



# Desarrollo de un algoritmo predictivo para la personalización de estilos de vida basado en genómica y biométrica

Maximiliano Hernandez S. y David Fernández Martín

Máster Universitario en Bioinformática

Universidad Internacional de La Rioja

# Índice

1 Introducción

2 Marco teórico

3 Metodología

4 Discusión

5 Conclusiones

6 Bibliografía

# 1. Introducción

## Motivación

Los avances recientes en inteligencia artificial han permitido su aplicación en la personalización de la salud, lo que ha generado una gran motivación para este proyecto.

## Problema a Resolver

La falta de recomendaciones personalizadas basadas en datos genómicos y biométricos.



# 1. Introducción

- 1 Diseño e implementación de la arquitectura
- 2 Búsqueda de información
- 3 Aplicar algoritmos al chatbot
- 4 Evaluar el desempeño
- 5 Analizar los resultados



## 2. Marco teórico



Comprender y generar textos de apariencia humana



Evaluación comparativa del rendimiento de los LLM en tareas médicas



Extraer información de las historias clínicas electrónicas y los datos de los pacientes



LLM especializado en medicina: Med-PaLM 2

## 2. Marco teórico

**Precisión de la información**

**Lenguaje médico complejo**

**Colaboración con Profesionales de  
la Salud**

**Monitoreo continuo y mejoras**

**Aceptación y confianza de los  
usuarios**

**Personalización y Relevancia**

## 3. Metodología

# Arquitectura del Sistema

**main.py**

Configuración del modelo y  
carga de documentos.

**custom\_agent.py**

Definición del agente  
personalizado para manejar  
consultas.

**app.py**

Interfaz de usuario con  
Streamlit.

# Configuración Inicial y Librerías (main.py)

1

### Importación de librerías esenciales

Se importan las librerías necesarias para el funcionamiento del sistema.

2

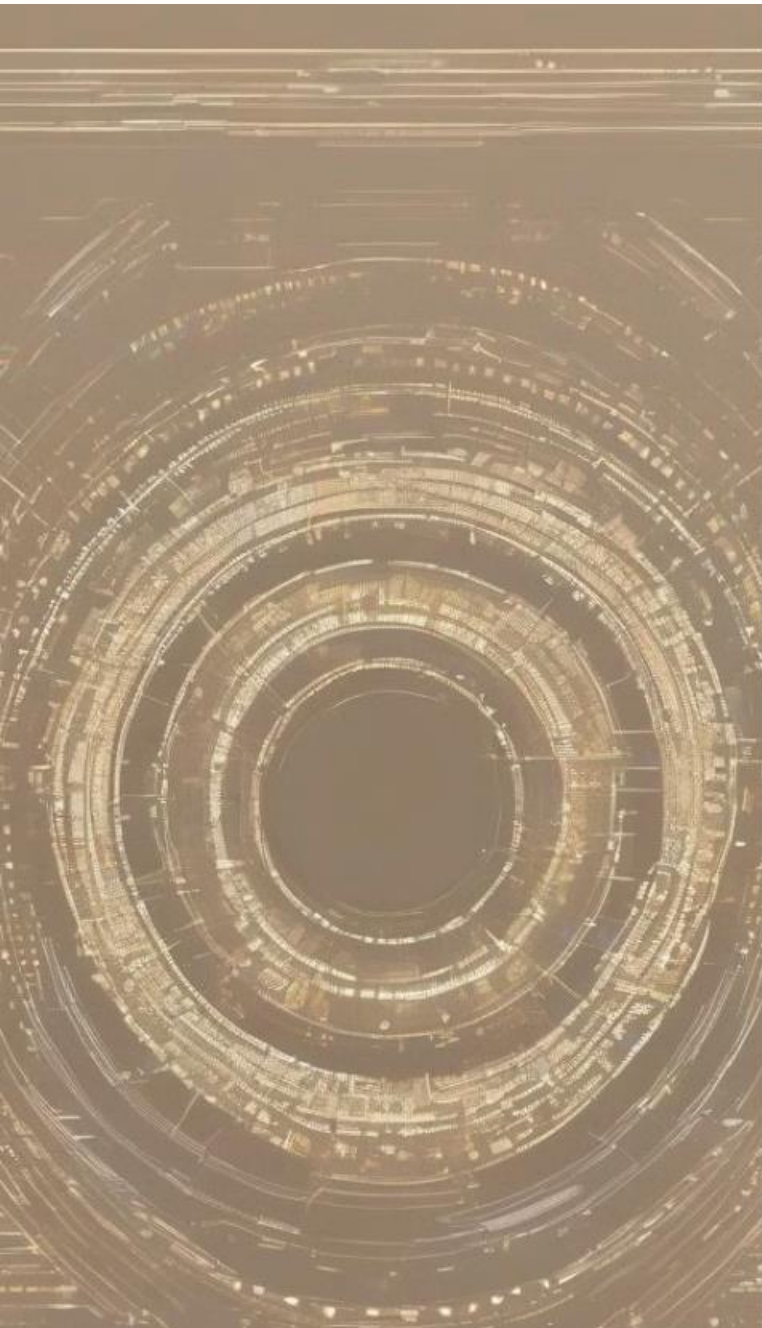
### Configuración de variables de entorno y caché

Se establecen las variables de entorno y se configura el uso de caché para mejorar la eficiencia.

3

### Inicialización de FastAPI

Se inicializa FastAPI para crear la API del servicio.





# Configuración del Modelo y Embeddings (main.py)



### Carga de la clave API

Se carga la clave API necesaria para el modelo de lenguaje ChatGroq.



### Preparación de embeddings

Se construye un índice vectorial usando FAISS para los embeddings del modelo.

# 3. Metodología

## Funciones Clave y Endpoints (main.py)

1

### Endpoint /ask

Se define el endpoint /ask para manejar las preguntas de los usuarios.

2

### Uso de caché

Se utiliza caché para almacenar las respuestas y mejorar la eficiencia.

3

### Carga y procesamiento de documentos

Se cargan y procesan documentos desde diversas fuentes para alimentar el sistema.



# 3. Metodología

## Tratamiento de pdf (documentos médicos de libre distribución)

1

**Función** `load_and_split_pdf`:

Carga y divide el PDF en fragmentos manejables.

2

**Función** `vectorize_and_store`:

Vectoriza los fragmentos de los documentos.

3

**Función** `load_documents_and_vectors`:

Carga y vectoriza el PDF una sola vez, cacheando los vectores.

4

**Evento** `startup`: Llama a

`load_documents_and_vectors` para cargar y vectorizar el libro médico al iniciar la aplicación

Esto asegura que el pdf, aún de longitud elevada, se procese eficientemente y esté disponible para consultas rápidas sin necesidad de recargarlo repetidamente.

## 3. Metodología



### Definición del Agente Personalizado (custom\_agent.py)

#### Clase CustomAgent

Se define la clase CustomAgent que hereda de LLMSingleActionAgent.

#### Métodos input\_keys y plan

Se implementan los métodos input\_keys y plan para planificar acciones basadas en pasos intermedios.

### 3. Metodología

## Parser de Salida Personalizado (custom\_agent.py)

1

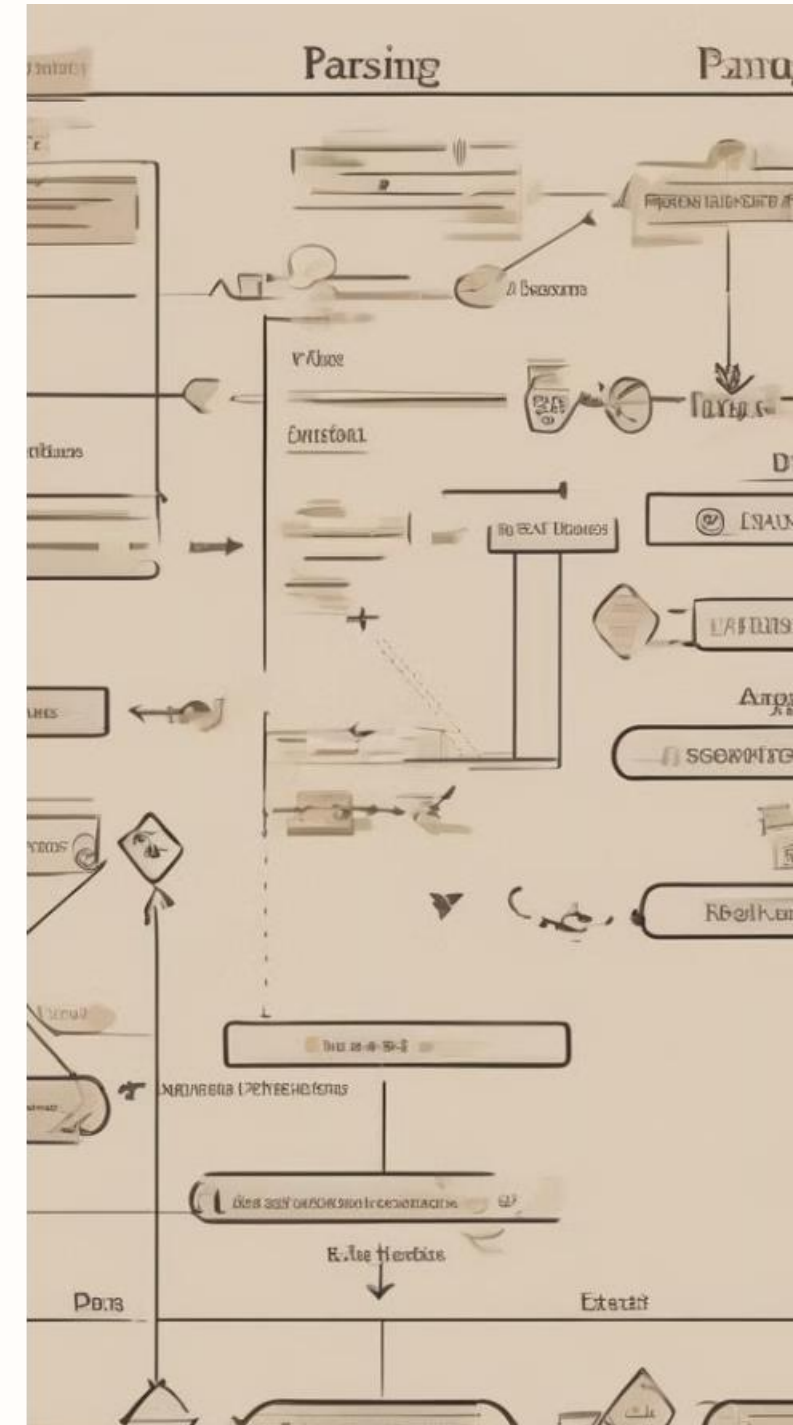
## CustomOutputParser

Se define la clase CustomOutputParser que analiza la salida del modelo de lenguaje.

2

## Decidir siguiente acción

Se implementan métodos para decidir si la respuesta es final o si se necesitan más acciones.





# 3. Metodología

## Creación del Agente y Ejecutor (custom\_agent.py)



### `create_custom_tools_agent`

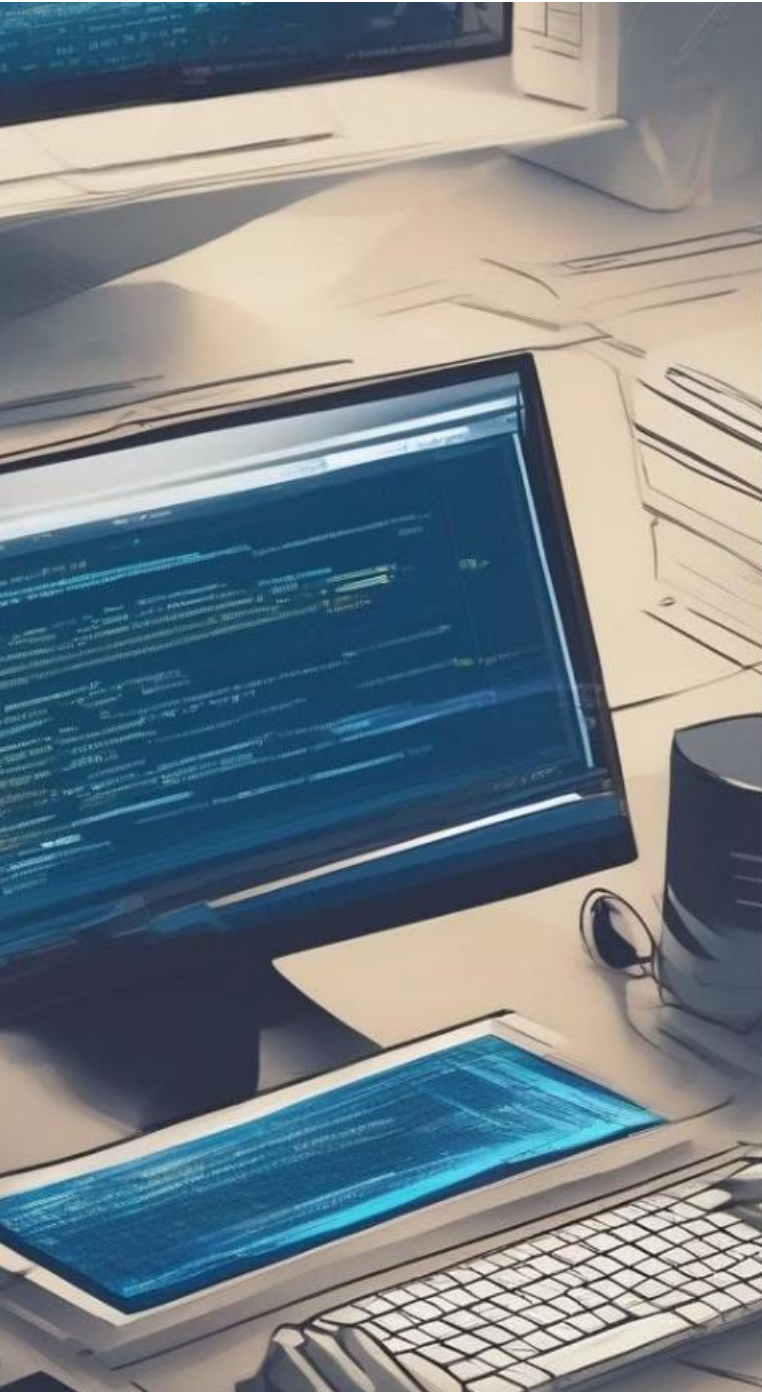
Se define la función para configurar el agente personalizado, integrando el modelo de lenguaje con herramientas externas.



### Preparación de prompts

Se preparan los prompts y se configura el ejecutor para el agente personalizado.

# 3. Metodología



## Nuevas Funcionalidades Implementadas

1

### Web Scraping con BeautifulSoup

Se añade la capacidad de hacer scraping en sitios web especializados como la Clínica Mayo para obtener información adicional sobre enfermedades.

2

### API de ClinVar

Se integra la API de ClinVar para obtener información genética específica sobre enfermedades.

3

### Traducción Automática

Las consultas se traducen al inglés antes de realizar búsquedas y se devuelven las respuestas traducidas al idioma original del usuario.



### 3. Metodología



## Interfaz de Usuario con Streamlit (app.py)

Configuración de la interfaz

Se configura la interfaz de usuario utilizando Streamlit.

Entrada de texto

Se agrega un campo de entrada de texto para que el usuario ingrese preguntas.

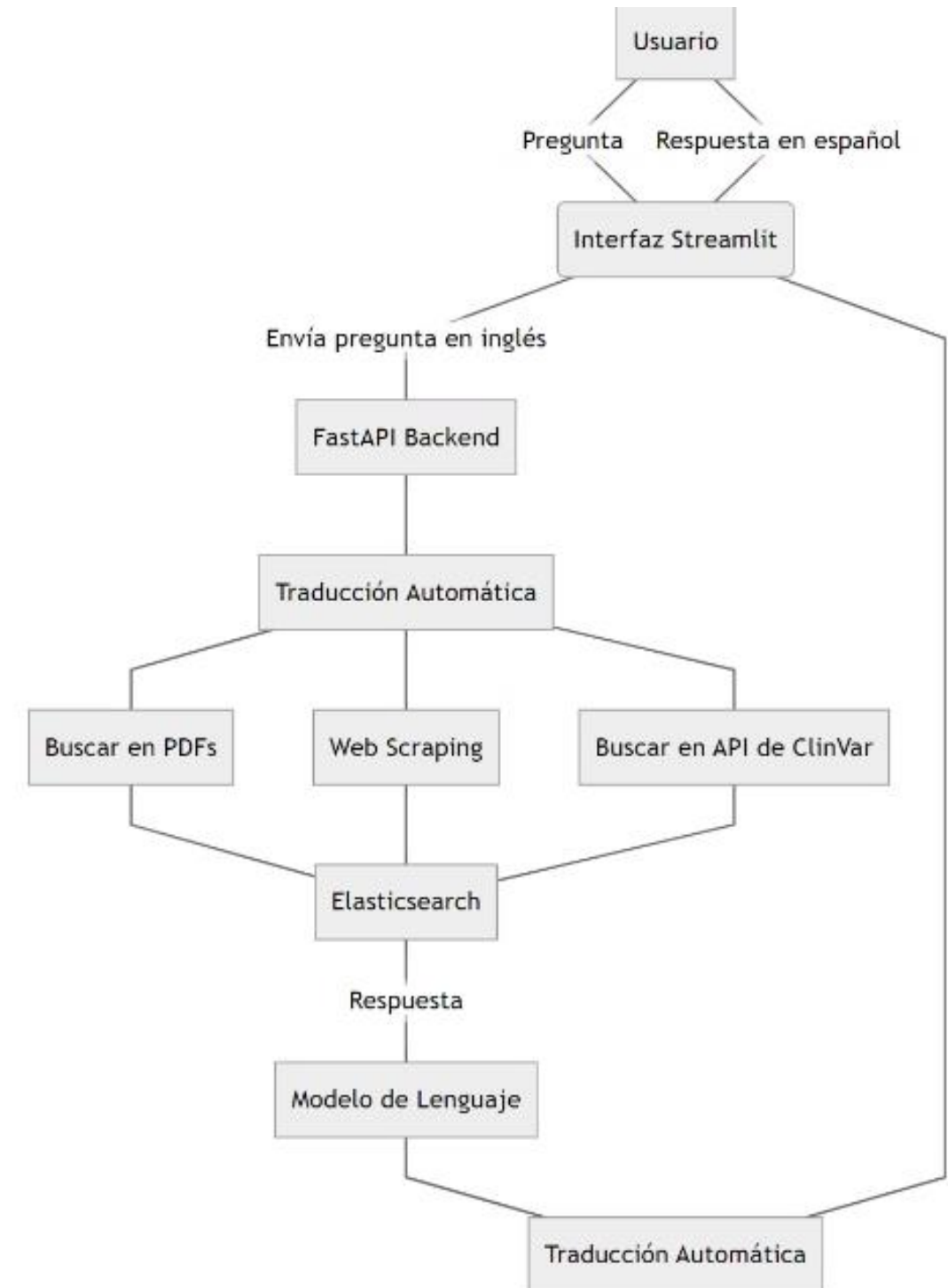
Botón de envío

Se incluye un botón para que el usuario pueda enviar las preguntas al backend.

# 3. Metodología

## Diagrama de funcionamiento

La interfaz de usuario se configura utilizando Streamlit, una biblioteca de Python que permite crear aplicaciones web interactivas de manera sencilla. La interfaz incluye un campo de entrada de texto para que el usuario pueda ingresar preguntas, así como un botón de envío para que pueda enviar esas preguntas al backend.



# 3. Metodología



## Características Clave del Sistema RAG

1

### Recopilación de Datos

El sistema RAG recopilará datos genómicos y biométricos de los usuarios a través de dispositivos y plataformas integradas.

2

### Análisis de Datos

Los datos recopilados serán analizados utilizando técnicas de inteligencia artificial y aprendizaje automático para identificar patrones y correlaciones.

3

### Generación de Recomendaciones

El algoritmo predictivo del sistema utilizará los resultados del análisis de datos para generar recomendaciones personalizadas de estilo de vida.



## 3. Metodología

# Beneficios del Sistema RAG

### Mejora de la Salud

Las recomendaciones personalizadas del sistema RAG ayudarán a los usuarios a adoptar estilos de vida más saludables, lo que se traducirá en una mejor salud y bienestar general.

### Prevención de Enfermedades

Al tener en cuenta la predisposición genética, el sistema podrá recomendar estrategias preventivas para reducir el riesgo de desarrollar ciertas enfermedades.

### Empoderamiento del Usuario

Los usuarios tendrán un mayor control y conocimiento sobre su salud, lo que les permitirá tomar decisiones más informadas y participar activamente en su bienestar.

### Eficiencia en la Atención Médica

El sistema RAG puede ayudar a los profesionales de la salud a brindar atención más personalizada y efectiva, optimizando los recursos y reduciendo los costos.

# Futuro del Sistema RAG

### 1 Expansión a Nuevas Áreas

El sistema RAG tiene el potencial de expandirse a otras áreas de la salud y el bienestar, como la prevención de enfermedades crónicas y la optimización del rendimiento deportivo.

### 3 Avances en Inteligencia Artificial

Los avances en técnicas de inteligencia artificial y aprendizaje automático mejorarán la precisión y la eficacia del algoritmo predictivo del sistema RAG.

1

2

### Integración con Dispositivos Wearables

La integración con dispositivos wearables y aplicaciones móviles permitirá una recopilación de datos más precisa y en tiempo real, mejorando la calidad de las recomendaciones.

3

## 4. Discusión

Seleccione el tipo de pregunta que desea simular:

Preguntas simples (Wikipedia)



Seleccione una pregunta para mostrar la respuesta:

¿Qué es la diabetes tipo 2?



**Respuesta:** La diabetes tipo 2 es una afección en la que el cuerpo no usa la insulina adecuadamente, lo que causa niveles altos de glucosa en sangre. Es común en adultos y se puede controlar con dieta, ejercicio y medicamentos.

## 4. Discusión

Seleccione el tipo de pregunta que desea simular:

Preguntas complejas (Mayo Clinic)



Seleccione una pregunta para mostrar la respuesta:

¿Qué es la diabetes tipo 2?



**Respuesta:** La diabetes mellitus (DM) es un trastorno metabólico crónico caracterizado por una hiperglucemia persistente. Puede deberse a una secreción deficiente de insulina, a una resistencia a las acciones periféricas de la insulina o a ambas. Según la Federación Internacional de Diabetes (IDF), aproximadamente 415 millones de adultos de entre 20 y 79 años padecían diabetes mellitus en 2015. La DM está demostrando ser una carga para la salud pública mundial, ya que se prevé que esta cifra aumente a otros 200 millones en 2040. La hiperglucemia crónica, en sinergia con otras aberraciones metabólicas en pacientes con diabetes mellitus, puede causar daños en varios sistemas orgánicos, lo que conduce al desarrollo de complicaciones de salud incapacitantes y potencialmente mortales, de las cuales las más destacadas son las complicaciones microvasculares (retinopatía, nefropatía y neuropatía) y macrovasculares que conducen a un riesgo de 2 a 4 veces mayor de enfermedades cardiovasculares. En esta revisión se ofrece una visión general de la patogénesis, el diagnóstico, la presentación clínica y los principios de tratamiento de la diabetes.

## 4. Discusión

### Simulación de preguntas complejas en Mayo Clinic

Diabetes tipo 2

1.30x m...

↑ 1.30x

Ataque al corazón

1.80x m...

↑ 1.80x

Hipertensión

2.20x m...

↑ 2.20x

Cáncer de pulmón

1.90x m...

↑ 1.90x



## 4. Discusión



Mejora el compromiso y la educación de los pacientes



Reduce la carga administrativa de los médicos



Mejora la monitorización remota del paciente

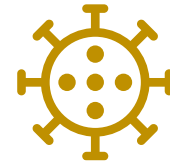


Fomenta la vida asistida en el entorno de las personas mayores

## 4. Discusión



Calidad de las fuentes de información



Falta de integración con los registros médicos



Interpretación de dudas ambiguas



Consideraciones éticas y legales

## 4. Discusión

Integrar más  
fuentes de datos

Personalización  
y adaptación al  
usuario

Integración con  
el historial  
clínico  
electrónico

Evaluación  
clínica y  
validación

Mejora  
continua  
basada en los  
comentarios de  
los usuarios

# 5. Conclusiones

1

Desarrollo de un chatbot medico especializado

2

Aplicación en el manejo automático de hallazgos secundarios en estudios

3

Integración de multiples fuentes de información médica

4

Evaluación exhaustiva

5

Identificación de limitaciones y desafíos

6

Expansión de la base de conocimientos

7

Integración de técnicas de aprendizaje por refuerzo

8

Exploración de enfoques multimodales

9

Realización de estudios clínicos a gran escala

10

Colaboración interdisciplinaria

# 6. Bibliografía

1. Egli A. ChatGPT, GPT-4, and Other Large Language Models: The Next Revolution for Clinical Microbiology? Clin Infect Dis. 11 de noviembre de 2023;77(9):1322-8.
2. Singhal K, Tu T, Gottweis J, Sayres R, Wulczyn E, Hou L, et al. Towards Expert-Level Medical Question Answering with Large Language Models [Internet]. arXiv; 2023 [citado 1 de junio de 2024]. Disponible en: <http://arxiv.org/abs/2305.09617>
3. Mökander J, Schuett J, Kirk HR, Floridi L. Auditing large language models: a three-layered approach. AI Ethics [Internet]. 30 de mayo de 2023 [citado 31 de marzo de 2024]; Disponible en: <https://link.springer.com/10.1007/s43681-023-00289-2>
4. Yu B. Evaluating Pre-Trained Language Models on Multi-Document Summarization for Literature Reviews.
5. Agapito G, Calabrese B, Guzzi PH, Cannataro M, Simeoni M, Care I, et al. DIETOS: A recommender system for adaptive diet monitoring and personalized food suggestion. En: 2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) [Internet]. New York, NY: IEEE; 2016 [citado 1 de abril de 2024]. p. 1-8. Disponible en: <http://ieeexplore.ieee.org/document/7763190/>
6. Iwendi C, Khan S, Anajemba JH, Bashir AK, Noor F. Realizing an Efficient IoMT-Assisted Patient Diet Recommendation System Through Machine Learning Model. IEEE Access. 2020;8:28462-74.
7. Open medical-LLM leaderboard - a hugging face space by openlifescienceai [Internet]. [citado 22 de junio de 2024]. Disponible en: [https://huggingface.co/spaces/openlifescienceai/open\\_medical\\_llm\\_leaderboard](https://huggingface.co/spaces/openlifescienceai/open_medical_llm_leaderboard)
8. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc. 1 de septiembre de 2018;25(9):1248-58.
9. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. Can J Psychiatry. julio de 2019;64(7):456-64.
10. Mackley MP, Fletcher B, Parker M, Watkins H, Ormondroyd E. Stakeholder views on secondary findings in whole-genome and whole-exome sequencing: a systematic review of quantitative and qualitative studies. Genet Med. marzo de 2017;19(3):283-93.
11. Levy KD, Blake K, Fletcher-Hoppe C, Franciosi J, Goto D, Hicks JK, et al. Opportunities to implement a sustainable genomic medicine program: lessons learned from the IGNITE Network. Genet Med. marzo de 2019;21(3):743-7.
12. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. mayo de 2015;17(5):405-24.
13. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 4 de enero de 2018;46(D1):D1062-7.
14. FastAPI - A modern, fast (high-performance), web framework for building APIs with Python 3.6+ based on standard Python type hints. [Internet]. [citado 1 de junio de 2024]. Disponible en: <https://fastapi.tiangolo.com>
15. Streamlit • A faster way to build and share data apps [Internet]. [citado 1 de junio de 2024]. Disponible en: <https://streamlit.io>
16. Groq builds the world’s fastest AI inference technology [Internet]. [citado 1 de junio de 2024]. Disponible en: <https://groq.com>



# 6. Bibliografía

17. Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media, Inc.
18. Robertson, S. (2004). Understanding inverse document frequency: on theoretical arguments for IDF. Journal of documentation, 60(5), 503-520.
19. DisGeNET - a database of gene-disease associations [Internet]. [citado 1 de junio de 2024]. Disponible en: <https://www.disgenet.org/>
20. Welcome to flask — flask documentation (3.0.X) [Internet]. [citado 1 de junio de 2024]. Disponible en: <https://flask.palletsprojects.com/en/3.0.x/>
21. Gobierno de Navarra. (2021). 1300 secuenciaciones de 2000 pacientes de Navarra refuerzan la estrategia de Medicina Personalizada. [citado 17 de junio de 2024]. Disponible en: <https://www.navarra.es/es/noticias/2021/10/07/1300-secuenciaciones-de-2000-pacientes-de-navarra-refuerzan-la-estrategia-de-medicina-personalizada>
22. Chow JCL, Sanders L, Li K. Design of an Educational Chatbot Using Artificial Intelligence in Radiotherapy. AI. 2 de marzo de 2023;4(1):319-32.
23. Yasunaga M, Leskovec J, Liang P. LinkBERT: Pretraining Language Models with Document Links [Internet]. arXiv; 2022 [citado 2 de junio de 2024]. Disponible en: <http://arxiv.org/abs/2203.15827>
24. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT - Reshaping medical education and clinical management. Pak J Med Sci [Internet]. 7 de febrero de 2023 [citado 2 de junio de 2024];39(2). Disponible en: <https://pjms.org.pk/index.php/pjms/article/view/7653>
25. Zúñiga Salazar G, Zúñiga D, Vindel CL, Yoong AM, Hincapie S, Zúñiga AB, et al. Efficacy of AI Chats to Determine an Emergency: A Comparison Between OpenAI's ChatGPT, Google Bard, and Microsoft Bing AI Chat. Cureus [Internet]. 18 de septiembre de 2023 [citado 2 de junio de 2024]; Disponible en: <https://www.cureus.com/articles/183679-efficacy-of-ai-chats-to-determine-an-emergency-a-comparison-between-openais-chatgpt-google-bard-and-microsoft-bing-ai-chat>
26. Morse S, Hagen J. Talkdesk talks about the benefits of being an early adopter of ChatGPT [Internet]. 2023 [citado 5 de junio de 2024]. Disponible en: <https://www.healthcareitnews.com/news/talkdesk-talks-about-benefits-being-early-adopter-chatgpt>
27. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. En: Artificial Intelligence in Healthcare [Internet]. Elsevier; 2020 [citado 5 de junio de 2024]. p. 25-60. Disponible en: <https://linkinghub.elsevier.com/retrieve/pii/B9780128184387000022>
28. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthc J. julio de 2021;8(2):188-94.
29. Pros & cons of artificial intelligence in medicine [Internet]. 2024 [citado 5 de junio de 2024]. Disponible en: <https://drexel.edu/ccj/stories/artificial-intelligence-in-medicine-pros-and-cons/>
30. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med. enero de 2019;25(1):30-6.
31. Kiseleva A, Kotzinos D, De Hert P. Transparency of AI in Healthcare as a Multilayered System of Accountabilities: Between Legal Requirements and Technical Limitations. Front Artif Intell. 30 de mayo de 2022;5:879603.

**Gracias por vuestra atención**