
This is the **published version** of the master thesis:

Ayala Cervantes, Juan; Suppi Boldrito, Remo, tut. Desarrollo de un Chat Bot para el laboratorio clínico. 2025. 7 pag. (Màster en Intel·ligència Artificial i Big Data en Salut)

This version is available at <https://ddd.uab.cat/record/318700>

under the terms of the  license

Desarrollo de un Chat Bot para el laboratorio clínico

Juan Ayala Cervantes

Resumen —

El artículo describe el desarrollo de un chatbot utilizando IA, concebido para reducir los errores de la fase preanalítica del laboratorio mediante respuestas 24/7 basadas en un catálogo validado de unas 4 000 pruebas, convertido a JSON y utilizando la técnica RAG (*Retrieval Augmented Generation*) sobre GPT-4o-mini. Ha sido implementado en Python utilizando librerías como LangChain, FAISS, entre otras, y dispone de una interfaz web para mejorar la accesibilidad de este, el sistema además dispone de un servidor de memoria para acelerar las respuestas, el bot también se integra con bases de datos que contienen imágenes de tubos y material de extracciones y se puede clasificar como Dispositivo Médico Clase I.

Se están realizando pruebas piloto en dos laboratorios donde un 80 % de los usuarios lo considera útil, aunque el 30 % encuentra dificultades al formular consultas, lo que subraya la necesidad de mejorar la experiencia de usuario y la detección automática de idioma. La metodología RAG ha permitido minimizar "alucinaciones" y actualizar contenidos sin re-entrenar, pero persisten retos técnicos (manejo heterogéneo de tablas), regulatorios (RGPD) y de explicabilidad clínica. Se prevé ampliar la herramienta a las fases analítica y post-analítica y evolucionar si se pudiera hacia funciones de soporte a la decisión clínica bajo normativa CE, reforzando así la calidad del laboratorio clínico y liberando al personal para tareas de mayor valor.

Palabras clave Chatbot, laboratorio clínico, RAG, preanalítica.

Abstract—

The article describes the development of an AI-based chatbot designed to reduce errors in the pre-analytical phase of the laboratory through 24/7 responses based on a validated catalog of about 4,000 tests, converted to JSON and using the RAG (*Retrieval Augmented Generation*) technique on GPT-4o-mini. It has been implemented in Python using libraries such as LangChain and FAISS and features a web interface to improve accessibility. The system also includes a memory server to speed up responses, and the bot integrates with databases containing images of tubes and extraction materials. It can be classified as Class I Medical Device Software.

Pilot tests are being conducted in two laboratories where 80% of users find it useful, although 30% encounter difficulties in formulating queries, highlighting the need to improve user experience and automatic language detection. RAG has helped minimize "hallucinations" and update content without retraining, but technical (heterogeneous table handling), regulatory (GDPR), and clinical explainability challenges persist. The tool is expected to be expanded to the analytical and post-analytical phases and potentially evolve towards clinical decision support functions under CE regulations, thus enhancing the quality of the clinical laboratory and freeing up staff for higher-value tasks.

Index Terms Chatbot, Clinical laboratory, RAG, pre-analytical

- *E-mail de contacto:* jayalac@lrc.cat
- *Trabajo tutorizado por:* Remo Suppi Boldrino
- *Curso:* 2025



1 INTRODUCCIÓN

El desarrollo de la IA generativa, especialmente de los modelos de lenguaje extendido (LLM), abre la puerta a múltiples aplicaciones en la práctica médica. Estos modelos están disponibles las 24 horas del día y son fácilmente accesibles desde cualquier ordenador, lo que representa una ventaja significativa.

La generación de una petición analítica da origen a una serie de pasos previos antes de que la muestra llegue al laboratorio.

Estos pasos incluyen:

- Una petición médica generada.
- Asignación del día y el centro de extracción.
- Las posibles preparaciones según la prueba solicitada.
- La extracción de la muestra.
- Cualquier preparación previa que la muestra necesite.
- Transporte y llegada al laboratorio.

Todos estos pasos conforman lo que se denomina la "fase preanalítica". Esta fase en un laboratorio de análisis clínicos es, sin duda, la etapa más crítica y propensa a errores. Es aquí donde se origina aproximadamente el 70% de las

incidencias de proceso que pueden comprometer la calidad y fiabilidad de los resultados.

Con el fin de abordar este desafío y mejorar la eficiencia del proceso, se ha desarrollado un chatbot (a la fecha no existe en España ningún desarrollo similar en este campo), innovador en el campo de los análisis clínicos, diseñado específicamente para responder preguntas relacionadas con esta fase crucial.

Este asistente conversacional, construido en Python y utilizando la API de ChatGPT, basado en RAG (*Retrieval Augmented Generation*). presenta un avance significativo en la gestión del conocimiento dentro del laboratorio clínico. Las ventajas son significativas ya que el chatbot está disponible 24/7, ofrece respuestas instantáneas y precisas a las dudas más comunes sobre las distintas pruebas médicas, preparación del paciente, la recolección de muestras, transporte y almacenamiento, entre otros aspectos fundamentales de la fase preanalítica, basándose en un catálogo de aproximadamente 4000 pruebas.

Esta herramienta no solo permite optimizar el flujo de trabajo y reducir los errores, sino que también permite empoderar al personal médico, de enfermería y del laboratorio con información accesible y fiable, contribuyendo a la disminución de incidencias de proceso y la mejora continua en la calidad a favor del paciente.

2 MARCO TEÓRICO:

Los LLM son una clase de modelos de inteligencia artificial (IA) generativa basados en redes neuronales profundas, diseñados para comprender, generar y procesar texto en lenguaje natural. Su origen se remonta a los avances en el Procesamiento del Lenguaje Natural (PLN) y el aprendizaje profundo, particularmente con la introducción de la arquitectura Transformer en 2017 por Vaswani et al. [17]. Esta arquitectura, que utiliza mecanismos de auto-atención (*self-attention*), permitió a los modelos manejar dependencias de largo alcance en el texto de manera más eficiente que las arquitecturas recurrentes (RNN) o convolucionales (CNN) previas.

La característica distintiva de los LLM es su gran escala, tanto en el número de parámetros (billones o incluso trillones) como en el volumen masivo de datos de entrenamiento (terabytes de texto y código de diversas fuentes de internet). Esta escala les permite aprender patrones lingüísticos complejos, semántica, gramática e incluso cierto grado de "conocimiento" general sobre el mundo.

2.1 FUNDAMENTOS Y ARQUITECTURA:

Los LLM se basan en los siguientes principios clave:

- **Modelado de lenguaje predictivo:** Su función principal es predecir la siguiente palabra en una secuencia de texto, dada una secuencia de palabras anteriores. Este proceso se realiza de manera

probabilística, asignando probabilidades a las posibles palabras subsiguientes.

- **Arquitectura Transformer:** Constituye el núcleo de la mayoría de los LLM. Los Transformers se componen de:
 - **Codificador (Encoder):** Procesa la secuencia de entrada para generar representaciones contextuales de cada palabra.
 - **Decodificador (Decoder):** Genera la secuencia de salida palabra por palabra, utilizando las representaciones del codificador y las palabras generadas previamente.
 - **Mecanismos de atención:** Permiten al modelo ponderar la importancia de diferentes partes de la secuencia de entrada al generar cada palabra de salida, capturando dependencias complejas y de largo alcance.
- **Aprendizaje por pre-entrenamiento y ajuste fino (fine-tuning):**
 - **Pre-entrenamiento:** Los LLM se entrena inicialmente en conjuntos de datos masivos y heterogéneos mediante tareas de auto-supervisión (como la predicción de la siguiente palabra o el enmascaramiento de palabras). Este proceso consume una enorme cantidad de recursos computacionales.
 - **Fine-tuning:** Una vez pre-entrenados, los LLM pueden ser "ajustados" para tareas específicas (como clasificación de texto, resumen, traducción o respuesta a preguntas) utilizando conjuntos de datos más pequeños y etiquetados. Esto permite adaptar el modelo a dominios particulares sin necesidad de entrenarlo desde cero.

2.2 CAPACIDADES Y APLICACIONES:

La escala y el entrenamiento extensivo de los LLM les confieren una amplia gama de capacidades:

- **Generación de texto coherente y fluido:** Pueden producir ensayos, artículos, resúmenes, poemas, código de programación y otros tipos de contenido textual que a menudo son indistinguibles del texto generado por humanos.
- **Comprendión del lenguaje natural:** Son capaces de responder preguntas, resumir documentos extensos, clasificar textos, extraer información clave y traducir entre idiomas.
- **Razonamiento y resolución de problemas:** Aunque no poseen una "comprensión" en el sentido humano, pueden exhibir capacidades emergentes de razonamiento al procesar y conectar información de su vasto corpus de entrenamiento, utilizando técnicas como la cadena de pensamiento y dando más tiempo para que el modelo elabore la

respuesta.

- **Conversación interactiva:** Permiten desarrollar chatbots y asistentes virtuales, capaces de mantener diálogos coherentes y contextuales.

2.3 PERSONALIZACIÓN DE LOS LLM:

Los LLM permiten una serie de ajustes como:

Ingeniería de Prompts (Prompt Engineering): la ingeniería de *prompts* se refiere al diseño de las entradas (llamados *prompts*) que se le dan a un LLM para guiar su respuesta hacia el resultado deseado. Dado que los LLM son modelos pre-entrenados con una vasta cantidad de texto, la forma en que se les formula una pregunta o instrucción puede influir drásticamente en la calidad, relevancia y precisión de su salida.

Generación Aumentada por Recuperación (Retrieval Augmented Generation – RAG): RAG es una técnica que combina las capacidades generativas de un LLM con la capacidad de recuperar información de una base de conocimiento externa y autorizada. Los LLM tradicionales están limitados al conocimiento con el que fueron pre-entrenados, lo que puede llevar a respuestas desactualizadas, imprecisiones ("alucinaciones") o la incapacidad de responder preguntas sobre datos muy específicos o propietarios.

Esta técnica tiene una serie de beneficios frente a las otras técnicas de personalización:

- **Reducción de alucinaciones:** Al brindar respuestas del LLM basadas en fuentes de información específicas y verificadas, minimizando la probabilidad de que genere respuestas incorrectas.
- **Acceso a información actualizada y específica:** Permite al LLM responder sobre datos que no estaban en su entrenamiento inicial.
- **Trazabilidad y citación:** RAG permite citar las fuentes de donde se extrajo la información, aumentando la confianza y la auditabilidad.
- **Menor coste que el fine-tuning:** Evita la necesidad de reentrenar o ajustar el modelo cada vez que hay una actualización de datos.

Ajuste fino (fine-tuning): es el proceso de tomar un LLM ya pre-entrenado y afinarlo aún más en un conjunto de datos más pequeño y específico para una tarea o dominio particular. A diferencia del pre-entrenamiento masivo, el fine-tuning actualiza los pesos internos del modelo para que se adapte mejor a las sutilezas, el vocabulario y los patrones de un dominio específico.

Existen diferentes tipos y enfoques de *fine-tuning*:

- **Fine-tuning Supervisado (Supervised Fine-tuning - SFT):** es el método más común, donde el modelo se entrena

con pares de entrada-salida etiquetados para una tarea específica (ej. pregunta-respuesta, instrucción-respuesta).

- **Fine-tuning con Instrucciones (Instruction Fine-tuning):** Se entrena el modelo para seguir instrucciones precisas, a menudo con datos en formato de "instrucción-respuesta".

- **Ajuste Fino Eficiente en Parámetros (Parameter-Efficient Fine-Tuning - PEFT):** incluyen técnicas como LoRA (*Low-Rank Adaptation*) que permiten ajustar un LLM utilizando solo una fracción de sus parámetros, lo que reduce drásticamente los recursos computacionales y el tiempo necesario, a la vez que se evita el "*catastrophic forgetting*" (olvido de conocimientos previos).

2.4 LIMITACIONES DE LOS LLM:

Entre las principales limitaciones de los LLM podemos encontrar:

Alucinaciones: Pueden generar información incorrecta, inventada o sin, lo que es crítico en aplicaciones sensibles como las médicas. Esto se debe a que su objetivo es generar texto plausible, no necesariamente veraz.

Sesgos: Heredan los sesgos presentes en los datos de entrenamiento o de recuperación. Si los datos reflejan estereotipos o discriminación, el LLM puede transferirlos de forma sistemática a sus respuestas.

Falta de "comprensión" real: Aunque simulan la comprensión, los LLM no tienen conciencia ni un entendimiento del mundo como los humanos y operan basándose en patrones estadísticos.

Dependencia de los datos de entrenamiento: Su conocimiento está limitado a los datos con los que fueron entrenados y pueden tener dificultades con información muy reciente o altamente especializada que no estaba en su conjunto inicial.

Coste computacional: El entrenamiento y, en menor medida, la inferencia de LLM son desde el punto de vista computacional muy intensivos y requieren infraestructuras potentes, caras y que ocupan una gran superficie.

Transparencia y explicabilidad: A menudo se consideran "cajas negras", lo que dificulta entender cómo llegan a sus respuestas, lo que es un factor crítico en entornos donde la confianza y la auditabilidad son esenciales (entornos clínicos, por ejemplo). Aunque actualmente existen técnicas y procedimientos que permiten verificar y/o obtener una determinada 'explicabilidad' de sus respuestas.

2.5 CONSIDERACIONES ÉTICAS Y DE SEGURIDAD

El uso de LLM en el ámbito médico conlleva consideraciones éticas y de seguridad cruciales:

Precisión y fiabilidad: Es fundamental garantizar que la información proporcionada sea precisa y no induzca a error y se requieren mecanismos de verificación y validación rigurosos.

Confidencialidad y privacidad: El manejo de datos sensibles de pacientes exige el cumplimiento estricto de normativas vigentes en protección de datos (RGPD), asegurando que la información personal no es comprometida.

Responsabilidad: Determinar la responsabilidad en caso de errores o daños derivados del uso de LLM es un área compleja y en desarrollo.

Supervisión humana: Los LLM deben ser herramientas de ayuda/apoyo a la toma de decisiones, no de sustitución del profesional ya que la supervisión y el juicio humano son indispensables, especialmente en decisiones críticas.

3 OBJETIVOS:

El presente trabajo tiene por objetivo el diseño, desarrollo, implementación y verificación inicial de un agente virtual, especializado en un catálogo de pruebas de laboratorio (aproximadamente de 4000 pruebas), que incluye una amplia información de cada prueba analítica y el cual ha sido validado por varios especialistas para las diferentes pruebas que lo conforman.

Como objetivo complementario con este desarrollo se pretende reducir las incidencias preanalíticas, evitando errores y problemas desde la solicitud de la prueba hasta que los tubos ingresan al laboratorio.

4 MATERIAL Y MÉTODOS

La primera prueba de concepto fue el diseño, desarrollo e implementación de una versión que utiliza el modelo GTP 4.0-mini de OpenAI, siendo un chatbot basado en un catálogo de pruebas del laboratorio en formato XML. Este diseño inicial fue desarrollado utilizando la técnica de *prompt* y la interfaz web de OpenAI. En pruebas preliminares se detectó que este tipo de desarrollo tiene una limitación muy importante, ya que el usuario final debe registrarse como usuario de OpenAI y, además, si no se dispone de una cuenta de pago, el uso está limitado a cierto número de preguntas. No obstante, esta versión está siendo validada por profesionales especialistas de 2 laboratorios que pertenecen a la red de laboratorios participantes, con el objetivo de ajustar las preferencias de uso, mirar su adaptabilidad y adecuación al usuario final entre otras.

Teniendo en cuenta las limitaciones del modelo basado en la técnica del *prompt*, se opta por desarrollar una versión utilizando la API de OpenAI con la finalidad de utilizar RAG. Es posible también utilizar *prompt* con la API de OpenAI, pero tiene la limitación de 200000 tokens como entrada (el catálogo supera los 500000 tokens).

	Valors de referència	Edat	Minim	Màxim
GENERAL			10 %	%
Paràmetre	INTERPRETACIÓ			
Informació Clínica				
Notes:	Valors de tall normals aplicats per al patró citomètric, extrets de l'estudi Fernández-Bañares F, et al. PLoS ONE 2014.			
Nom prova:	HLA B-27 CRIBRATGE Sang			
Codi prova:	5700			
Num mostra etiqueta:	51			
Tipus de mostra:	SANG TOTAL	Termini de lluirament:	3 dies laborables	
Mètode:	Citometria de flux	Tipus de contingut:	T. VERD HU 4 ml	
		Volum de mostra:	1500 Microlitres	
Paràmetre	HLA B-27 CRIBRATGE Sang			
Valors de referència		Edat		
GENERAL				
Paràmetre	HLA B-27 CONFIRMACIÓ Sang			
Valors de referència		Edat		
GENERAL				
Paràmetre	HLA B-27 CONFIRMACIÓ (%) Sang			
Valors de referència		Edat		
GENERAL				
Informació Clínica				
Nom prova:	HLA B-27 CONFIRMACIÓ Sang			
Codi prova:	5701			
Num mostra etiqueta:	51	Termini de lluirament:	15 dies laborables	
Tipus de mostra:	SANG TOTAL	Tipus de contingut:	T. VERD HU 4 ml	
Mètode:	PCR	Volum de mostra:	1500 Microlitres	
Paràmetre	HLA B-27 CONFIRMACIÓ Sang			
Valors de referència		Edat		
GENERAL				
Paràmetre	HLA B-27 CONFIRMACIÓ (%) Sang			
Valors de referència		Edat		
GENERAL				
Informació Clínica				

Fig. 1. Estructura del catálogo en PDF

Uno de los problemas encontrados durante el diseño fue la estructura del catálogo del laboratorio, que es un documento en PDF con cierto orden y que varía según cada prueba (ver Fig 1).

Para transformar este documento en datos, se utilizaron diversas librerías en Python como PyPDF2, pdfminer.six o pdfplumber, pero no se obtuvo un resultado satisfactorio, ya que debido a la estructura del PDF (principalmente por las tablas) no se reconocían los caracteres de forma correcta y muchas veces se incluía información de una prueba mezclada con la prueba siguiente lo cual hacía que la calidad de las respuestas no fuera la más adecuada en muchos casos. También se utilizó una versión en XML, generada con Acrobat (Adobe), la cual presentaba menos errores en cuanto a tablas y caracteres, pero incluía errores en cuanto a algunas claves, estructura de los valores de referencia y también errores al identificar caracteres numéricos.

Teniendo en cuenta estas limitaciones se optó por utilizar el formato JSON, ya que la estructura de etiquetas y jerarquía brinda muy buen resultado al momento de hacer las consultas. Además, la bibliografía especializada sugiere JSON o YAML como los formatos más adecuados para mejorar el desempeño de RAG. [1]

No obstante, esta decisión tuvo como consecuencia reestructurar todo el catálogo a formato JSON, lo cual ha significado un tiempo y esfuerzos adicionales. Para el presente trabajo se utilizó un catálogo JSON con un número reducido de estas pruebas laboratoriales, los resultados preliminares muestran mejor calidad de respuestas y ausencia de alucinaciones, validando así el esfuerzo realizado.

Para la nueva versión del chabot basado en RAG y que utiliza este catálogo reducido como base de conocimiento, se ha encontrado que la mayor dificultad radica en las tablas de los valores de referencia, pues debido a la variedad de estas (número de filas y jerarquías) que pueden contener, significa personalizar el código para cada tabla.

En relación con aspectos técnicos para el desarrollo se ha utilizado el intérprete Python en un entorno Conda sobre Linux y la API de OpenAI (modelo 4o-mini) utilizando librerías Python como Langchain, sqlite3, FAISS, FASTA-PI, uvicorn, entre otras. El modelo ha sido implementado con la técnica de RAG y utiliza FAISS (*open source*) para el almacenamiento y la consulta de los *embeddings* generados, y como modelo para estos se ha utilizado *Text-embedding-3-small* de 1536 dimensiones y con un máximo de *tokens* entrada de 8191.

Para mejorar las prestaciones, la arquitectura propuesta utiliza un servidor de memoria(FATSAPI) que permite almacenar las preguntas y respuestas por identificador de sesión, permitiendo obtener mejores velocidades de tiempo de respuesta (ver Fig. 2). Las respuestas que proporciona el chatbot también muestran imágenes de tubos y contenedores necesarios para las muestras, estas imágenes son obtenidas de una base de datos externa.

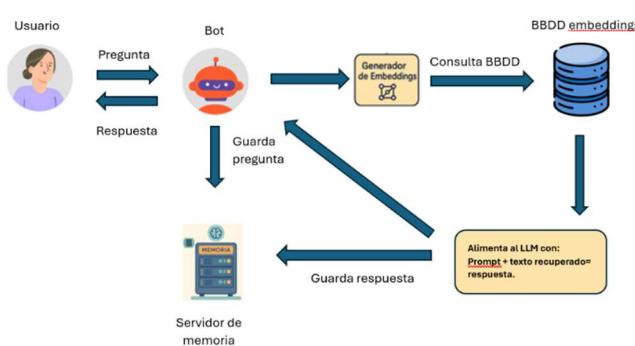


Fig. 2 Esquema de funcionamiento del bot.

En la fase actual del desarrollo y validación, se están realizando pruebas piloto para validar su desempeño y adecuarlo al catálogo disponible de pruebas para minimizar las respuestas ‘inventadas’ (*alucinaciones*) que sufren estos modelos. La reducción de estas respuestas tiene como base la estructura del catálogo en formato JSON que permite, como se indicó anteriormente, realizar las búsquedas con más precisión por parte del LLM.

Dentro de diseño propuesto se ha desarrollado e implementado una versión preliminar funcional con voz (*speech-to-text -S2T-* y *text-to-speech -T2S-*) con la finalidad de permitir establecer una conversación con los usuarios, utilizando las librerías como *whisper* de OpenAI (S2T) y Google Voice (T2S). [2]

Una vez validada la versión final y realizado su despliegue en preproducción, se someterá a una validación por especialistas, técnicos de laboratorio, personal de enfermería y personal administrativo. En función de la evolución del presente proyecto, se ha planificado desarrollar una versión para pacientes con un catálogo con menor información sobre las pruebas, ya que no debe dar más

información que la necesaria para evitar confusiones o malas interpretaciones.

Este desarrollo se ha valorado como Dispositivo Médico Clase 1 (*Medical Device Software -MDSW- clase 1*) por el tipo de información que gestiona el chabot y que en caso de fallo se produciría una incidencia de proceso cuya solución sería solicitar nueva muestra. [14]

5 RESULTADOS

El modelo basado en RAG con un catálogo reducido (sobre formato JSON) está siendo validado en el laboratorio del Hospital de Mar y el Laboratorio Core de Mas Blau tanto por personal administrativo como por técnicos de laboratorio. Las pruebas preliminares el diseño de la interfaz web ha sido valorada satisfactoriamente por el grupo de usuarios consultados clasificándola de “agrable, sencilla y comprensible” (Fig. 3).



Fig. 3 Interfaz gráfica del chatbot.

La aceptación ha sido mayoritaria por parte de todo el personal implicado y ha permitido recoger sugerencias de mejora del chabot, así también respuestas erróneas que se hayan podido obtener de éste.

En las primeras estadísticas de uso, el 80% de los usuarios encuentran la herramienta ‘útil’ y solo un 30% de los usuarios ha tenido problemas al hacer las preguntas.

Con estos resultados preliminares y las sugerencias recibidas se ha planificado realizar una segunda versión que incluya más pruebas de catálogo y que solucione los problemas detectados.

6 DISCUSIÓN

Los chatbots aplicados al sector sanitario son herramientas muy útiles y en constante evolución, tanto en los servicios ofrecidos, como en las tecnologías utilizadas (LLM). [15]

El objetivo principal de los chatbots debe ser garantizar una experiencia agradable y accesible para todos los usuarios con la finalidad que los chatbots puedan utilizarse en todo su potencial. En una revisión sistemática de chatbots aplicados al sector sanitario [15], se identificaron una serie de sugerencias y las cuales se han incluido en el desarrollo del presente chabot:

- Interacción fluida y natural: para abordar mejor la forma en que interactúan los usuarios que no están familiarizados con la tecnología.
- Interfaz de usuario: ya sea basada en la web o aplicación móvil debe estar centrada en el usuario, ser simple y con la información estrictamente necesaria utilizando todas las tecnologías de asistencia.
- Flexibilidad: en cuanto al tipo de preguntas que el usuario puede hacer y por lo tanto el chatbot puede interpretar.
- Lenguaje simple y sencillo: con palabras cortas de uso frecuente y oraciones cortas.
- Personalización: la interfaz debe ofrecer una interacción multimodal lo más cercana posible a las necesidades del usuario (incluido el idioma hablado). [16]

Para dar respuesta a todas estas premisas se desarrolló una versión piloto basada en la técnica del *prompt*, pero presentó dificultades ya que el usuario debe estar registrado en OpenAI. Es por lo que se decidió implementar una versión a través de la API y basada en RAG, que puede ser ejecutada en una plataforma web a la que se puede acceder sin necesidad de registros previos, siendo compatible con múltiples dispositivos (ordenadores, tabletas y móviles).

Esta última versión desarrollada basada en RAG y en validación por expertos ha tenido como premisa de diseño delimitar exhaustivamente la fuente de conocimiento así también como dar más flexibilidad al lenguaje empleado. En la presente versión, el idioma de la respuesta coincide con el idioma en que se realiza la consulta, pero debido a la estructura del catálogo se ha detectado que en determinadas ocasiones el modelo tiene problemas al dar las respuestas (combina parte del texto en español y catalán). Para solucionar este problema, se está valorando poder identificar primero el idioma con librerías externas (*langdetect*, *aterium*) para así poder forzar a que el modelo responda en el idioma que identifican dichas librerías (se dispone de un chabot orientado a instrucciones de laboratorio que utiliza estas librerías y que presentan buenos resultados ya que las respuestas son acordes al idioma

que se le realiza la pregunta).

Como reflexión final se puede afirmar que en el ámbito de la medicina y especialmente en el de los laboratorios, los chatbots tienen el potencial de crecimiento ya que permiten mejorar la automatización de las respuestas a las consultas rutinarias relacionadas con las pruebas de laboratorio, así como de facilitar la realización de pruebas de laboratorio y ayudas en la interpretación de resultados analíticos. Además, pueden proporcionar sugerencias de pruebas analíticas según el diagnóstico del paciente complementando las ayudas de otras herramientas y proporcionando información como un CDS (sistema de ayuda en la decisión clínica). Es importante que todas estas herramientas sean validadas por expertos y se realicen todos los trámites y acciones necesarias para cumplir con normativa CE.

7 CONCLUSIONES

La implementación de un chatbot en la fase preanalítica del laboratorio clínico ayuda a reducir las incidencias causadas por carencia de información y permite garantizar que los pacientes y el personal médico tengan acceso inmediato a instrucciones claras y precisas, mejorando así la calidad de las muestras y la eficiencia del laboratorio. También permite automatizar y optimizar la comunicación, y así liberar el personal para centrarse en actividades que requieren intervención humana, elevando el nivel general de servicio y satisfacción de los usuarios.

La presente versión de chabot desarrollado cumple con las premisas prefijadas, está siendo evaluada por expertos de dos laboratorios y las encuestas preliminares a usuarios demuestran que el 80% de los usuarios encuentran la herramienta 'útil' y solo un 30% de los usuarios tiene dificultades para hacer las preguntas y obtener la respuesta adecuada.

Como trabajo futuro se pueden identificar otras áreas del laboratorio clínico que se pueden beneficiar de este desarrollo incluyendo las áreas analíticas y post-analíticas. Versiones futuras también se podría incluir información de los PNT'S (procedimientos y normas técnicas) correspondientes a cada laboratorio de la red de laboratorios.

AGRADECIMIENTOS

Mis Agradecimientos a los Dr. Óscar Díaz y Bonaventura Ruiz (DIBI) por su apoyo, al Dr. Remo Suppi (UAB) por su orientación, al Dr. Felipe Lumbreras (CSIC) por toda la ayuda prestada y a mi hija Sabina acicate de mi vida.

REFERENCIAS

- [1] O. Caelen, M. A. Blete, "Desarrollo de aplicaciones con GPT-4 y ChatGPT." 2023. Editorial O'Reilly-Anaya
- [2] D. Foster, "Deep Learning generativo" 2023. Editorial

- O'Reilly-Anaya.
- [3] W.Mckinney, "Python para análisis de datos" 2023. Editorial O'Reilly-Anaya.
- [4] C. Huyen, "Diseño de sistemas de Machine Learning" 2023. Editorial Marcombo
- [5] C. Pineda, "Aprendizaje automático y profundo en Python" 2022. Editorial Ra-Ma.
- [6] A. Géron, "Aprende Machine Learning con Scikit-Learn, Keras y TensorFlow 3era edición 2023, editorial O'Reilly-Anaya
- [7] J. Ortega, "Big Data, machine lerning y data science en Python" 2023, editorial Ra-Ma
- [8] S. Raschka, Y. Liu, V. Mirjalili "Machine Learning con Pythorch y Scikit-Learn" 2023, editorial Marcombo
- [9] J. Torres, "Introducción al aprendizaje por refuerzo profundo" 2021, editorial Watch this Space
- [10] W. Ramírez, C. Ramírez, "Programación de inteligencia artificial" 2023, editorial Ra-Ma
- [11] J. Bobadilla, "Machine Learning y Deep Learning" 2021, editorial Ra-Ma
- [12] S. Aussó, C. Raventós, V. Gómez, A. Berenguer, J. Aznar, et al. Guia per a la qualificació i classificació d'una aplicació informàtica basada en IA, https://ticsalutsocial.cat/wp-content/uploads/2025/03/Qualificacio_Classificacio_MDSW_Guia-CAT-1.pdf, 2025
- [13] S. Aussó, D. Domínguez, M. Quintana, Guia sobre l'explicabilitat en la Intel·ligència Artificial, <https://ticsalutsocial.cat/wp-content/uploads/2025/02/230226-Informe-Explicabilitat.pdf>, 2022
- [14] S. Aussó, C. Raventós, V. Gómez, A. Berenguer, J. Aznar, et al, Guia per a la qualificació i classificació d'una aplicació informàtica basada en IA, https://ticsalutsocial.cat/wp-content/uploads/2025/03/Qualificacio_Classificacio_MDSW_Guia-CAT-1.pdf, 2025
- [15] E. Grassini, M. Buzzi, B. Leporini, A. Vozna, A systematic review of chatbots in inclusive healthcare: insights from the last 5 years, https://link.springer.com/article/10.1007/s10209-024-01118-x?utm_source=chatgpt.com 2024
- [16] S. Yang, F. Wang , M. Greenblatt , S. Huang , Y. Zhang, Chatbots de IA en medicina de laboratorio clínico: fundamentos y tendencias, Química Clínica, volumen 69, número 11, noviembre de 2023, páginas 1238–1246
- [17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones et al, Attention Is All You Need, <https://arxiv.org/abs/1706.03762>, 2017