

Desarrollo de un chatbot inteligente con LLM para diagnóstico médico de la enfermedad pulmonar obstructiva crónica

Development of an intelligent chatbot with LLM for the medical diagnosis of chronic obstructive pulmonary disease

Adrian A. VEGA ¹

Freddy D. ZAMBRANO ²

Joffre J. CARTUCHE ³

Dixys L. HERNÁNDEZ ⁴

¹ Universidad Técnica de Machala. Machala, Ecuador. avega12@utmachala.edu.ec

² Universidad Técnica de Machala. Machala, Ecuador. fzambrano2@utmachala.edu.ec

³ Universidad Técnica de Machala. Machala, Ecuador. jcartuche@utmachala.edu.ec

⁴ Universidad Técnica de Machala. Machala, Ecuador. dhernandez@utmachala.edu.ec

RESUMEN

La Enfermedad Pulmonar Obstructiva Crónica (EPOC) constituye un reto por su detección tardía. Este trabajo presenta un prototipo de chatbot inteligente construido con LLM, que integra faiss y embeddings semánticos aplicados a la fragmentación de datos clínicos. El sistema procesa consultas en lenguaje natural y genera recomendaciones fundamentadas. Los resultados muestran coherencia validada por especialista y el despliegue local confirma su utilidad práctica. Se busca ofrecer una herramienta accesible para apoyar el diagnóstico temprano y el manejo clínico.

Palabras clave: epoc, faiss, embeddings semánticos, fragmentación

ABSTRACT

Chronic obstructive pulmonary disease (COPD) is challenging to detect due to its late onset. This paper presents a prototype intelligent chatbot built with LLM, which integrates faiss and semantic embeddings applied to clinical data fragmentation. The system processes queries in natural language and generates informed recommendations. The results show consistency validated by specialists, and local deployment confirms its practical usefulness. The aim is to offer an accessible tool to support early diagnosis and clinical management.

Key words: epoc, faiss, semantic embeddings, chunking

Recibido: 09/09/2025

Aprobado: 17/10/2025

Publicado: 30/11/2025

1. INTRODUCCIÓN

La enfermedad pulmonar obstructiva crónica (EPOC) es un grave problema de salud pública que afecta a millones de personas en todo el mundo produciendo problemas de respiración. Es una afección del tracto respiratorio progresivo e irreversible con limitación crónica del flujo aéreo que se expone normalmente con disnea, tos crónica, y producción de esputo. La EPOC resulta frecuentemente subdiagnosticada o incorrectamente diagnosticada, sobre todo en sus inicios, lo cual reduce la eficacia de las intervenciones terapéuticas y aumenta de manera importante el riesgo de complicaciones, discapacidad y muerte. La enfermedad ocupa un puesto de prioridad en sistemas de salud debido a su prevalencia elevada, su impacto económico y social, y a la destrucción de calidad de vida en sus pacientes (Global Initiative for Chronic Obstructive Lung Disease, 2025).

El diagnóstico de la EPOC, de acuerdo con los lineamientos internacionales de la iniciativa GOLD, requiere valoración clínica integral y pruebas funcionales como la espirometría (Global Initiative for Chronic Obstructive Lung Disease, s. f.). No obstante, en zonas rurales o con recursos limitados, estos procedimientos suelen ser inaccesibles, lo que retrasa el diagnóstico y conduce a tratamientos en fases avanzadas de la enfermedad (Cunha et al., 2024).

Frente a estas limitaciones estructurales del sistema sanitario, se hace necesario explorar alternativas innovadoras que puedan complementar la labor médica y facilitar el acceso temprano a información diagnóstica confiable. Por ello, la inteligencia artificial se ha vuelto esencial para el desarrollo de soluciones digitales basadas en tecnologías de la información (TI) orientadas en la atención médica. De hecho, en 2024, el 66 % de los médicos en EE. UU. reportaron usar herramientas de inteligencia artificial (IA) en su práctica clínica, en comparación con el 38 % del año anterior (American Medical Association, 2025, p. 3). Dentro de esta amplia área, los modelos de lenguaje natural, incluidos los Modelos de Lenguaje de Gran Escala (LLM), han avanzado significativamente en términos de comprensión, procesamiento y generación de lenguaje humano. Un estudio Delphi resalta que LLM tiene el potencial de cambiar significativamente en la atención médica, pero también presentan desafíos en cuanto a su implementación y uso (Denecke et al., 2024).

En este marco, los modelos de lenguaje de gran escala como GPT-4 destacan por su capacidad para conducir entrevistas clínicas estructuradas mediante técnicas de prompt engineering, para la identificación de síntomas clave (Denecke et al., 2024; Meskó, 2023). En EPOC, pueden recopilar antecedentes y síntomas relevantes para generar recomendaciones preliminares. Aunque no sustituyen el juicio médico, constituyen un recurso de apoyo que promueve la detección temprana y la educación en salud (American Medical Association, 2025).

Varios estudios recientes se han centrado en la aplicación de tecnologías digitales en la gestión de la EPOC, sentando las bases de pasos importantes hacia la interacción paciente-tecnología. Wegener et al., (2024) crearon un agente conversacional diseñado basado en agentes para adultos mayores con EPOC, destinado a facilitar la autogestión en el hogar. Aunque su diseño se basaba en una lógica de conversación estructurada, sus resultados demostraron una aceptación positiva y el potencial de este tipo de herramientas para fomentar el autocuidado. Por el contrario, Glynn et al., (2025) evaluaron clínicamente una aplicación de autogestión de pacientes con EPOC para móviles, e informaron de una mejora significativa de los síntomas, la calidad de vida y el cumplimiento del tratamiento tras doce meses de seguimiento. En ambos casos, sin embargo, las plataformas utilizadas carecen de capacidades avanzadas de comprensión contextual y generación lingüística. Por el contrario, la aplicación de modelos lingüísticos a gran escala ofrece la posibilidad de dotar a dichos sistemas de una interacción más flexible, empática y clínicamente informada, abriendo nuevas vías para su aplicación en procesos de orientación diagnóstica.

A nivel global, Robertson et al., (2024) plantean que la inteligencia artificial puede mejorar la clasificación de la severidad, predecir exacerbaciones y superar las barreras de acceso en países de ingresos bajos y medios, proponiendo un enfoque multimodal que integre imágenes, espirometría y biomarcadores. De forma complementaria, Schachner et al., (2020) evidencian que la mayoría de agentes conversacionales en enfermedades crónicas son prototipos tempranos con limitada estandarización y uso de técnicas avanzadas, lo que abre la oportunidad para propuestas basadas en LLM aplicadas a la EPOC.

En este trabajo, se presenta un chatbot médico que utiliza GPT-4 para ayudar en el diagnóstico preliminar de la EPOC mediante la interacción en lenguaje natural y la generación de recomendaciones orientadas por guías clínicas. Este proceso se organiza siguiendo la metodología CRISP-DM, conocida por su enfoque estructurado y reproducible en proyectos de Inteligencia Artificial (Zia et al., 2022). Esto nos permite evaluar de manera realista como los modelos de lenguaje grande (LLM) pueden ser útiles en la atención primaria, especialmente en situaciones donde el acceso es limitado.

2. METODOLOGÍA

El objetivo de este trabajo es el desarrollo de un chatbot para el diagnóstico temprano de la EPOC (Enfermedad Pulmonar Obstructiva Crónica), con el fin de proporcionar recomendaciones médicas precisas y bien fundamentadas para apoyar a los profesionales de la salud. Para ello, se aplica la metodología CRISP-DM, que orientará cada paso del desarrollo, desde la recopilación y el procesamiento de datos clínicos hasta la creación del modelo predictivo y la evaluación del chatbot. A través de este enfoque estructurado, se busca asegurar que el chatbot no solo sea técnicamente robusto, sino que también pueda generar respuestas personalizadas basadas en datos reales de pacientes y adaptarse a las necesidades clínicas de los médicos.

Para materializar este enfoque, se diseñó una arquitectura que conecta la extracción de datos clínicos en PDF con su indexación en FAISS, una biblioteca optimizada para búsquedas de similitud en vectores embebidos y la posterior generación de respuestas mediante GPT-4o, integrando opcionalmente información web. Este flujo adopta un esquema de Recuperación-Aumentada-de-Generación (RAG), el cual combina mecanismos de recuperación de información y generación de texto para mejorar precisión y factualidad (Gupta et al., 2024)

En la Figura 1 se muestra esta arquitectura, organizada según las fases de CRISP-DM, lo que facilita comprender la función de cada componente y garantiza coherencia metodológica en el desarrollo del chatbot.

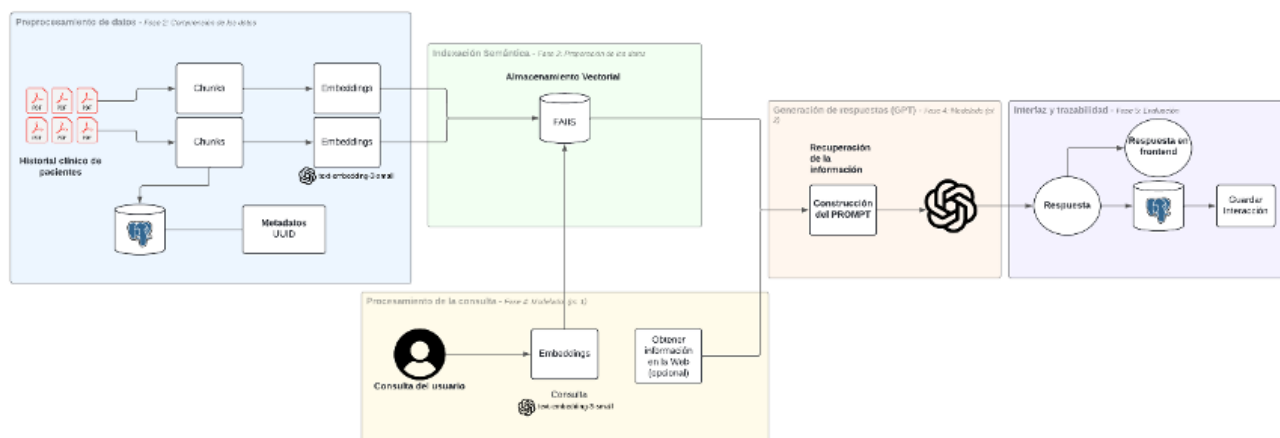


Figura 1 Arquitectura general del prototipo de chatbot para EPOC basada en CRISP-DM, RAG y LLM

La arquitectura del chatbot para recomendaciones médicas sobre la EPOC se presenta en la Figura 1. Esta estructura integra diversas tecnologías para asegurar que las respuestas generadas sean precisas, contextualizadas y fundamentadas en información confiable. Este sistema combina procesamiento de documentos clínicos en formato PDF, almacenamiento vectorial mediante FAISS, consulta web para información actualizada, y generación conversacional utilizando el modelo GPT-4o de OpenAI.

Los modelos de lenguaje de gran tamaño (LLMs) han demostrado un notable desempeño en generación de texto, pero enfrentan desafíos como la generación de respuestas erróneas, limitaciones en la actualización del conocimiento y la falta de trazabilidad en las fuentes utilizadas (Denecke et al., 2024). Además, la técnica de generación aumentada por recuperación (RAG) ha sido propuesta para anclar las salidas de los modelos de lenguaje grande (LLM) en datos externos, lo que potencialmente reduce las alucinaciones y mejora la confiabilidad de las respuestas (Omar et al., 2025).

El flujo de trabajo del sistema comienza con la recepción de la consulta del usuario, la cual es convertida en un vector de embeddings para realizar una búsqueda semántica eficiente en el índice vectorial FAISS. Los fragmentos textuales más relevantes son recuperados y complementados con información proveniente de la web; esta información aumentada se utiliza para construir un prompt enriquecido que es enviado al modelo GPT-4o para la generación de la respuesta final, garantizando precisión y contextualización.

El desarrollo de este sistema sigue la metodología CRISP-DM, que permite alinear las fases del proyecto con los componentes técnicos de la arquitectura, asegurando un proceso estructurado y coherente en la construcción del chatbot.

2.1. Comprensión del negocio

Esta fase se enfocó en definir claramente el propósito del chatbot, el apoyo a médicos y personal de salud en la toma de decisiones relacionadas con la Enfermedad Pulmonar Obstructiva Crónica (EPOC), brindando recomendaciones preliminares en base a guías clínicas. Se identificó que, para ser clínicamente útil, el sistema debía ser capaz de reconocer y manejar la clasificación GOLD (Global Initiative for Chronic Obstructive Lung Disease), la cual divide a los pacientes en cuatro grados de severidad (GOLD 1 - GOLD 4) según los valores obtenidos en la espirometría y datos de sintomatología. Asimismo, se consideró esencial integrar las escalas de valoración como el Modified Medical Research Council (mMRC) y el COPD Assessment Test (CAT). Este último fue especialmente relevante, ya que su estructura estandarizada de ocho ítems sirvió como base para enriquecer los prompts del chatbot, asegurando que las recomendaciones incluyeran aspectos de calidad de vida del paciente.

La integración de estos elementos permite que el chatbot no solo procese datos clínicos generales, sino que también ofrezca recomendaciones orientadas al grado de severidad de la EPOC. Por ejemplo, sugerir la inclusión de broncodilatadores de acción prolongada en GOLD 2 o plantear oxigenoterapia en GOLD 4, siempre como apoyo al juicio clínico profesional.

Se reconocieron además los desafíos de los modelos de lenguaje generativo, como la generación de respuestas no verificables y descontextualizadas, lo que motivó la selección de una arquitectura híbrida que combine recuperación de información con generación conversacional. Esta decisión conceptual fundamenta la adopción de una arquitectura RAG, que se muestra en la Figura 1, y que guiará el desarrollo técnico posterior.

2.2. Comprensión de los datos

En esta fase se recopilaron y analizaron documentos clínicos en formato PDF que contienen información detallada sobre pacientes diagnosticados con Enfermedad Pulmonar Obstructiva Crónica (EPOC). Cada caso clínico presenta diferentes apartados como sus antecedentes personales y de exposición, incluyendo su historial de tabaquismo y contacto prolongado con biomasa, datos demográficos como edad y sexo, así como resultados de espirometría necesarios para determinar la severidad de la obstrucción. La clasificación GOLD de 2025 se utiliza para categorizar la enfermedad en grados del 1 al 4, mientras que las escalas clínicas mMRC (Modified Medical Research Council) y CAT (COPD Assessment Test) permiten evaluar la disnea y el impacto de la enfermedad en la calidad de vida del paciente. Además, los planes de manejo y la frecuencia de exacerbaciones documentados en los casos aportan información esencial para contextualizar las recomendaciones generadas por el sistema.

Por ejemplo, una paciente de 68 años con tabaquismo prolongado y FEV1 del 51% fue clasificada como GOLD 2, mMRC 2 y CAT 10, ubicándose en el grupo E según GOLD 2025. Este tipo de insumos clínicos constituye la base para el entrenamiento y validación del sistema.

Este tipo de información es extraído de los documentos clínicos en formato PDF, tal como se ilustra en la Figura 2, donde los archivos PDF se procesan para convertirlos en chunks de texto. Estos fragmentos son convertidos en embeddings mediante el modelo de OpenAI, permitiendo que el sistema recupere la información relevante de forma eficiente en las fases posteriores del flujo.

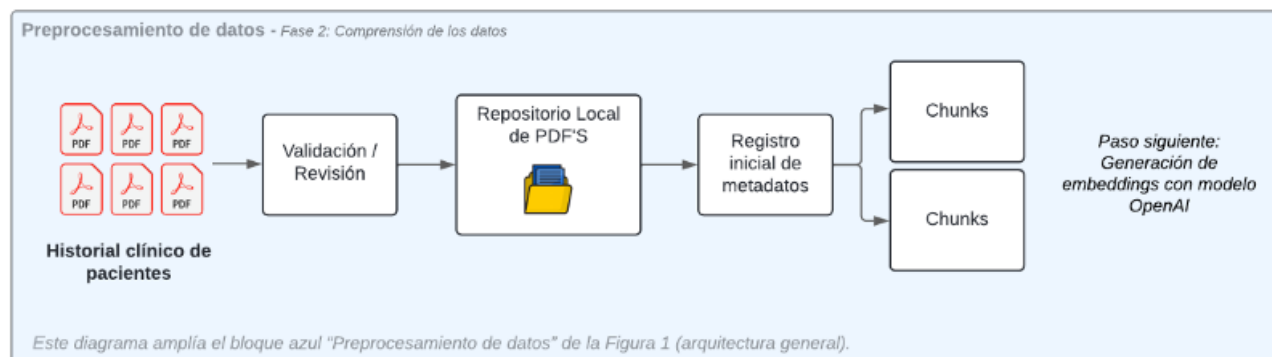


Figura 2 Proceso de recopilación y análisis de datos clínicos en PDF para construir la base de conocimiento del chatbot

Asimismo, se analizó la diversidad y la cantidad de datos disponibles para asegurar que fueran suficientes para entrenar y evaluar el sistema. Este análisis permitió detectar posibles sesgos o vacíos de información que podrían afectar el desempeño del chatbot y planificar acciones correctivas.

2.3. Preparación de los datos

En esta fase, los documentos clínicos recopilados fueron sometidos a un proceso automatizado de extracción y estructuración de la información. Mediante el uso de la librería PyMuPDF, se extrajo el texto de los archivos PDF, aplicando posteriormente técnicas de limpieza y normalización para eliminar caracteres innecesarios y garantizar la coherencia semántica de los datos. El contenido textual fue dividido en fragmentos denominados chunks, diseñados para preservar el contexto clínico de cada sección, como antecedentes, diagnóstico y plan de manejo, asegurando así que las consultas futuras puedan acceder a información relevante de forma precisa.

Cada chunk generado fue transformado en un vector numérico utilizando el modelo preentrenado text-embedding-3-small de OpenAI, lo que permite representar semánticamente su contenido para búsquedas posteriores, este modelo ofrece una mejora notable en precisión semántica pasando de un 31.4 % a un 44.0 % en el benchmark multilingüe MIRACL, y de 61.0 % a 62.3 % en MTEB en comparación con su predecesor, *text-embedding-ada-002*, además de ser más eficiente en costos y recursos (OpenAI, 2025). Estos vectores fueron indexados en la base vectorial FAISS, mientras que los metadatos asociados a cada fragmento incluyendo identificador único, datos del paciente, evaluaciones CAT e historial de chats se almacenaron en PostgreSQL.

Este fragmento es extraído del documento PDF, procesado con PyMuPDF para la extracción de texto, y luego dividido en fragmentos más pequeños o chunks. Como se muestra en la Figura 3, estos chunks se convierten en vectores semánticos a través del modelo text-embedding-3-small de OpenAI, lo que permite que el sistema recupere de manera eficiente fragmentos relacionados en base a la consulta del usuario.

Una vez procesado, el vector semántico del fragmento se almacena en la base vectorial FAISS para búsquedas rápidas y efectivas. Este proceso asegura que la información relevante esté organizada y lista para ser recuperada en fases posteriores, facilitando la generación de respuestas específicas y personalizadas durante la fase de modelado, como se ilustra claramente en la Figura 3.

El proceso de preparación asegura la calidad y consistencia de los datos, mediante técnicas de limpieza y normalización, para que la información sea accesible y útil para la consulta semántica. Esta fase es crítica para minimizar errores y maximizar la precisión en las fases posteriores.

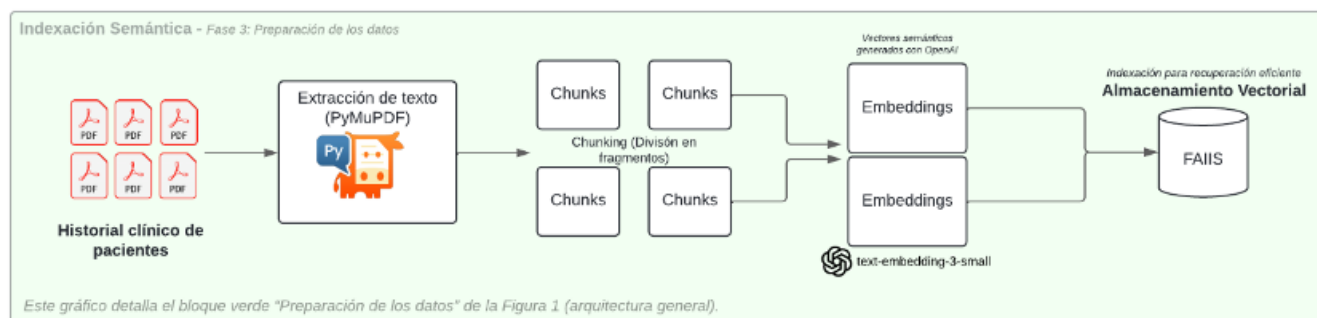


Figura 3 Extracción, limpieza, fragmentación y conversión de documentos clínicos en embeddings semánticos indexados en FAISS

2.4. Modelado

En esta etapa, el sistema recibe una consulta del usuario, la cual es transformada en un embedding mediante el modelo text-embedding-3-small de OpenAI. Este embedding es utilizado para realizar una búsqueda en la base vectorial FAISS y recuperar los fragmentos más relevantes que contienen información clínica procesada en fases anteriores. Opcionalmente, el sistema puede complementar este contexto con resultados obtenidos de búsquedas web confiables.

Tal como se muestra en la Figura 4, el flujo de procesamiento se divide en dos partes: la primera (bloque amarillo) corresponde a la fase de recuperación, donde se identifican y seleccionan los datos relevantes desde FAISS y la web; la segunda (bloque naranja) representa la fase de generación, el prompt final se nutrió tanto de la clasificación GOLD y la escala mMRC como de los puntajes del CAT, lo que permitió generar respuestas ajustadas no solo a la severidad funcional, sino también al impacto percibido en la calidad de vida.

Por ejemplo, ante una consulta como "¿Cuál es el manejo recomendado para un paciente GOLD 3, grupo E, con mMRC 2?", el sistema es capaz de recuperar fragmentos con información sobre casos clínicos similares, que se integran junto con los resultados web en el prompt antes de la generación de la respuesta.

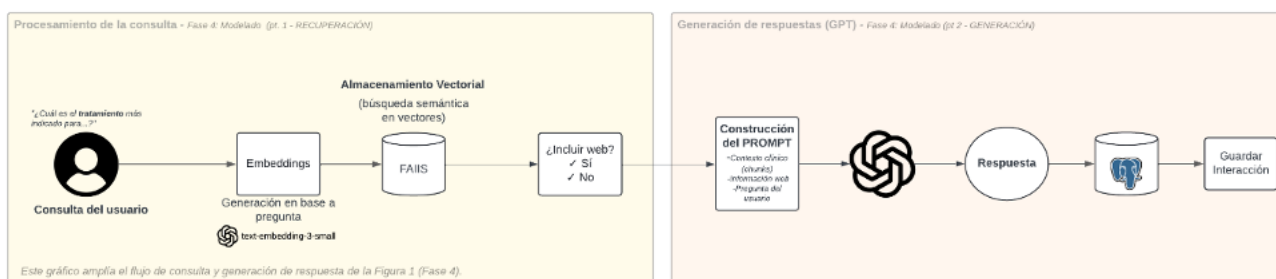


Figura 4 Flujo de procesamiento de consultas médicas en lenguaje natural con recuperación aumentada y generación de respuestas

Se implementan estrategias de manejo de contexto y optimización del prompt para maximizar la relevancia y precisión de las respuestas. Además, el sistema almacena en PostgreSQL tanto la consulta como la respuesta generada y los fragmentos utilizados, estableciendo mecanismos para garantizar la trazabilidad de las fuentes consultadas y fortaleciendo la confianza en las recomendaciones proporcionadas.

2.5. Evaluación

La validación del sistema se estructuró en un proceso escalonado que permite medir tanto la exactitud clínica de las respuestas como el desempeño general del chatbot. Este procedimiento, que combina la aplicación de un banco de preguntas clínicas y la valoración de un especialista, se ilustra en la **Figura 5**.

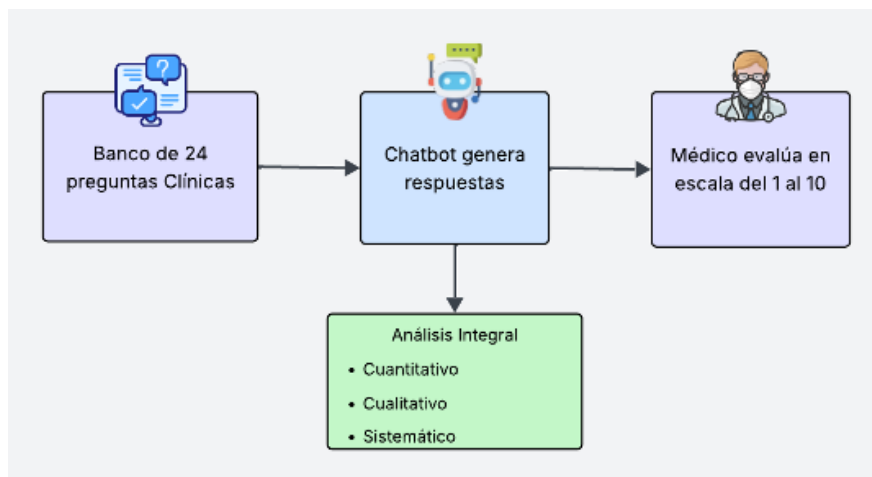


Figura 5 Procedimiento escalonado de evaluación del prototipo de chatbot mediante banco de preguntas clínicas y validación experta

En la primera etapa, se utilizó un banco de 24 preguntas clínicas diseñadas específicamente para cubrir un amplio espectro de escenarios en pacientes con EPOC. Estas preguntas consideran diferentes niveles de severidad GOLD, puntajes en las escalas mMRC y CAT, así como la clasificación en grupos de riesgo. El propósito de este banco fue asegurar que nuestro chatbot fuera evaluado bajo distintas situaciones y en condiciones representativas de la práctica clínica.

En la segunda etapa, el chatbot procesó cada uno de los escenarios y generó respuestas fundadas en la información indexada en la base vectorial, en conjunto con guías internacionales para el manejo de la EPOC. De esta manera, se aseguró que las recomendaciones fueran congruentes con la evidencia clínica disponible.

La tercera etapa correspondió a la evaluación del especialista en EPOC, quien calificaría cada respuesta en una escala del 1 al 10. El número uno reflejó respuestas escasamente pertinentes, mientras que el valor diez representó una respuesta precisa, completa y clínicamente útil. Este paso fue esencial para integrar la evaluación y para comparar el rendimiento del sistema frente a protocolos profesionales.

Finalmente, los resultados se analizarían en tres dimensiones:

- **Cuantitativa:** calculando el promedio de calificaciones, la dispersión de puntajes y la distribución de puntajes.
- **Cualitativa:** se consideraron las apreciaciones del especialista en cuanto a claridad, relevancia y utilidad clínica.
- **Sistémica:** se consideran los parámetros técnicos del tiempo de respuesta, la estabilidad operativa y el lidiar con el contexto de la conversación.

De esta manera, este enfoque mixto de pruebas clínicas simuladas y validación experta garantiza una evaluación completa y realista del rendimiento del chatbot, lo que brinda un equilibrio entre la precisión clínica y la robustez técnica del sistema.

2.6. Despliegue

Como se señaló, la fase de despliegue significa la integración de todos los módulos desarrollados y la ejecución del sistema en un entorno local. El sistema se inserta en una API REST desarrollado en Flask para conectividad en tiempo real entre los subcomponentes de la arquitectura. Esa API es llamada por un frontend en Vue.js para la implementación de la interfaz de usuario y por un backend basado en Flask, que interactúa con la base de datos PostgreSQL, Faiss, y el modelo GPT-4o.

Como se indica en la Figura 6, el flujo comienza con la consulta del usuario desde la interfaz. La solicitud se envía al backend en Flask, donde se orquesta la investigación y generación aumentadas para la recuperación (RAG): la información relevante se extrae primero de FAISS, luego se genera la respuesta

con el modelo de lenguaje y, finalmente, se registra la interacción en la base de datos. El resultado procesado se devuelve al frontend, donde se presenta al usuario de manera clara y estructurada.

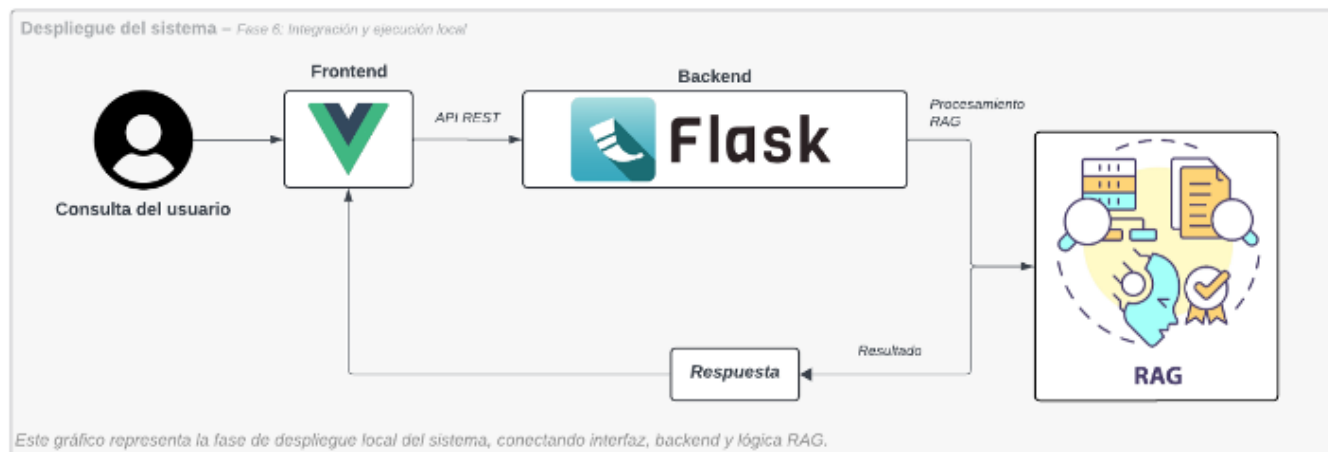


Figura 6 Integración completa de módulos en entorno local: backend en Flask, frontend en Vue.js y base de datos PostgreSQL

Con el objetivo de ilustrar el funcionamiento del prototipo en situaciones clínicas simuladas, se incluyen ejemplos de interacción.

La **Figura 7** muestra la respuesta inicial al ingreso de datos clínicos básicos (*GOLD 2, mMRC 1, tratamiento con LABA/LAMA y antecedente de exacerbación*). El bot reconoce estos parámetros, genera un anclaje clínico coherente y propone un plan de seguimiento.

○ Consulta general ● Chat por paciente

Código o ID del paciente: 12 Cargar historial

Asistente Médico – EPOC

Paciente con EPOC GOLD 2 y un mMRC de 1. Actualmente está en tratamiento con LABA/LAMA (tiotropio/olodaterol) y hace un mes presentó una exacerbación que fue tratada con antibiótico.

El paciente presenta EPOC GOLD 2 y un mMRC de 1. Actualmente está en tratamiento con LABA/LAMA (tiotropio/olodaterol) y ha tenido una exacerbación tratada con antibiótico hace un mes. En su última revisión se observó un mMRC de 1, lo que indica una disnea leve. Es importante monitorear los siguientes aspectos:

- **Controlar síntomas respiratorios:** Evaluar si hay empeoramiento de la disnea o aumento de la tos.
- **Revisar tratamiento:** Asegurarse de que el paciente esté usando correctamente el dispositivo de inhalación.
- **Vacunas:** Considerar la vacunación contra la gripe y el neumococo si no se han administrado.
- **Seguimiento regular:** Programar revisiones periódicas para ajustar el tratamiento según sea necesario.

Es fundamental que el paciente consulte con el médico si presenta **síntomas nuevos o empeoramiento** de su condición, como aumento de la disnea, fiebre o esputo purulento.

Información adicional: Según datos de la web, la EPOC GOLD 2 se asocia a un riesgo moderado de exacerbaciones y es crucial mantener un seguimiento adecuado para prevenir complicaciones.

Haz una pregunta sobre la EPOC... Enviar

Figura 7 Ejemplo de interacción: ingreso de datos clínicos iniciales del paciente y generación de recomendaciones preliminares

En la **Figura 8** se observa la adaptación del sistema ante un incremento sintomático (*mMRC 2, CAT 22 y tratamiento con LABA+ICS*). En este caso, la respuesta se orienta hacia el control de síntomas, la técnica de inhalación y medidas no farmacológicas como la rehabilitación pulmonar, además de un recordatorio sobre la necesidad de ajuste terapéutico.

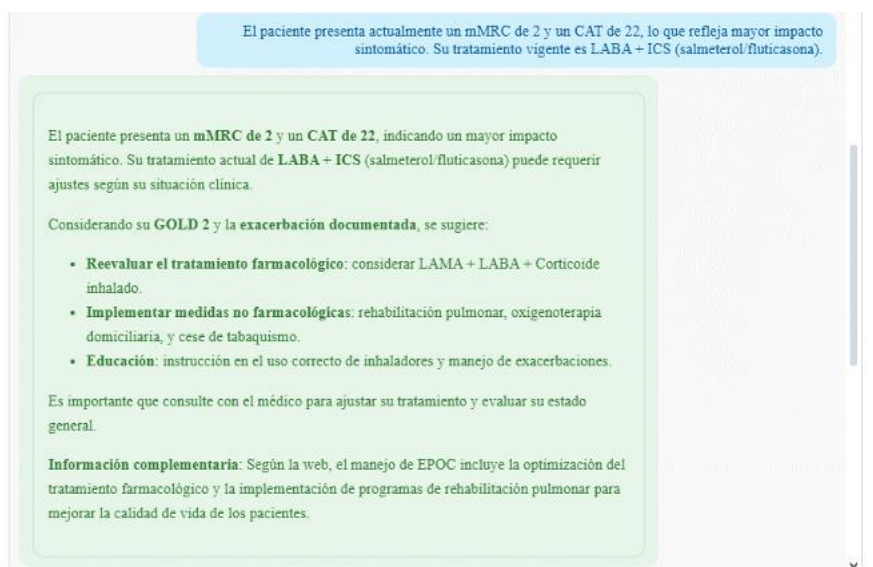


Figura 8 Ejemplo de interacción: actualización del estado clínico del paciente con incremento sintomático y respuesta adaptada del chatbot

Finalmente, la Figura 9 evidencia la capacidad del prototipo de contextualizar situaciones de mayor gravedad clínica (GOLD 3, mMRC 2, CAT 22 y exacerbación reciente tratada con prednisona). La salida integra recomendaciones de seguimiento más estrecho, prevención de nuevas exacerbaciones, posibles ajustes a terapia triple y la indicación explícita de acudir al médico ante signos de alarma.

Estos ejemplos muestran que el prototipo no se limita a generar texto genérico, sino que interpreta los datos clínicos ingresados y adapta sus recomendaciones según la severidad y evolución del paciente.

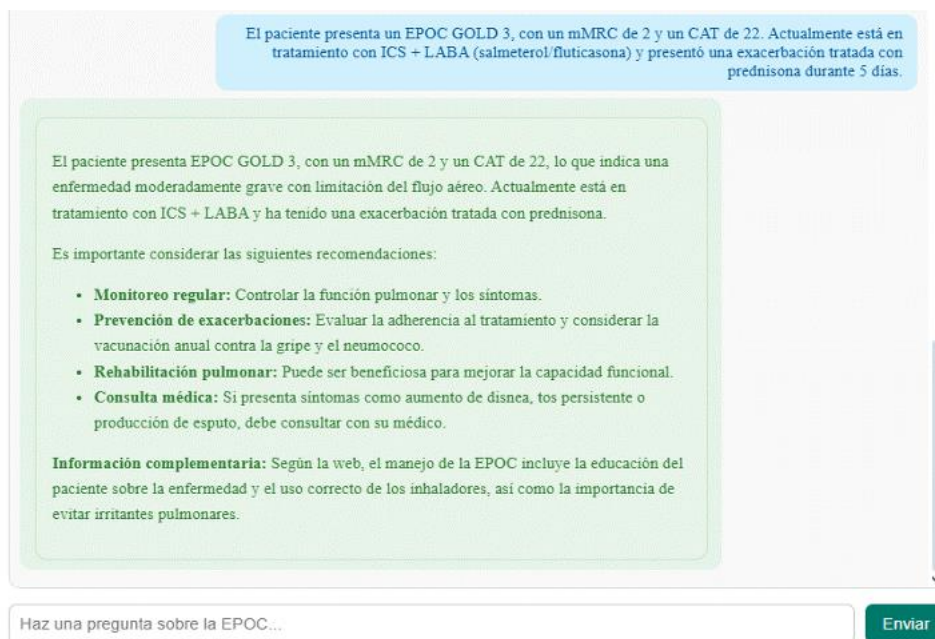


Figura 9 Ejemplo de interacción: respuesta del chatbot ante un escenario clínico de mayor gravedad en paciente con EPOC

3. RESULTADOS Y DISCUSIÓN

La evaluación del sistema mediante el cuestionario de 24 preguntas clínicas permitió obtener una visión detallada del desempeño del chatbot frente a distintos escenarios de pacientes con EPOC. El Cuadro 1 recoge la calificación otorgada por el especialista en cada una de las preguntas, con valores que oscilaron entre 6 y 9 puntos en la escala de 1 a 10.

Cuadro 1 Banco de preguntas

Nº	Pregunta	Puntaje (1-10)
1	¿Cuál es el tratamiento más adecuado para un paciente GOLD 3 con mMRC 2?	9
2	¿Qué broncodilatadores se recomiendan para un paciente GOLD 2 con CAT 15?	8
3	¿Cómo se clasifica un paciente GOLD 4 con mMRC 4 y CAT 30?	9
4	¿Es necesario usar corticosteroides inhalados en un paciente GOLD 1 con mMRC 0?	7
5	Un paciente con GOLD 2 y CAT 10, ¿requiere LAMA o LABA?	8
6	¿Qué medidas adicionales se recomiendan para un paciente GOLD 4 con mMRC 3?	9
7	¿Cuál es el enfoque terapéutico para un paciente con CAT 25 y mMRC 1?	8
8	¿Se recomienda rehabilitación pulmonar en un paciente GOLD 3 con mMRC 2 y CAT 18?	9
9	¿Cómo se estratifica el riesgo en un paciente GOLD 2, mMRC 3 y CAT 21?	8
10	¿Qué tipo de seguimiento es ideal para un paciente GOLD 1 con CAT 5?	7
11	¿Debe hospitalizarse un paciente GOLD 4 con mMRC 4 y CAT 32?	9
12	¿En qué grupo ABCD cae un paciente GOLD 3 con CAT 10 y mMRC 2?	8
13	¿Qué diferencias hay en el manejo entre GOLD 2 y GOLD 3 con síntomas similares?	8
14	¿Es recomendable oxigenoterapia domiciliaria para GOLD 4 con CAT 35?	9
15	¿Cuándo se considera añadir corticosteroides inhalados (ICS) en EPOC?	7
16	Tengo un paciente GOLD 2, mMRC 1, CAT 8. ¿Qué tratamiento inicial recomiendas?	8
17	¿Qué medidas no farmacológicas se recomiendan para grupo C?	8
18	¿Cómo debería manejar a un paciente del grupo D con frecuentes exacerbaciones?	9
19	Si un paciente tiene GOLD 1 y mMRC 0, ¿cuál sería su grupo de riesgo?	7
20	¿Qué hacer si un paciente con EPOC tiene CAT 20 pero mMRC 0?	8
21	¿Se considera grave un paciente GOLD 3 con CAT 28 y mMRC 4?	9
22	¿Qué criterios se usan para cambiar de grupo B a D en el seguimiento?	8
23	¿Puede un paciente con CAT bajo, pero mMRC alto estar en grupo C?	7
24	¿Es válido usar solo el mMRC para clasificar los síntomas en EPOC?	6

Fuente: Elaboración propia

Posteriormente, para facilitar la interpretación de los resultados, se elaboró un **Cuadro 2** que muestra los promedios obtenidos por categorías. Los resultados reflejan un mejor desempeño en las preguntas relacionadas con **tratamiento (8.6)**, mientras que los aspectos de **clasificación y diagnóstico** alcanzaron valores ligeramente menores (7.5–7.8), lo que indica áreas donde el sistema podría optimizarse en futuras iteraciones.

Cuadro 2 Promedios de banco de preguntas

Categoría	Nº de preguntas	Puntaje promedio (1-10)
Tratamiento	12	8.6
Clasificación	6	7.8
Diagnóstico	3	7.5
Seguimiento	3	8.0
Global	24	8.2

Fuente: Elaboración propia

Además del análisis cuantitativo, la prueba incluyó la participación de un médico especialista en EPOC, quien interactuó directamente con el chatbot. El especialista destacó que las respuestas generadas eran clínicamente coherentes, con un nivel de precisión útil para la práctica, y calificó la experiencia de uso

como “muy satisfactoria”, subrayando el potencial de la herramienta para servir como apoyo en contextos de atención primaria.

Estos resultados permiten concluir que el chatbot logra un desempeño sólido en la mayoría de escenarios clínicos planteados, especialmente en la orientación terapéutica. Si bien aún existen oportunidades de mejora en preguntas de clasificación diagnóstica, la validación experta respalda la viabilidad del sistema como recurso complementario en la práctica clínica.

4. CONCLUSIONES

El presente trabajo demostró la viabilidad de implementar un chatbot inteligente basado en modelos de lenguaje de gran escala (LLM) para apoyar el diagnóstico y manejo preliminar de la Enfermedad Pulmonar Obstructiva Crónica (EPOC). La metodología CRISP-DM permitió estructurar de manera ordenada las fases de desarrollo, desde la recopilación y procesamiento de datos clínicos hasta el modelado y validación del sistema.

Los resultados de la evaluación, aplicando un banco de 24 preguntas clínicas y la validación directa de un especialista en EPOC, mostraron un desempeño consistente con un promedio global de 8.2/10. El chatbot alcanzó sus mejores calificaciones en las recomendaciones terapéuticas, mientras que en aspectos de clasificación diagnóstica aún existen márgenes de mejora. No obstante, la opinión experta fue altamente positiva, destacando la coherencia y utilidad clínica de las respuestas generadas.

El despliegue del sistema en un entorno local, acompañado de ejemplos de interacción clínica, evidenció que el prototipo es capaz de interpretar parámetros clínicos ingresados por el usuario y adaptar sus respuestas en función de la severidad y evolución del paciente. Esta característica lo posiciona como una herramienta prometedora para apoyo a la toma de decisiones en atención primaria, especialmente en entornos con recursos limitados.

Finalmente, el sistema desarrollado integró de manera efectiva la clasificación clínica, la validación experta y el despliegue funcional en un prototipo coherente y útil para la práctica médica. Los resultados obtenidos demuestran que la combinación de técnicas de RAG y modelos LLM constituye una alternativa confiable para apoyar el diagnóstico y el manejo preliminar de la EPOC en entornos clínicos reales.

Criterios éticos y transparencia

El presente estudio se desarrolló siguiendo principios éticos de integridad académica y transparencia en la investigación, garantizando el uso responsable de las fuentes y el adecuado manejo de los datos considerados. Se declara que no se han vulnerado derechos de autor, confidencialidad ni normativas relacionadas con la protección de información sensible. En relación con el uso de herramientas de Inteligencia Artificial (IA), estas se emplearon únicamente como apoyo para la redacción y revisión de estilo del manuscrito, así como para la organización de referencias, sin sustituir el análisis crítico ni las aportaciones intelectuales de los autores.

REFERENCIAS BIBLIOGRÁFICAS

- American Medical Association. (2025, febrero 26). 2 in 3 physicians are using health AI—up 78% from 2023. American Medical Association. <https://www.ama-assn.org/practice-management/digital-health/2-3-physicians-are-using-health-ai-78-2023>
- Cunha, A. S., Raposo, B., Dias, F., Henriques, S., Martinho, H., & Pedro, A. R. (2024). Management of Chronic Obstructive Pulmonary Disease: Constraints in Patient Pathway and Mitigation Strategies. *Portuguese Journal of Public Health*, 42(2), 93-100. <https://doi.org/10.1159/000535474>
- Denecke, K., May, R., LLMHealthGroup, & Rivera Romero, O. (2024). Potential of Large Language Models in Health Care: Delphi Study. *Journal of Medical Internet Research*, 26, e52399. <https://doi.org/10.2196/52399>
- Global Initiative for Chronic Obstructive Lung Disease. (2025). Global strategy for the prevention, diagnosis and management of COPD: 2025 Report. GOLD. https://goldcopd.org/wp-content/uploads/2024/11/GOLD-2025-Report-v1.0-15Nov2024_WMV.pdf

- Glynn, L., Moloney, E., Lane, S., McNally, E., Buckley, C., McCann, M., & McCabe, C. (2025). A Smartphone App Self-Management Program for Chronic Obstructive Pulmonary Disease: Randomized Controlled Trial of Clinical Outcomes. *JMIR mHealth and uHealth*, 13, e56318. <https://doi.org/10.2196/56318>
- Meskó, B. (2023). Prompt Engineering as an Important Emerging Skill for Medical Professionals: Tutorial. *Journal of Medical Internet Research*, 25, e50638. <https://doi.org/10.2196/50638>
- Robertson, N. M., Centner, C. S., & Siddharthan, T. (2024). Integrating Artificial Intelligence in the Diagnosis of COPD Globally: A Way Forward. *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation*, 11(1), 114-120. <https://doi.org/10.15326/jcopdf.2023.0449>
- Schachner, T., Keller, R., & V Wangenheim, F. (2020). Artificial Intelligence-Based Conversational Agents for Chronic Conditions: Systematic Literature Review. *Journal of Medical Internet Research*, 22(9), e20701. <https://doi.org/10.2196/20701>
- Wegener, E. K., M Bergschöld, J., Kramer, T., Schmidt, C. W., & Borgnakke, K. (2024). Co-Designing a Conversational Agent With Older Adults With Chronic Obstructive Pulmonary Disease Who Age in Place: Qualitative Study. *JMIR Human Factors*, 11, e63222. <https://doi.org/10.2196/63222>
- Zia, A., Aziz, M., Popa, I., Khan, S. A., Hamedani, A. F., & Asif, A. R. (2022). Artificial Intelligence-Based Medical Data Mining. *Journal of Personalized Medicine*, 12(9), 1359. <https://doi.org/10.3390/jpm12091359>
- Omar, M., Agbareia, R., Glicksberg, B. S., Nadkarni, G. N., & Klang, E. (2025). Benchmarking the Confidence of Large Language Models in Answering Clinical Questions: Cross-Sectional Evaluation Study. *JMIR Medical Informatics*, 13, e66917. <https://doi.org/10.2196/66917>
- Gupta, S., Ranjan, R., & Singh, S. N. (2024). *A Comprehensive Survey of Retrieval-Augmented Generation (RAG): Evolution, Current Landscape and Future Directions* (Versión 1). arXiv. <https://doi.org/10.48550/ARXIV.2410.12837>
- OpenAI. (2024). *Nuevos modelos de integración y actualizaciones de API*. <https://openai.com/es-ES/index/new-embedding-models-and-api-updates/>
- Xu, C., & Jiang, X. (2022). *iMedBot: A Web-based Intelligent Agent for Healthcare Related Prediction and Deep Learning* (Versión 1). arXiv. <https://doi.org/10.48550/ARXIV.2210.05671>



Esta obra está bajo una Licencia Creative Commons
Atribución-NoComercial 4.0 Internacional