https://doi.org/10.1016/j.teler.2024.100138