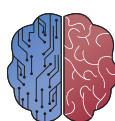




UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería de la Salud



INGENIERÍA
DE LA SALUD

**TFG del Grado en Ingeniería de la
Salud**

**Aplicación de chatbot con
LLM y RAG para la gestión
de información científica de
COVID-19 en PubMed
Documentación Técnica**

Presentado por Daniel de Lara Pérez
en Universidad de Burgos

3 de julio de 2024

Tutores: José Ignacio Santos Martín – Virginia Ahedo
García

Índice general

Índice general	i
Índice de figuras	iii
Índice de tablas	iv
Apéndice A Plan de Proyecto Software	1
A.1. Introducción	1
A.2. Planificación temporal	4
Apéndice B Documentación de usuario	15
B.1. Requisitos software y hardware para ejecutar el proyecto. . .	15
B.2. Instalación / Puesta en marcha	15
B.3. Manuales y/o Demostraciones prácticas	16
Apéndice C Manual del programador.	19
C.1. Estructura de directorios	19
C.2. Compilación, instalación y ejecución del proyecto	19
C.3. Pruebas del sistema	20
C.4. Instrucciones para la modificación o mejora del proyecto. . .	20
Apéndice D Descripción de adquisición y tratamiento de datos	21
D.1. Descripción formal de los datos	21
D.2. Descripción clínica de los datos.	21
Apéndice E Manual de especificación de diseño	23
E.1. Planos	23
E.2. Diseño arquitectónico	23

Apéndice F Especificación de Requisitos	31
F.1. Diagrama de casos de uso	33
F.2. Explicación casos de uso.	34
F.3. Prototipos de interfaz o interacción con el proyecto.	34
Apéndice G Estudio experimental	41
G.1. Cuaderno de trabajo.	41
G.2. Configuración y parametrización de las técnicas.	43
G.3. Detalle de resultados.	45
Apéndice H Anexo de sostenibilización curricular	65
H.1. Introducción	65
H.2. Competencias de sostenibilidad	65
Bibliografía	69

Índice de figuras

A.1. Ilustración del proceso de Scrum	2
A.2. Sprint Inicial	7
A.3. Sprint Prueba Modelos	7
A.4. Sprint creación del primer modelo funcional	7
A.5. Sprint Mejoras	8
A.6. Sprint Marco Teórico	8
A.7. Sprint Anexos	8
A.8. Sprint GUI	9
A.9. Finalización de la documentación	9
A.10. Salario de un Programador Junior en España	10
B.1. Requisitos de GPU	16
B.2. Demostración de aceleración por hardware	17
B.3. Interfaz gráfica de usuario	18
B.4. Interfaz gráfica de usuario tras una ejecución	18
E.1. Diagrama de secuencia del caso de uso número 1	24
E.2. Diagrama de secuencia del caso de uso número 2	26
E.3. Diagrama de secuencia del caso de uso número 3	27
E.4. Diagrama de secuencia del caso de uso número 4	28
E.5. Diagrama de secuencia del caso de uso número 5	28
E.6. Diagrama de componentes del proyecto	29
F.1. Diagrama de casos de uso	34

Índice de tablas

F.1. CU-1 Generar Respuesta.	35
F.2. CU-2 Actualizar el modelo.	36
F.3. CU-3 Modificar la información especializada.	37
F.4. CU-4 Repetir el prompt.	38
F.5. CU-5 Eliminar conversación.	39
G.1. Tabla de puntuaciones para validar un modelo de lenguaje especializado en COVID-19.	62
G.2. Notas de validación del modelo de lenguaje especializado en COVID-19.	63
G.3. Notas de validación del modelo de lenguaje especializado en COVID-19.	63
G.4. Notas de validación del modelo de lenguaje especializado en COVID-19.	63
G.5. Notas de validación del modelo de lenguaje especializado en COVID-19.	64
G.6. Notas de validación del modelo de lenguaje especializado en COVID-19.	64

Apéndice A

Plan de Proyecto Software

A.1. Introducción

La intención de este anexo es la de dar a entender el plan de proyecto que se ha seguido a la hora de realizar este trabajo.

Primero se hablará sobre la planificación temporal seguida y más adelante su viabilidad tanto económica como legal

Las metodologías ágiles (especialmente Scrum) utilizadas en el desarrollo de este proyecto son aquellas que permiten ir adaptando la forma y carga de trabajo a medida que se desarrolla el proyecto, analizando los nuevos retos que pueden surgir durante la realización del mismo lo que otorga flexibilidad al proyecto durante la etapa de desarrollo.[Hayat et al., 2019]

Scrum

Scrum es un marco ligero que ayuda a las personas, equipos y organizaciones a generar valor a través de soluciones adaptables para problemas complejos.

[Ken Schwaber and Jeff Sutherland, 2020]

En la figura A.1 se puede ver una ilustración del proceso de Scrum

Está basado en el empirismo y emplea un enfoque iterativo e incremental para el desarrollo del producto minimizando riesgos ya que en el caso de un desastre simplemente habría que dar marcha atrás hasta la última versión estable.

Esta metodología tiene dos pilares clave, el equipo Scrum y los eventos.

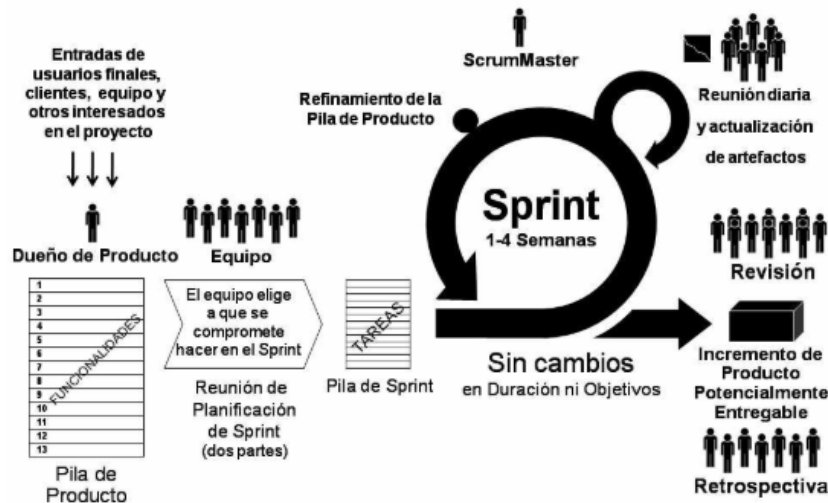


Figura A.1: Ilustración del proceso de Scrum

Equipo Scrum

Capital humano que se encarga de desarrollar el proyecto. Constan de desarrolladores, product owner y Scrum Master.

Desarrolladores

Son los encargados de crear cada incremento al producto en cada sprint.

Sus tareas dentro del ciclo Scrum son la de crear un plan para el sprint llamado Sprint Backlog, Adaptar su plan de desarrollo día a día según los objetivos del sprint y responsabilizarse del trabajo de sus compañeros desarrolladores para garantizar un buen trabajo de todas las partes.

Product owner

El propietario del producto, es el responsable de sacar el máximo valor al trabajo del equipo Scrum así cómo gestionar la Backlog comunicando los objetivos a los que se desea llegar, se asegura de que se alcanza una buena calidad del producto final y gestionando los proyectos pendientes a realizar en el sprint.

Scrum Master

El máximo responsable de cumplir con la guía de Scrum, explicando cuidadosamente su teoría al resto del equipo así como coordinando su puesta en práctica.

Se responsabiliza de la efectividad del equipo Scrum eliminando las barreras entre los miembros, formándolos en la metodología y ayudándolos en sus planificaciones y tareas.

Eventos de Scrum

Los eventos en Scrum son oportunidades señaladas en el tiempo para revisar y adaptar los artefactos de Scrum, se crean para crear una regularidad y minimizar reuniones imprevistas.

El Sprint

Eventos de longitud fija donde ocurre todo el trabajo y en el que se fijan distintos objetivos.

En estos sprints iterativos se buscan realizar cambios que no deterioren el objetivo del sprint ni disminuya su calidad y clarificar los objetivos revaluando las tareas pendientes a medida que se desarrolla el proyecto.

Planificación de Sprint

Este evento inicia el Sprint estableciendo los objetivos a realizar en el mismo, este evento requiere de la participación del equipo de Scrum al completo.

Se pretende responder las siguientes cuestiones:

- ¿Por qué este Sprint es valioso?
- ¿Qué se puede hacer este Sprint?
- ¿Cómo se realizará el trabajo elegido?

Scrum diario

Su propósito es el análisis del avance hacia el objetivo final del Sprint. Es un evento de corta duración para los desarrolladores en los que discuten las diversas tareas del día.

Revisión del Sprint

Este evento que sucede al finalizar un Sprint tiene como propósito analizar este mismo, se revisan los logros conseguidos, los cambios que han tenido lugar y objetivos futuros.

La retrospectiva del Sprint

El objetivo de este último evento es el de aumentar la calidad del proyecto analizando el anterior sprint determinando que fue bien en el mismo, así como dificultades y la resolución o no de éstas.

Este evento concluye el Sprint

Artefactos de Scrum

Son la representación de los objetivos tanto a corto como a largo plazo del proyecto.

Pila del producto

Es una lista ordenada de lo que se necesita para mejorar el producto, los ítems en la pila del producto se consideran listos para su selección en un evento de planificación de Sprint.

Pila del Sprint

El trabajo pendiente por hacer en el Sprint. Se puede entender como el trabajo que aún le queda a los desarrolladores para alcanzar el objetivo de ese sprint.

Incremento

El incremento es cada pequeño objetivo cumplido que aporta un paso más hacia completar el objetivo del Sprint. Cada incremento es aditivo y deben funcionar todos juntos para llegar a la meta final. [Schwaber and Sutherland, 2020]

A.2. Planificación temporal

Para la realización de este trabajo se ha seguido la metodología Scrum mediante la aplicación de Zube.io lo que ha dado lugar a la siguiente planificación mediante Sprints.

Sprint 1: Sprint inicial

En este primer Sprint se pretendía tener una toma de contacto inicial con las herramientas que se iban a desarrollar a lo largo del proyecto, en especial se asignaron dos objetivos principales, familiarizarse con el entorno de Zube así como investigar las tecnologías de RAG y de LLMs.

A.2

Sprint 2: Prueba de modelos

Segundo Sprint del proyecto, el objetivo era investigar los diferentes modelos de LLM, sus respuestas ante datos clínicos y cómo implementarlos, así como sus formas de pago si lo requiriese.

A.3

Sprint 3: Creación del primer modelo funcional

En este Sprint se creó el primer chatbot funcional con la primera implementación del proceso de RAG con los datos finales sacados de PubMed. También se empezó con el desarrollo de la memoria.

A.4

Sprint 4: Mejoras

En este pequeño Sprint se creó el repositorio en GitHub y se realizaron mejoras al código del modelo, eliminando código innecesario.

A.5

Sprint 5: Marco Teórico

Este Sprint tenía como objetivo inicial la creación de una interfaz gráfica web en streamlit, lamentablemente aparecieron multitud de problemas con la GPU permitida y se decidió dejarla en la backlog list. Al final en este sprint se realizó el desarrollo del marco teórico de la memoria.

A.6

Sprint 6: Finalización del Marco Teórico y actualización de Anexos

En este Sprint se finaliza lo empezado en el anterior sprint y se empieza a trabajar en los anexos, sobre todo en el de Planificación y Datos así cómo se investigó el anexo de sostenibilidad.

A.7

Sprint 7: Creación de la interfaz gráfica de usuario, validación del modelo y trabajo en los anexos

El objetivo principal de este sprint fue el de acabar la herramienta añadiendo la interfaz gráfica de usuario, es decir la parte visual que comunicará al usuario con el software del programa para hacerlo más manejable gestionando las entradas y mostrando las salidas, además se empezó a trabajar en la validación del modelo, creando una escala numérica en la que el desarrollador, otros LLMs y personas ajenas al proyecto que cumplen el rol de futuros usuarios, dieron su opinión interactuando con el modelo y puntuándolo en dicha escala numérica. Mientras este trabajo se desarrollaba también se actualizaban los anexos de la documentación del proyecto.

A.8

Sprint 8: Finalización de la documentación

Sprint final en el que se busca finalizar la documentación para presentar el proyecto y actualizar el repositorio para mejorar la profesionalidad del trabajo.

A.9

Sprint board

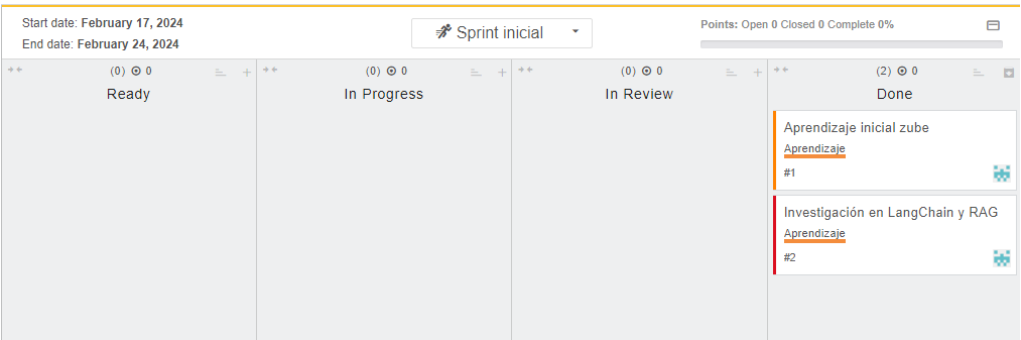


Figura A.2: Sprint Inicial

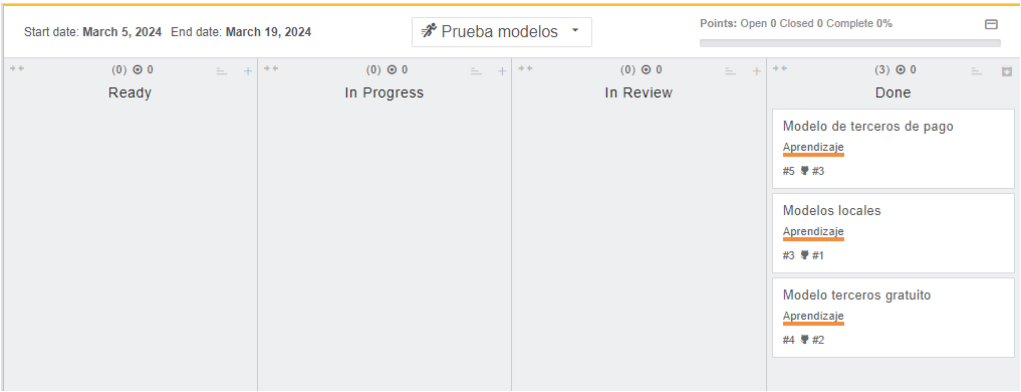


Figura A.3: Sprint Prueba Modelos

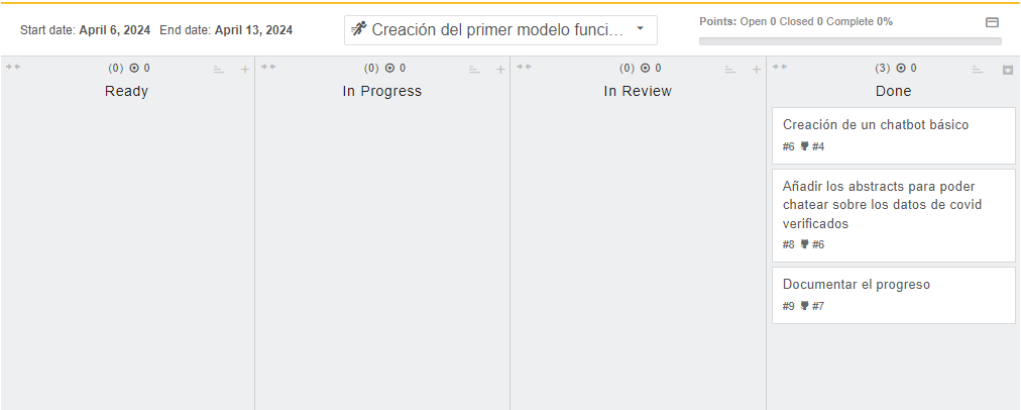


Figura A.4: Sprint creación del primer modelo funcional

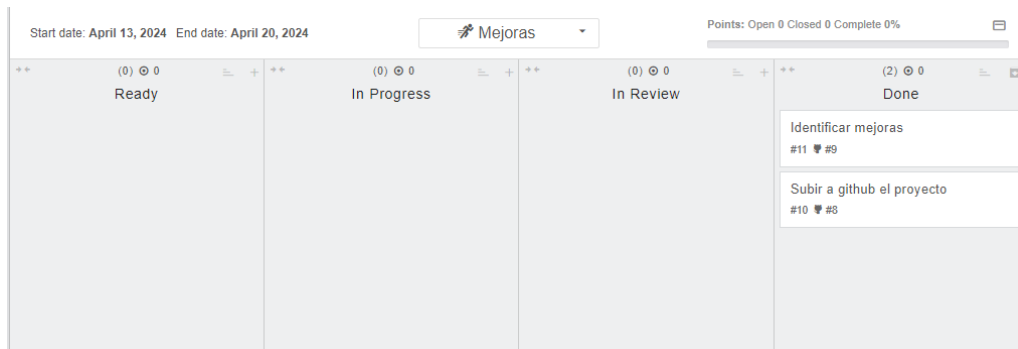


Figura A.5: Sprint Mejoras

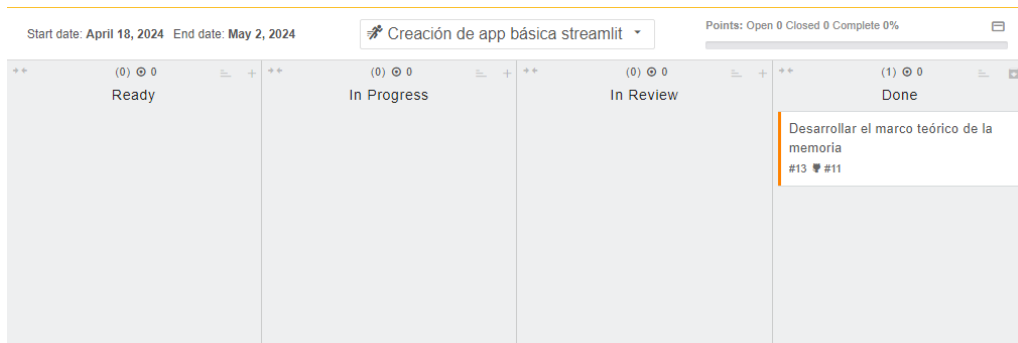


Figura A.6: Sprint Marco Teórico

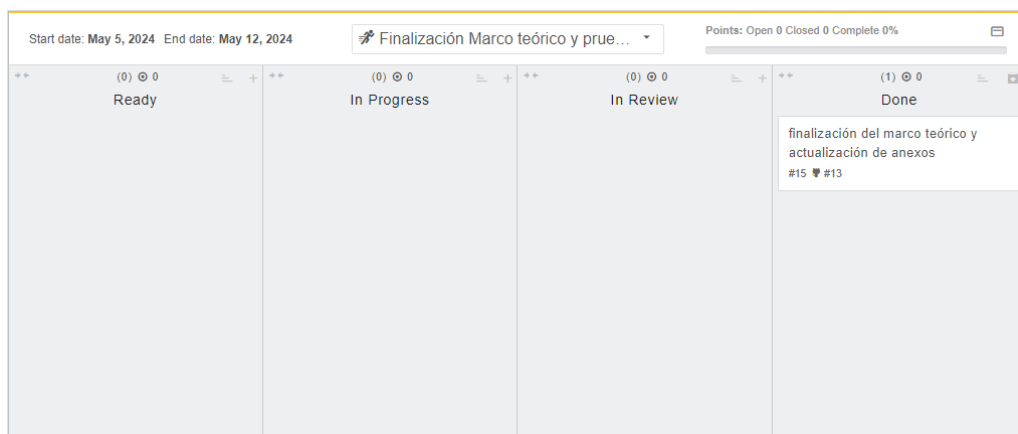


Figura A.7: Sprint Anexos

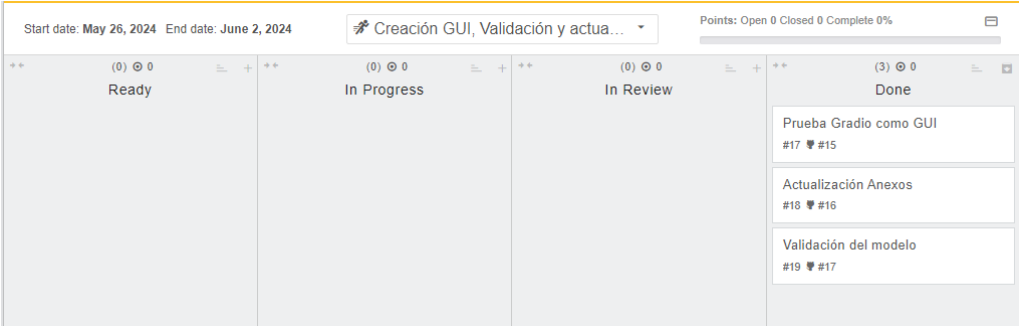


Figura A.8: Sprint GUI

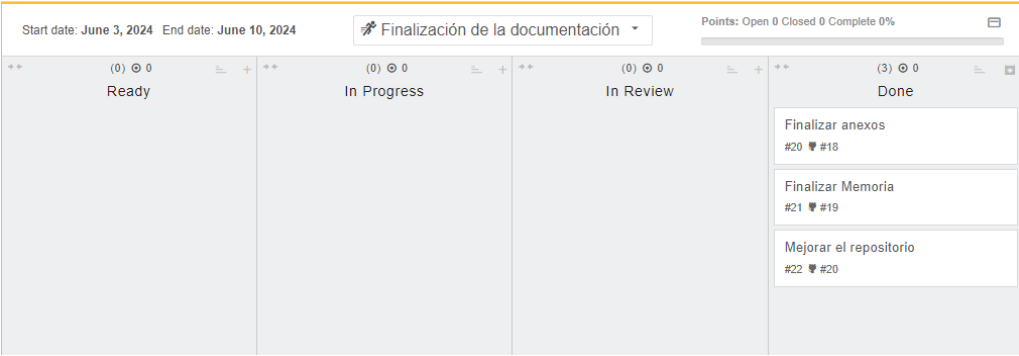


Figura A.9: Finalización de la documentación

Planificación económica

En este apartado se analizarán los costes monetarios de la implementación del proyecto

Librerías

Las librerías empleadas en el desarrollo han sido de código abierto por lo que no suman ningún coste.

Contratación

Se estima que el sueldo promedio de un programador Junior en España sería de 1750 euros brutos al mes a la fecha de realización de este proyecto como indica en la figura A.10, asumiendo que los pluses y complementos así como la parte proporcional de las pagas extra está incluida. [tal, 2024]



Figura A.10: Salario de un Programador Junior en España

También asumimos de manera arbitraria ya que esto cambiará dependiendo de multitud de condiciones un 9% de Retención de IRPF y unas cotizaciones, enfermedad profesional y accidentes de trabajo del 35%.

El resultado del cálculo sería una base de cotización de 1750 euros a los que se le tendrán que sumar los costes de la seguridad social que ascienden hasta los 525 ($0,3 \cdot 1750$) euros calculados teniendo en cuenta lo siguiente:

- 23.60 % contingencias comunes
- 5.5 % tipo general de desempleo para un contrato indefinido
- 0,20 % FOGASA (Fondo de Garantía Salarial)
- 0,70 % formación profesional

El coste de contratar a un programador junior al mes asciende a los 1750 euros + 525 euros = 2275 euros al mes. Si la duración del proyecto ha sido de 3 meses, el coste total de mano de obra es de 2275 euros * 3 meses = 6825 euros. [Bil, 2014]

Web

Si se quisiera desplegar la aplicación Gradio ofrece hosting a través de los spaces de Hugging Face que puede ser gratuito o de pago. Teniendo en cuenta la GPU necesaria para ejecutar los modelos, sería conveniente emplear las opciones de pago que facilitan la aceleración de GPU, Hugging Face ofrece el espacio NVidia T4 small * 4vCPU * 15GB por 0,40 Dólares la hora, declarándose "dormidos" tras una hora de inactividad, este tiempo no se cobra.

[Hug, 2016]

Amortización de equipos

El valor total de adquisición de los equipos informáticos utilizados es de XXX euros. Suponiendo un periodo de amortización de 5 años, los costes de amortización para la duración de este proyecto de 3 meses son $(XXX/5) \cdot (3/12) = YYY$

Viabilidad legal

Licencias

Las librerías empleadas son Open Source.

Las licencias de las librerías empleadas son las siguientes:

- Licencia de Python Software Foundation
- Licencia BSD modificada
- Licencia Apache 2.0
- Licencia MIT

El código publicado en el repositorio está licenciado con licencia MIT.

Legislación

Según la Ley de inteligencia artificial publicada el 13 de marzo de 2024 las inteligencias artificiales se pueden clasificar en cuatro grupos según su riesgo:

- Riesgo inaceptable
- Riesgo alto
- Riesgo limitado
- Riesgo mínimo o nulo

Cuando una inteligencia artificial presente una clara amenaza para la seguridad y/o los derechos de las personas estas estarán catalogadas como de riesgo inaceptable y, por lo tanto, prohibidas.

Las IAs que se empleen en infraestructuras críticas, formación profesional, componentes de seguridad, gestión de empleo, servicios públicos y privados esenciales, administración de justicia o gestión de la migración serán considerados de alto riesgo.

Estas IAs estarán sujetas a estrictas regulaciones como, exhaustivos controles de calidad, registros de actividad constantes y medidas adecuadas de supervisión humana.

Los riesgos limitados incluyen a los que puedan presentar problemas relacionados a la falta de transparencia en el uso de las mismas. Por ello sólo se permitirán aquellas IAs de Riesgo limitado las cuales aporten la suficiente información a su usuario. Un ejemplo de esto sería un chatbot que sólo estaría permitido en el caso de que no hubiese ninguna en la interacción hombre-máquina, es decir, que el usuario siempre sea consciente que no está chateando con una persona real.

Las IAs de riesgo nulo cómo los filtros de spam o las usadas en videojuegos no tienen restricciones.

El proyecto cumple con la Ley de inteligencia artificial ya que se trata de una inteligencia artificial de Riesgo limitado en la que los usuarios en todo momento conocen que están interactuando con una máquina.[[Europea, 2024](#)]

Para entender la viabilidad legal de los datos hay que consultar su fuente original, PubMed, los artículos ahí publicados son recursos gratuitos que apoyan la investigación y recuperación de literatura biomédica.[[H, 2023](#)]

Apéndice B

Documentación de usuario

B.1. Requisitos software y hardware para ejecutar el proyecto.

Requisitos software

Las librerías necesarias para este proyecto se encuentran al principio del proyecto para su instalación en servicios en línea con aceleración de GPU pero en el caso de su instalación en local se encontrarán en el archivo requirements.txt, también para ejecutarlo en local se deberá tener Python y pip instalado

Requisitos hardware

No se recomienda la ejecución en local del programa, sin embargo se ha realizado una medición de la gpu necesaria para ejecutar el proyecto ya que es el aspecto más exigente, se puede ver en la figura [B.1](#).

B.2. Instalación / Puesta en marcha

Debido a que no se creó un espacio en Hugging Face que pueda soportar de forma permanente los grandes requisitos de GPU del proyecto se recurre a demos temporales.

Para visualizar la prueba de concepto simplemente hay que ejecutar el código presente en el repositorio en un entorno con suficiente GPU, por



Figura B.1: Requisitos de GPU

ejemplo, Kaggle o Google colab y acceder al link que se creará al final de la ejecución.

Para su puesta en marcha en local se deberá clonar el repositorio en github e instalar en el entorno deseado el archivo requirements.txt con el siguiente comando `pip install -r requirements.txt` siempre y cuando se tenga instalado tanto Python como pip en el sistema.

B.3. Manuales y/o Demostraciones prácticas

En este ejemplo se ejecutará el proyecto desde google colab empleando la aceleración por hardware T4 la cual será suficiente para la ejecución del proyecto, esto se podrá seleccionar desde la opción en el panel superior Entorno de ejecución -> Cambiar tipo de entorno de ejecución y seleccionar un acelerador por hardware suficiente para ejecutar el proyecto, T4 es suficiente como se ve en la figura B.2.

Una vez ejecutado el proyecto se abrirá al final una ventana con la que se podrá interactuar con el modelo y un link con el que se podrá interactuar con el en línea y compartirlo para que se pueda acceder a él desde distintos dispositivos, el ejemplo de la interfaz vacía se puede ver en la figura B.3.

Una vez ejecutada simplemente falta insertar el prompt deseado y seleccionar submit, luego, el modelo arrojará una respuesta con los abstracts que haya interpretado como los más relevantes según el prompt del usuario y la respuesta del prompt enriquecido con dichos abstracts obteniendo así la

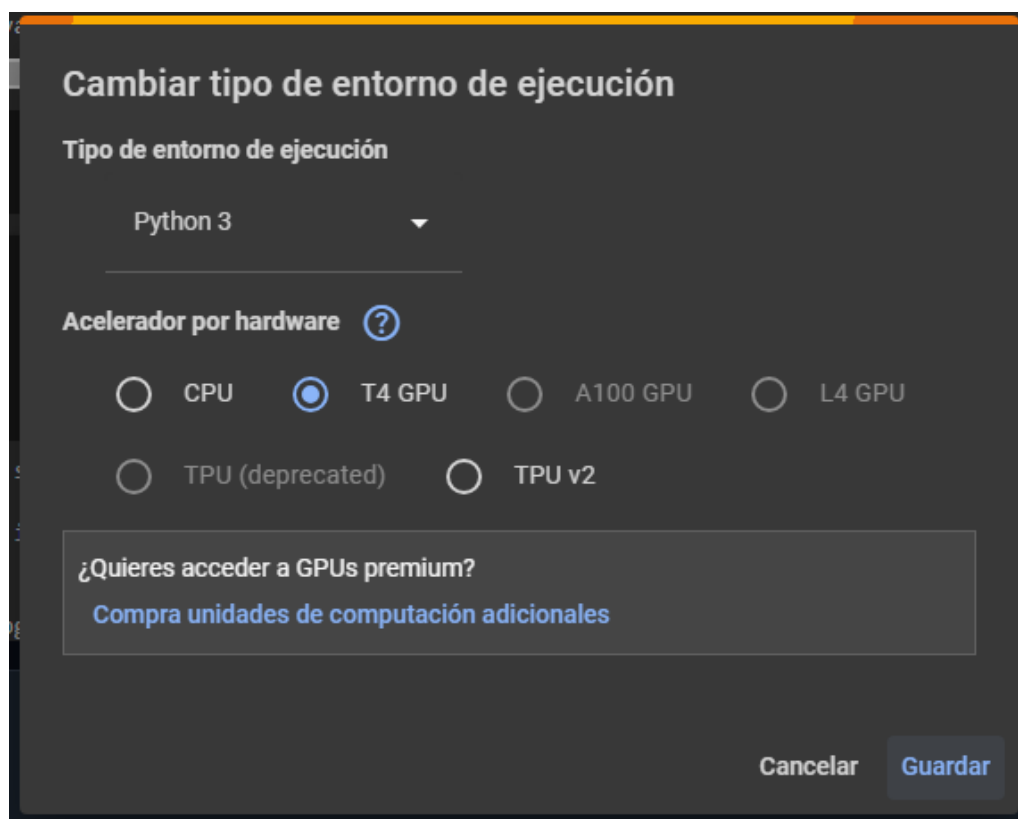


Figura B.2: Demostración de aceleración por hardware

salida esperada de la respuesta al prompt inicial y bibliografía que la apoye, se puede ver en la figura B.4

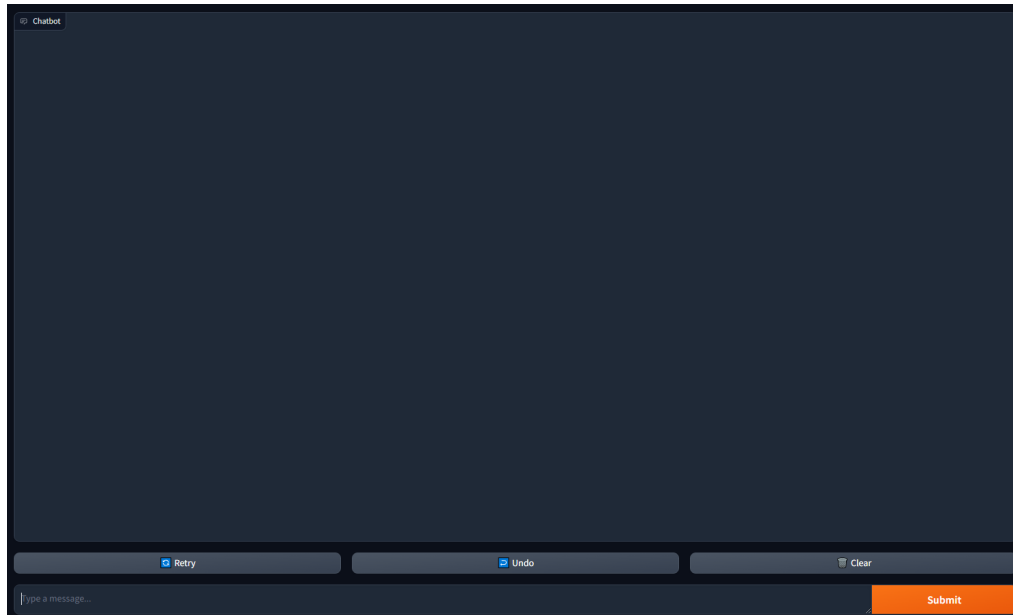


Figura B.3: Interfaz gráfica de usuario

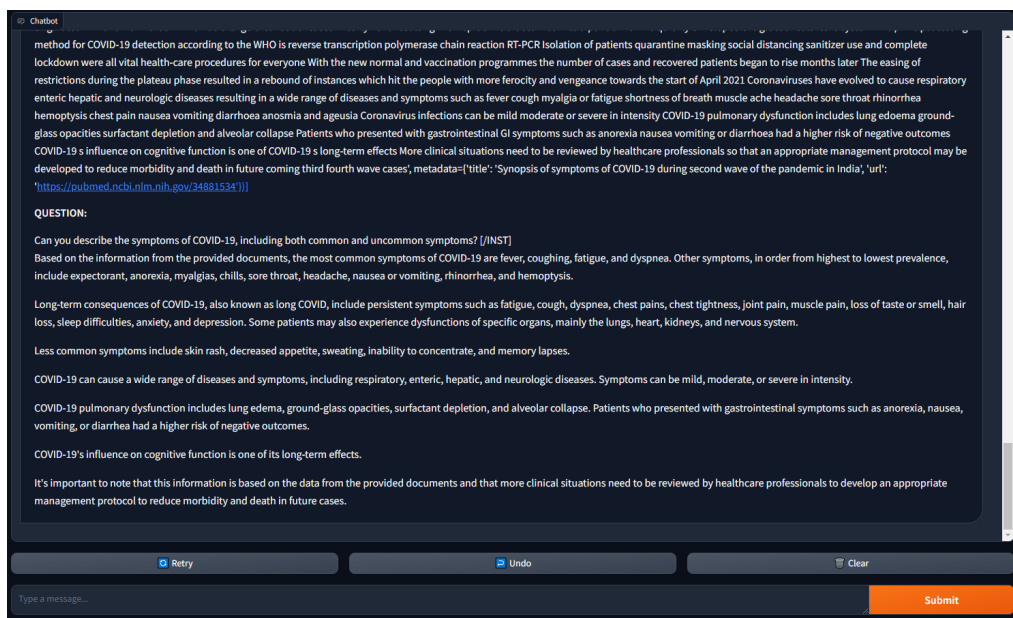


Figura B.4: Interfaz gráfica de usuario tras una ejecución

Apéndice C

Manual del programador.

C.1. Estructura de directorios

Descripción de los directorios y ficheros entregados.

- **/:** Directorio base, en el se encuentran el resto de directorios y los archivos readme dónde se encuentra información relevante del proyecto, las licencias y el código de conducta.
- **/.github/ISSUE_TEMPLATE/:** Carpeta que contienen archivos que indican unas guidelines en caso de bugs o peticiones de nuevas características.
- **/application/:** Carpeta que contiene el código fuente para ejecutar el proyecto.
- **/data/:** Directorio en el que están presentes los datos que posteriormente se vectorizarán para realizar el proceso de RAG
- **/documentación/:** Carpeta que contiene documentación relevante para entender la totalidad del proyecto, desde aspectos relevantes de su desarrollo hasta guías de ejecución o ejemplos de iteraciones.

C.2. Compilación, instalación y ejecución del proyecto

Para visualizar la prueba de concepto simplemente hay que ejecutar el código presente en el repositorio en un entorno con suficiente GPU, por

ejemplo, Kaggle o Google colab y acceder al link que se creará al final de la ejecución.

C.3. Pruebas del sistema

Para comprobar que la aplicación se haya ejecutado correctamente simplemente hay que comprobar que la ejecución devuelva la interfaz gráfica de usuario con su enlace para ejecutarlo en web.

C.4. Instrucciones para la modificación o mejora del proyecto.

Para la modificación del código simplemente hay que acceder al código presente en el repositorio de GitHub ya sea clonando el repositorio o descargándolo y luego editar el archivo en un editor de texto. Se recomienda para facilitar las pruebas, emplear entornos que faciliten aceleración de GPU como Kaggle o Google colab o bien ejecutarlo en una máquina potente que aguante los altos requisitos del proyecto.

Para la modificación del código es necesario un access token de Hugging Face, cualquier persona puede generarla de manera gratuita con una cuenta de Hugging Face, por lo que el primer paso será acceder a [Hugging Face](#). registrarse e iniciar sesión con la cuenta creada. Una vez creada la cuenta se debe acceder a Settings -> Access Token -> New Token y copiar el Token ahí presente. Luego se accede al código y, en la sección de login, insertar el token, tendrá este formato:

```
from huggingface_hub import login

login("pegar aquí el token de acceso")
```

Apéndice D

Descripción de adquisición y tratamiento de datos

D.1. Descripción formal de los datos

Los datos empleados en el proceso de RAG para refinar los resultados es un gran dataset publicado en Kaggle en formato csv. Este Dataset consta de tres campos:

- title: el título del paper
- abstract: el abstract del paper
- url: URL que dirige al paper

El dataset contiene 10000 papers en total

En esta prueba de concepto sólo son de interés los abstract, las otras dos columnas no se utilizarán.

[H, 2023]

D.2. Descripción clínica de los datos.

Papers científicos

Un paper científico es un documento escrito cuyo objetivo es la divulgación de la investigación realizada por profesionales. Los papers científicos

forman el soporte estructural de toda la ciencia ya que estos documentos no sólo aportan valor en la época que son publicados sino que permanecen en el tiempo convirtiéndose en parte fundamental del conocimiento. Los papers usualmente contienen una estructura formada por abstract, introducción, materiales y métodos, resultados, discusión, conclusión y referencias.[Katz, 1985]

Abstracts

El abstract es la parte más importante del paper científico puesto que, además de servir como resumen del mismo, es la carta de presentación del paper, siendo la primera sección que los interesados consultarán.

Un abstract debe ser escrito según las guías de la revista científica en la que será publicado el artículo y preferiblemente deberá contener entre 150 y 250 palabras. Las preguntas fundamentales que un abstract debe responder son las siguientes:

- ¿Porqué se empezó la investigación?
- ¿Qué se hizo en la investigación?
- ¿Qué encontramos?
- ¿Qué significan los resultados?

[Meo, 2018]

Apéndice E

Manual de especificación de diseño

E.1. Planos

No proceden en este proyecto

E.2. Diseño arquitectónico

En esta sección se mostrarán los diagramas de secuencia de cada caso de uso lo que aportará claridad a su ejecución además de un diagrama de componentes que esquematizará los distintos miembros constituyentes presentes en el proyecto.

Diagramas de secuencia

CU-1 Generar Respuesta

Ver imagen [E.1](#)

CU-2 Actualizar el modelo

Ver imagen [E.2](#)

CU-3 Modificar la información especializada

Ver imagen [E.3](#)

CU-4 Repetir el prompt

Ver imagen [E.4](#)

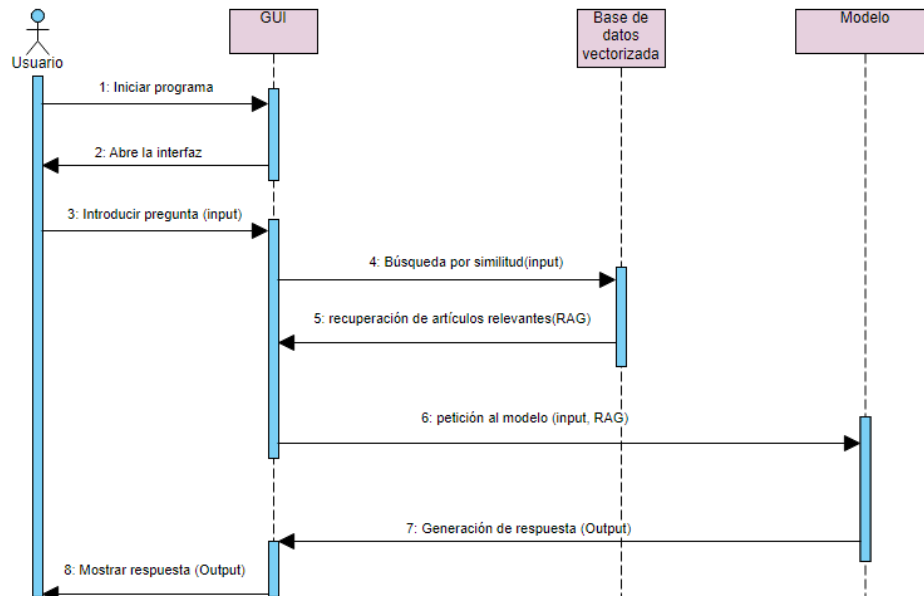


Figura E.1: Diagrama de secuencia del caso de uso número 1

CU-5 Eliminar conversación

Ver imagen [E.5](#)

Diagrama de componentes

En la figura [E.6](#) se puede apreciar el diagrama de componentes del proyecto que pretende ilustrar la disposición de cada parte relevante de la aplicación así como sus interacciones. A continuación se explicará brevemente cada componente.

- **Aplicación:** código general que tiene función estructural y de control del resto de componentes del proyecto.
- **HuggingFace:** comunidad de inteligencia artificial que permite el acceso mediante tokens a modelos de manera gratuita con alguna restricción.
- **Mistral 7B:** modelo Grande de Lenguaje (LLM) creado por Mistral, destaca por su pequeño tamaño y gran optimización en comparación a otros modelos similares lo que permite una relación coste/calidad superior.

- **GUI:** interfaz de usuario desarrollada con el framework de Gradio que permite al usuario interactuar de manera cómoda e intuitiva con el proyecto.
- **Base de datos vectorizada:** base de datos de tipo índice FAISS que contiene la información de entrenamiento para el proceso de RAG vectorizada para calcular su similitud matemática con el prompt de entrada.
- **Generación de embeddings:** proceso realizado por un modelo de machine learning que representa la información originalmente pasada en lenguaje natural como un vector medible y comparable.
- **Datos de entrenamiento:** conjunto de datos en .CSV (podrían tener otros formatos) sobre los que se quieren obtener respuestas especializadas.

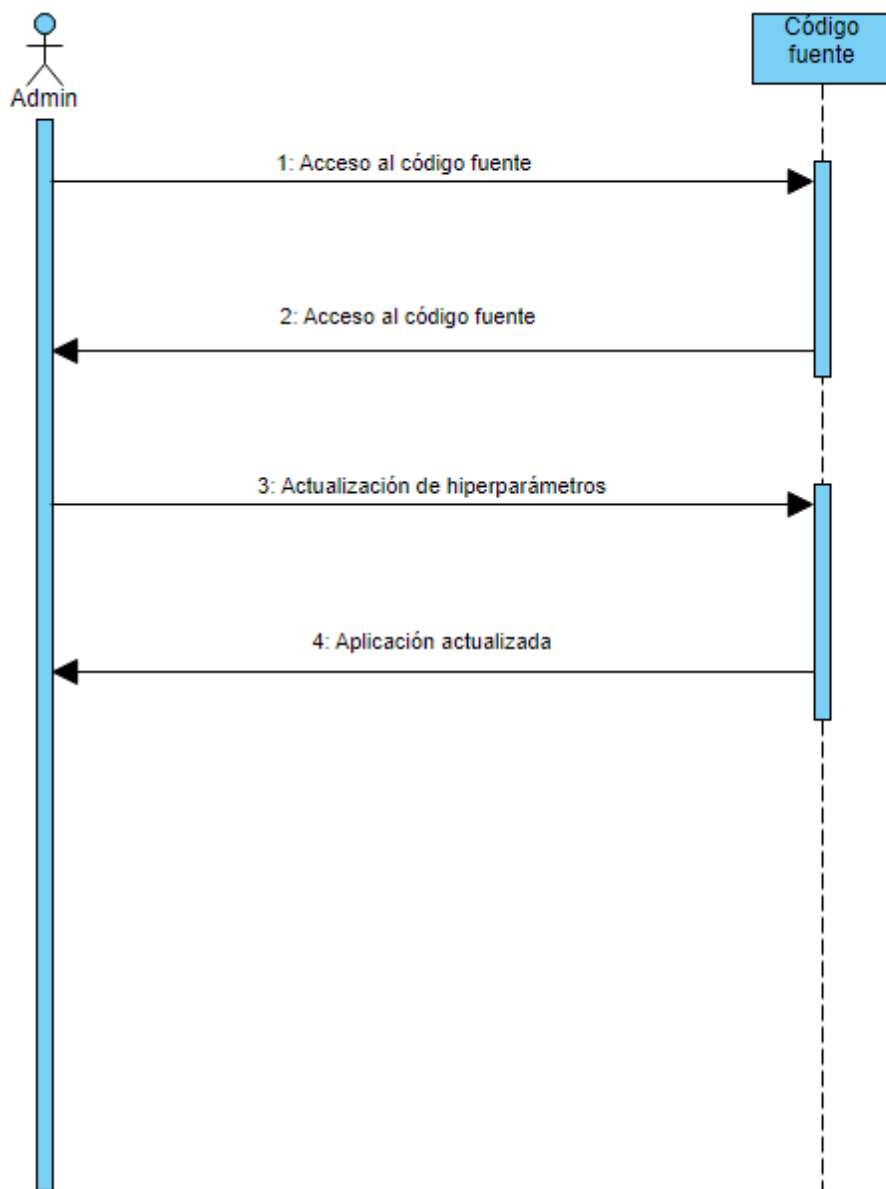


Figura E.2: Diagrama de secuencia del caso de uso número 2

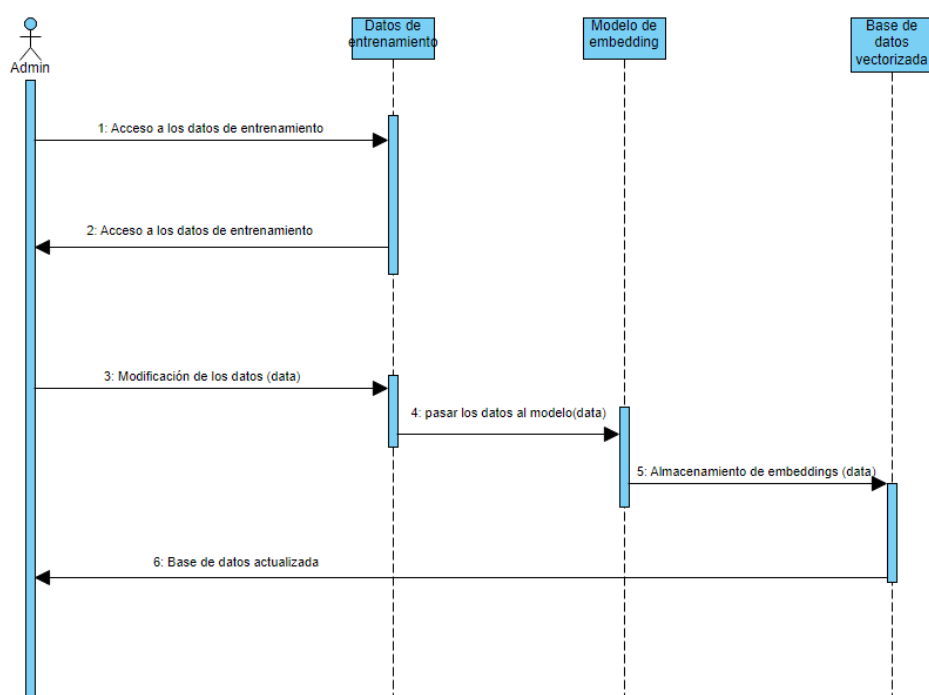


Figura E.3: Diagrama de secuencia del caso de uso número 3

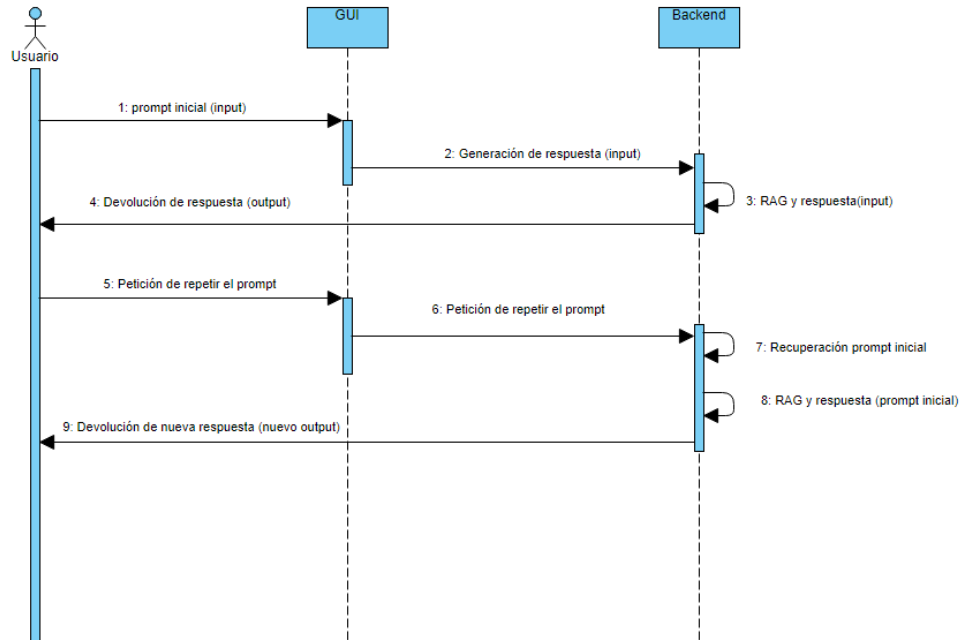


Figura E.4: Diagrama de secuencia del caso de uso número 4

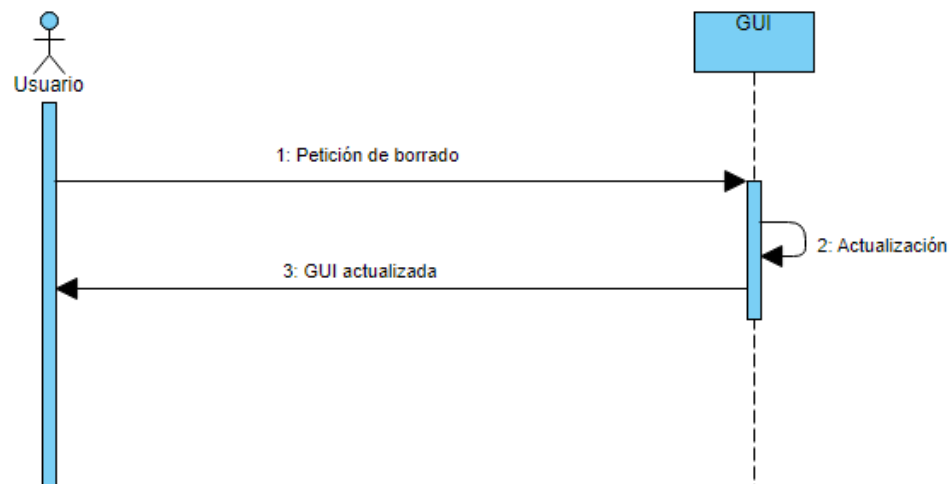


Figura E.5: Diagrama de secuencia del caso de uso número 5

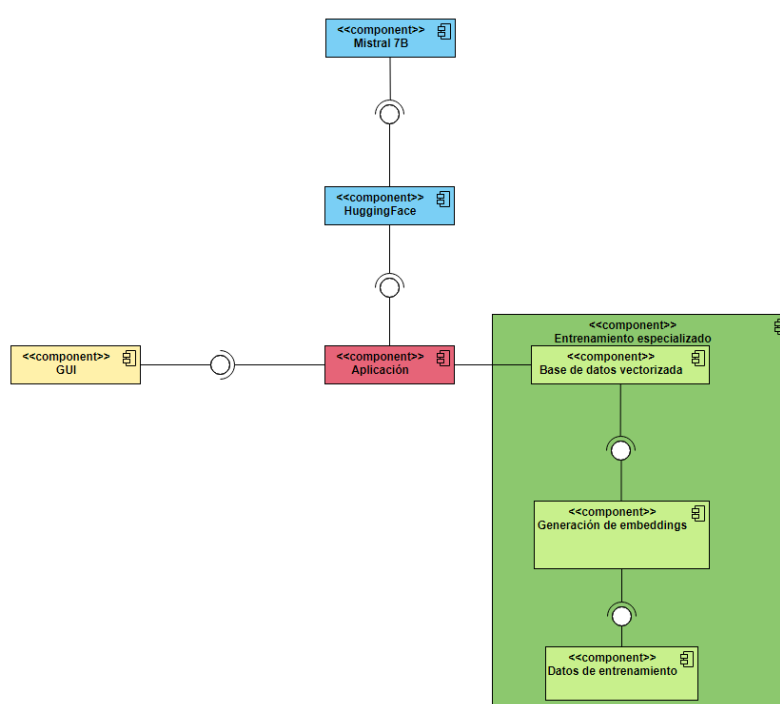


Figura E.6: Diagrama de componentes del proyecto

Apéndice *F*

Especificación de Requisitos

Los requisitos funcionales describen el comportamiento que se desea que el programa tenga, es decir, todas las tareas, servicios o funciones que debe realizar el programa. [Malan et al., 2001]

Por otra parte los requisitos no funcionales son aquellos que no definen lo que el software hará sino que se centran en cómo lo hará, por poner unos ejemplos pueden ser requisitos de rendimiento, limitaciones de diseño o características de calidad. [Chung et al., 2000]

Catálogo de requisitos

Requisitos funcionales

- **RF-1 Interacción con el modelo:** la interacción con el modelo debe poder realizarse mediante la introducción de prompts al mismo.
- **RF-2 Procesamiento del lenguaje natural:** toda interacción con el modelo debe poder realizarse mediante lenguaje natural sin necesidad de código o comandos de ningún tipo, a su vez, las respuestas proporcionadas por el programa deberán estar presentadas en lenguaje natural.
- **RF-3 Recuperación de embeddings más significativos:** el programa deberá vectorizar el prompt proporcionado por el usuario, compararlo con los abstract vectorizados en la base de datos y devolver los más significativos, es decir, los abstracts vectorizados cuya distancia al prompt embebido sea menor.

- **RF-4 Devolución de respuesta al usuario:** el programa deberá, con el prompt y los abstracts de mayor relevancia, llamar al modelo y generar una respuesta enriquecida la cual será presentada al usuario.
- **RF-5 Referencia a la fuente mediante URL en la respuesta:** todos los abstracts recuperados por el modelo deberán ir acompañados con el hipervínculo que redirija al texto completo del artículo mencionado, con ello el usuario puede consultar el texto completo en caso de que quiera profundizar la información con fuentes contrastadas.
- **RF-6 Informe de error:** en el caso de que haya algún error en la ejecución del programa o si el usuario introduce un prompt inválido se debe informar al usuario arrojándole el mensaje de error pertinente.
- **RF-7 Borrado de conversaciones:** el usuario debe ser capaz de eliminar la última interacción con el programa.
- **RF-8 Repetición del prompt:** el usuario debe ser capaz de repetir el prompt previamente introducido de manera cómoda para poder comparar entre diferentes salidas.
- **RF-9 Modificación de base de datos vectorizada:** el administrador del programa debe ser capaz de modificar en su totalidad la base de datos vectorizada mediante la cual se realiza el proceso de RAG.
- **RF-10 Afinación de hiperparámetros:** el administrador debe poder modificar el comportamiento del modelo cambiando sus hiperparámetros.
- **RF-11 Inicio de interfaz gráfica:** la interfaz gráfica debe mostrarse tras la ejecución del programa.
- **RF-12 Despliegue en línea:** tras la ejecución del programa, este debe generar un enlace para ser consultado en línea así como compartido entre distintos usuarios.

Requisitos no funcionales

- **RNF-1 Usabilidad:** todo el programa debe contar con funcionalidades lo suficientemente intuitivas como para que un profesional en el ámbito de la salud lo pueda usar sin ningún conocimiento previo en el campo de la computación.

- **RNF-2 Mantenibilidad:** el mantenimiento del proyecto debe de estar al alcance de cualquier administrador cualificado encargado de mantener el programa actualizado. También se deben poder realizar mejoras siguiendo el manual del programador incluido en este mismo documento.
- **RNF-3 Portabilidad:** el programa debe poder ejecutarse en distintos dispositivos.
- **RNF-4 Simultaneidad:** el programa debe poder procesar peticiones provenientes de distintos usuarios.
- **RNF-5 Fidelidad:** los artículos presentes en la aplicación deben de provenir de fuentes fiables y contrastadas, al tratarse información sanitaria no se debe dejar al azar las fuentes de información.
- **RNF-6 Documentación:** la aplicación debe tener manuales de soporte que faciliten su comprensión y uso.
- **RNF-7 Adecuación:** la respuesta proporcionada por el modelo debe estar en sincronía con el prompt proporcionado por el usuario, respondiendo a las preguntas que este propone y realizando las tareas demandadas siempre y cuando estén dentro de las facilidades que el modelo puede proporcionar.
- **RNF-8 Adaptabilidad:** el modelo debe responder correctamente ante modificaciones en la base de datos vectorizada.
- **RNF-9 Decencia:** las respuestas proporcionadas por el modelo deben obedecer a las normas de conducta básicas sin caer en insultos, lenguaje malsonante o desprecio a minorías aunque esto se encuentre en el prompt proporcionado por el usuario.

F.1. Diagrama de casos de uso

En la figura [F.1](#) se puede ver un diagrama de casos de uso con tres actores principales, el usuario que es cualquier persona que utilice el programa para generar respuestas a sus cuestiones, el administrador que será el encargado del correcto funcionamiento del proyecto, gestionando los parámetros del modelo así cómo la información que dispone para el enriquecimiento y la base de datos vectorizada donde se almacenan los embeddings de la información con la que queramos enriquecer los prompts con el RAG.

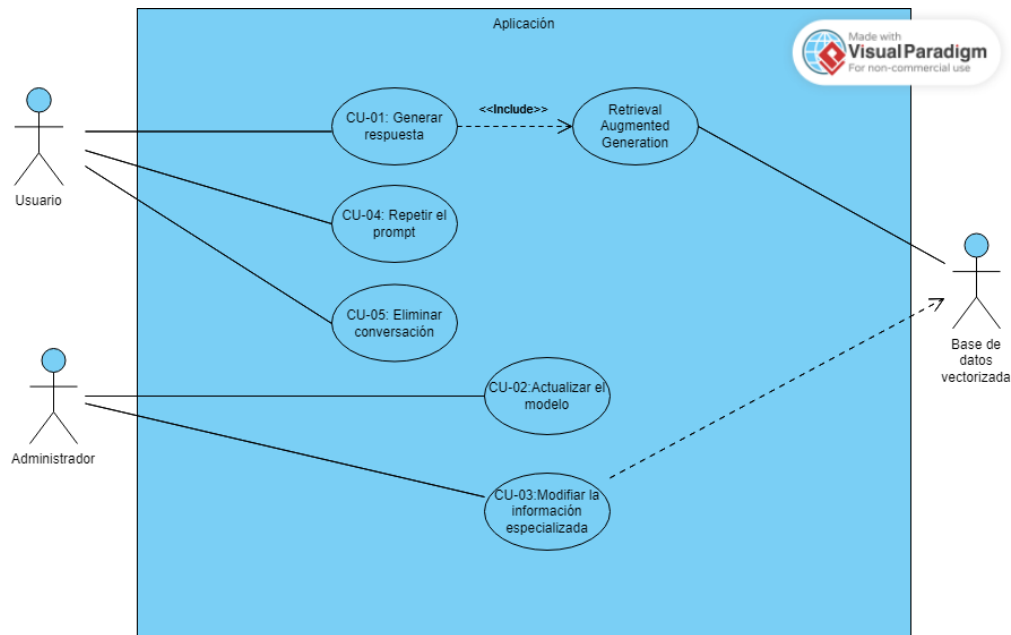


Figura F.1: Diagrama de casos de uso

F.2. Explicación casos de uso.

Se puede describir mediante el uso de tablas o mediante lenguaje natural.

Una muestra de cómo podría ser una tabla de casos de uso:

F.3. Prototipos de interfaz o interacción con el proyecto.

Para este proyecto se ha desarrollado una interfaz de usuario básica que se puede ver en las imágenes [B.3](#) y [B.4](#)

CU-1	Generar Respuesta
Versión	1.0
Autor	Daniel de Lara Pérez
Requisitos asociados	RF-1, RF-2, RF-3, RF-4, RF-5, RF-6
Descripción	El usuario introduce, empleando lenguaje natural, una pregunta o solicitud al programa, esta pregunta se enriquecerá y un modelo generará una respuesta que se devuelve al usuario por pantalla
Precondición	El proyecto está iniciado
Acciones	<ol style="list-style-type: none"> 1. El programa recibe la solicitud introducida por el usuario. 2. Se vectoriza la solicitud para tener una representación matemática de la información. 3. El vector se compara, calculando su distancia, con los presentes en la base de datos vectorizada. 4. Se devuelve, junto con la solicitud inicial, la información cuya distancia al embedding de la solicitud sea más pequeña. 5. La solicitud inicial junto con el prompt son pasados al modelo para enriquecerlo y generar una respuesta especializada. 6. Se devuelve la respuesta al usuario. 7. El programa queda a la espera de una nueva solicitud.
Postcondición	La respuesta generada se devuelve al usuario.
Excepciones	Si el programa recibe una solicitud que no es capaz de procesar como caracteres especiales o lenguaje sin sentido se le informará al usuario
Importancia	Alta

Tabla F.1: CU-1 Generar Respuesta.

CU-2	Actualizar el modelo
Versión	1.0
Autor	Daniel de Lara Pérez
Requisitos asociados	RF-10
Descripción	El administrador modifica el código fuente afinando los valores que se le pasan al modelo como hiperparámetros
Precondición	El administrador tiene acceso al código fuente
Acciones	<ol style="list-style-type: none"> 1. El administrador modifica los hiperparámetros por los que considere necesarios. 2. Se vuelve a ejecutar el código fuente.
Postcondición	Futuras iteraciones del modelo generan respuestas con los nuevos hiperparámetros.
Excepciones	
Importancia	Alta

Tabla F.2: CU-2 Actualizar el modelo.

CU-3	Modificar la información especializada
Versión	1.0
Autor	Daniel de Lara Pérez
Requisitos asociados	RF-10
Descripción	El administrador modifica la información contenida en la base de datos vectorizada, eliminando posibles datos erróneos, añadiendo nueva información o modificando el tema de la especialización.
Precondición	El administrador tiene acceso al archivo dónde se guarda la información
Acciones	<ol style="list-style-type: none"> 1. El administrador modifica el archivo añadiendo o eliminando según considere para la tarea que se quiera realizar. 2. Se vuelve a cargar el archivo con la información (Si el formato en el que se encuentra presentado cambia es posible que haya que modificar también el cargador) 3. Se tokeniza la información si no se encuentra previamente en tokens establecidos. 4. Se vuelve a ejecutar el código fuente.
Postcondición	Futuras iteraciones del modelo generan respuestas especializadas con la nueva información.
Excepciones	
Importancia	Alta

Tabla F.3: CU-3 Modificar la información especializada.

CU-4	Repetir el prompt
Versión	1.0
Autor	Daniel de Lara Pérez
Requisitos asociados	RF-8
Descripción	El usuario repite la misma pregunta para poder contrastar entre dos respuestas diferentes.
Precondición	El proyecto está iniciado y se ha generado una respuesta
Acciones	<ol style="list-style-type: none">1. El usuario selecciona la opción de repetir el prompt2. El programa realiza la misma consulta ya realizada previamente
Postcondición	El usuario recibe una nueva consulta con el prompt previamente seleccionado.
Excepciones	No se ha introducido ningún prompt
Importancia	Media

Tabla F.4: CU-4 Repetir el prompt.

CU-5	Eliminar conversación
Versión	1.0
Autor	Daniel de Lara Pérez
Requisitos asociados	RF-7
Descripción	El usuario elimina la última interacción con el modelo o la conversación en su totalidad.
Precondición	El proyecto está iniciado y se ha generado, al menos, una respuesta
Acciones	<ol style="list-style-type: none">1. El usuario selecciona la opción de eliminar última interacción o toda la conversación2. El programa elimina del registro la o las interacciones designadas
Postcondición	La o las consultas seleccionadas no se muestran por pantalla.
Excepciones	No se ha introducido ningún prompt
Importancia	Baja

Tabla F.5: CU-5 Eliminar conversación.

Apéndice G

Estudio experimental

En este apéndice se mostrarán y explicarán las distintas pruebas que se realizaron así cómo las condiciones en las que se realizaron y su discusión.

G.1. Cuaderno de trabajo.

Se enumeran todos los métodos probados con resultados positivos o no.

1.- Primero se intentó realizar el proyecto en un entorno local mediante el editor de texto Visual Studio y un modelo pequeño de Mistral así cómo una versión preliminar de una aplicación Web empleando Streamlit, sin embargo los resultados que se obtuvieron con ese modelo eran bastante malos y no se parecían a las salidas esperadas.

2.- Tras el fracaso del primer modelo se probó a actualizar al Mistral 7B sin embargo, las librerías necesarias para ejecutarlo así cómo el tamaño del modelo hizo imposible la ejecución en local por lo que se pasó a los entornos virtuales, en este caso Google Colab que permite la aceleración de gpu de forma gratuita.

3.- Se cambió la interfaz gráfica de usuario de Streamlit ya que, el gran tamaño de la tecnología empleada hacía que fuera imposible lanzar la aplicación a la web sin emplear espacios de pago con gran GPU por lo que al final se decantó por emplear demos temporales en Gradio con posibilidad de mejora a entornos permanentes.

4.- En las primeras pruebas del modelo se usaron prompts sencillos y sin técnicas de prompt engineering y se obtenían resultados a veces buenos

pero otras veces el lenguaje de la salida era confuso y con los espacios mal puestos un ejemplo:

Prompt: Specify the symptoms of covid-19

Salida: According to multiple scientific publications, including those retrieved through bibliographical searches in PubMed and Google Scholar (as mentioned in two documents provided), the symptoms of covid-19 can be broadly categorised into general and organ-specific manifestations. The most common persists symptoms after recovery. Covid signs also include skin rash, the descended appetite, sweat, inability to concentrate, memory lapse. In addition, Analytic concluding that fever, sore throat, myalgia, stiffness, headache, recurrent nasal fissures, pricking or fiery path, a higher risk for negative outcomes. It should be noted however that the meta-analytics mentioned earlier ranks fever as the most common symptom in both the dynamic phases and convalescent periods afterwards: Fever, wet cough, fluid retaining around legs, difficulty breathing, high temperature, low grade fever, tiredness, aching limbs, joint swellings, sleep disturbances, digital clubbing, changes in gait, pattern, anaemia, hirsutism, splitting haemorrhages, sudden weight gain, new sensory perceptions like tinnitus, rash, palpitations, pressure changes, constipation, excessive thirst, dark urine, colicky abdominal discomfort, poor nutrition, reduced libido, extreme mood swings, trouble concentrating, seeing spots, floaters, double vision, light headedness, dizziness, emesis, projectile vomiting, water brash, ophthalmos, periorbital disease, trembling, voice change, in yellowish eyes, black stools, rectal bleeding, rhythmic irregularity, cardiac arrhythmia, liver cirrhosis, stroke, paralyzing weakness, convulsions, electrolyte imbalance, dementia. Although not included in the number of documents provided here, it's worth noting that long term effects on cognitive function has also been identified as a significant consequence of covid-19 infection especially if left untreated promptly. References available upon request.

5.- Tras los malos resultados y, buscando la mejora en los resultados se emplearon técnicas de prompt engineering para comunicarse de manera eficiente con el modelo.

Prompt: Can you describe the symptoms of COVID-19, including both common and uncommon symptoms?

Salida: Based on several publications indexed in PubMed, the most frequent symptoms of COVID-19 are fever (up to 78 %), coughing (around 60 %), fatigue (approximately 33 %), and dyspnea (about 26 %) [1]. These findings come from a systematic review and meta-analysis covering over 132,000 infected individuals across multiple continents [1][2]. Common additional symptoms include but aren't restricted to:

Expectoration - mucous secretion when coughing; approximately 22 %.

Anosmia – loss of sense of smell; roughly 21 %, especially among those not previously diagnosed through traditional methods like temperature checks

or CT scan results [1], although other sources suggest upwards of 65 % depending upon population demographics [3].

Ageusia— lack of ability to perceive tastes; nearly half of affected persons report experiencing significant changes in flavor perception alongside olfactory issues [1].

Myalgias / Arthralgias – muscular & bone pain respectively; affecting almost 17 %; often accompanied by weakness and stiffness [1].

Chills – shivers indicative of body cooling down involuntarily; present in close to 15 % of sufferers [1].

Headaches – occurring in just under 11 % of afflicted populations [1]; sometimes precede viral illness development [4].

Nausea or Vomiting – affects slightly above 8 % [1].

Rhinorrhea – runny nose; occurs within 8 % of subjects surveyed [1].

Hemoptysis – bloodied sputum expelled via coughing; only seen in fewer than 3 % of respondents [1].

Longer term post-recovery impacts known colloquially as "long covid" consist of various lingering complaints lasting weeks to years past initial resolution of primary Covid-related symptoms. Such chronic problems span numerous organ systems—respiratory, cardiovascular, digestive, immunological, neuropsychologic etc., necessitating further research attention regarding diagnosis, prognosis, prevention strategies, treatments options, and overall impact assessment [2][3].

6.- Para mejorar aún más los resultados se realizaron técnicas de optimización de hiperparámetros, así como se refinó el método de retrieval, lo que mejoró aún más los ya buenos resultados obtenidos.

G.2. Configuración y parametrización de las técnicas.

Para la generación de respuestas se empleó el modelo Mistral 7B con los siguientes hiperparámetros:

- Tarea: Generación de Texto
- Temperatura 0.2

- Penalización por repetición: 1
- Tokens máximos nuevos: 1000

Se seleccionó un valor bajo de la temperatura para que se ajuste a la información proporcionada y no genere datos inventados a partir de otra información que posea el modelo.

Para el entorno virtual en el que se realizaron las pruebas se seleccionó Google Colab con el Tipo de entorno de ejecución en Python 3 y el acelerador por hardware T4 GPU para ayudar a ejecutar las tareas costosas que ejecutará el modelo.

Para las tareas de retrieval y generación de Embeddings se usó el modelo sentence-transformers/all-mpnet-base-v2 especializado en la generación de los mismos, para el RAG se usó el sistema de búsqueda similarity_score_threshold con un valor de threshold del 0.4 es decir, sólo devolverá los abstracts que presenten una similitud de coseno mayor de 0.4.

Para generar los prompts se emplearon las siguientes técnicas de prompt engineering:

- **Claridad y Especificidad:** los prompts están formulados de manera clara y específica para que el modelo entienda exactamente qué tipo de información se espera. Por ejemplo, se especifica que se deben incluir tanto síntomas comunes como menos comunes.
- **Contexto y Rol Asignado:** al asignar un rol específico al modelo (por ejemplo, eres un experto en enfermedades infecciosas o eres un experto en salud pública), se le proporciona un contexto que le ayuda a generar una respuesta más informada y precisa.
- **Instrucciones Detalladas:** se incluyen instrucciones detalladas para guiar la respuesta del modelo. Por ejemplo, se pide mencionar variaciones en los síntomas según la gravedad de la enfermedad o entre diferentes grupos de población.
- **Formato y Estructura:** la estructura del prompt está diseñada para ser lógica y coherente, comenzando con una petición clara y luego proporcionando detalles adicionales que deben ser incluidos.
- **Preguntas Abiertas y Específicas:** se utilizan preguntas abiertas que permiten una respuesta detallada, pero también se asegura de que sean lo suficientemente específicas para evitar respuestas demasiado generales o fuera de tema.

G.3. Detalle de resultados.

Para la validación se generó la tabla [G.1](#) donde validar el modelo y se generaron varias pruebas, tanto de parte del desarrollador cómo de terceros.

Los resultados obtenidos en el proyecto son muchas veces los deseados, sin embargo en ocasiones hay que refinar el prompt por fallos en el modelo como por ejemplo, la pérdida de consistencia en los espacios de la salida aunque la información si que sea la correcta. Esto implica que para la obtención del resultado esperado se deban realizar varios intentos en algunos casos hasta obtener el mejor output posible.

Validación por parte del desarrollador

Prompt: What are the molecular and cellular mechanisms by which SARS-CoV-2, the virus responsible for Covid-19, enters human cells and evades the immune system?

Contexto: [Document(page_content='After two previous episodes in 2002 and 2012 when two highly pathogenic coronaviruses SARS MERS with a zoonotic origin emerged in humans and caused fatal respiratory illness we are today experiencing the COVID-19 pandemic produced by SARS-CoV-2 The main question of the year 2021 is if naturally- or artificially-acquired active immunity will be effective against the evolving SARS-CoV-2 variants This review starts with the presentation of the two compartments of antiviral immunity-humoral and cellular innate and adaptive-underlining how the involved cellular and molecular actors are intrinsically connected in the development of the immune response in SARS-CoV-2 infection Then the SARS-CoV-2 immunopathology as well as the derived diagnosis and therapeutic approaches will be discussed', metadata='title': 'Antiviral Immunity in SARS-CoV-2 Infection From Protective to Deleterious Responses', 'url': 'https://pubmed.ncbi.nlm.nih.gov/34946179'), Document(page_content='Background The spread of severe acute respiratory syndrome coronavirus 2 SARS-CoV-2 has precipitated the worst global pandemic in a century which has caused millions of infections and deaths as well as massive economic repercussions Objective As with any pathogenic virus it is crucial to understand its unique interactions with the human immune system so that pharmaceutical and prophylactic interventions can be deployed to effectively control the pandemic Methods A literature search by using PubMed was conducted in 2020 with variants of the terms COVID-19 SARS-CoV-2 and immunological response English language articles that presented original data about the immunologic response to coronavirus disease 2019 COVID-19 were selected for review This article

reviewed the current understanding of the innate and adaptive immune responses to SARS-CoV-2 infection including their relationship to current therapeutic and diagnostic strategies Results SARS-CoV-2 uses several unique molecular techniques to evade detection by the innate immune system early in the course of infection and upregulation of these innate immune pathways may possibly accelerate the time to recovery and prevent severe disease Although the majority of cases results in the patients recovery a significant proportion of infections result in deaths prompted by the host s inflammatory overreaction to the infection a response that can be attenuated with corticosteroids and potentially other immune modulators Conclusion Current work by the scientific community to further understand how SARS-CoV-2 interacts with the human immune system will be invaluable to our response and preparedness for future coronavirus pandemics', metadata='title': 'The immunologic response to severe acute respiratory syndrome coronavirus 2', 'url': 'https://pubmed.ncbi.nlm.nih.gov/34871157'), Document(page_content='Infection by severe acute respiratory syndrome coronavirus-2 SARS-CoV-2 provokes a potentially fatal pneumonia with multiorgan failure and high systemic inflammation To gain mechanistic insight and ferret out the root of this immune dysregulation we modeled by in vitro coculture the interactions between infected epithelial cells and immunocytes A strong response was induced in monocytes and B cells with a SARS-CoV-2-specific inflammatory gene cluster distinct from that seen in influenza A or Ebola virus-infected cocultures and which reproduced deviations reported in blood or lung myeloid cells from COVID-19 patients A substantial fraction of the effect could be reproduced after individual transfection of several SARS-CoV-2 proteins Spike and some nonstructural proteins mediated by soluble factors but not via transcriptional induction This response was greatly muted in monocytes from healthy children perhaps a clue to the age dependency of COVID-19 These results suggest that the inflammatory malfunction in COVID-19 is rooted in the earliest perturbations that SARS-CoV-2 induces in epithelia', metadata='title': 'A virus-specific monocyte inflammatory phenotype is induced by SARS-CoV-2 at the immune-epithelial interface', 'url': 'https://pubmed.ncbi.nlm.nih.gov/34969849'), Document(page_content='Life implies adaptation This is one of the fundamental principles that has permitted most living species to survive through ages in an ever-changing environment Spontaneously occurring events have shaped also virus populations and their fitness Thanks to their plasticity viruses have thrived in extremely dissimilar conditions Unsurprisingly SARS-CoV-2 the etiological agent of COVID-19 is no exception Thanks to an unprecedented rate of molecular tracing and sequence scrutiny the virus was followed in all its changes and shown to evolve in such a way as to possibly

determine subsequent waves of infection after the first global and massive outbreak This review illustrates the major modifications occurred to the virus since its discovery We describe the potential advantages that these changes conveyed as regards SARS-CoV-2 transmissibility resistance to host innate and adaptive barriers and molecular diagnosis', metadata='title': 'Evolution of viruses and the emergence of SARS-CoV-2 variants', 'url': 'https://pubmed.ncbi.nlm.nih.gov/34942015']

Respuesta: According to multiple studies analyzed in the provided documents, SARS- CoV-2 utilizes specific molecular techniques to enter human cells and evade the initial detection by the Innate Immune System during the early stages of infection (PubMed document < <https://pubmed.ncbi.nlm.nih.gov/34946179> >). One mechanism involves binding to Angiotensin-Converting Enzyme II (ACE2) receptors present primarily in alveolar type I andII pulmonary epithelium cells, allowing entry into those cells via endocytosis (<<https://pubmed.ncbi.nlm.nih.gov/34969849>>). Once inside, SARS-CoV-2 employs various methods like suppressing Interferon production - essential cytokines released upon recognition of foreign invaders – thereby inhibiting downstream signaling cascades necessary for activating both arms of the Antiviral Response; Humoral & Cellular (<<https://pubmed.ncbi.nlm.nih.gov/34946179>, <https://pubmed.ncbi.nlm.nih.gov/34871157>). Additionally, certain structural components within the Virus itself contribute significantly towards eluding Host Defense Mechanisms. For instance, spikes found on the surface of SARS-CoV-2 facilitate attachment onto target cells while minimizing interaction with Neutralizing antibodies generated due to prior exposure or vaccination (<https://www.scielo.br/j/rbio/a/ZXJKQNHGYBW>).. Furthermore, NonStructuralProteins play roles in manipulating RNA metabolism processes leading to impaired Translation Efficiencyand thus hampering efficient RecognitionbyHostCellMachinery(<https://pubmed.ncbi.nlm.nih.gov/34942015>). Overall,thesemechanisms enable SARS-Co V-2to successfullyenterhuman-cellsmaintainpersistentinfectionswhileevadinghostimmunesystemdetectioninitially.

Puntuación: **G.2**

Validación por otros llms, en este caso GPT

Prompt: What are the primary mechanisms underlying the evasion of immune responses by SARS-CoV-2 variants, and how might this impact the development and deployment of future vaccines?

Contexto: [Document(page_content='The emergence of SARS-CoV-2 variants may cause resistance at the immunity level against current vaccines Some emergent new variants have increased transmissibility infectivity hospitalization and mortality Since the administration of the first SARS-CoV-2

vaccine to a human in March 2020 there is an ongoing global race against SARS-CoV-2 to control the current pandemic situation Spike S glycoprotein of SARS-CoV-2 is the main target for current vaccine development which can neutralize the infection Companies and academic institutions have developed vaccines based on the S glycoprotein as well as its antigenic domains and epitopes which have been proven effective in generating neutralizing antibodies The effectiveness of SARS-CoV-2 vaccines and other therapeutics developments are limited by the new emergent variants at the global level We have discussed the emergent variants of SARS-CoV-2 on the efficacy of developed vaccines Presently most of the vaccines have been tremendously effective in severe diseases However there are still noteworthy challenges in certifying impartial vaccines the stories of re-infections are generating more stressful conditions and this needs further clinical evaluation', metadata='title': 'SARS-CoV-2 Emergence of New Variants and Effectiveness of Vaccines', 'url': 'https://pubmed.ncbi.nlm.nih.gov/34970509'), Document(page_content='The SARS-CoV-2 pandemic has highlighted how an emergent disease can spread globally and how vaccines are once again the most important public health policy to combat infectious disease Despite promising initial protection the rise of new viral variants calls into question how effective current SARS-CoV-2 vaccines will be moving forward Improving on vaccine platforms represents an opportunity to stay ahead of SARS-CoV-2 and keep the human population protected Many researchers focus on modifying delivery platforms or altering the antigens presented to improve the efficacy of the vaccines Identifying mechanisms of natural immunity that result in the control of infection and prevent poor clinical outcomes provides an alternative approach to the development of efficacious vaccines Early and current evidence shows that SARS-CoV-2 infection is marked by potent lung inflammation and relatively diminished antiviral signaling which leads to impaired immune recognition and viral clearance essentially making SARS-CoV-2 too hot to handle', metadata='title': 'Translating known drivers of COVID-19 disease severity to design better SARS-CoV-2 vaccines', 'url': 'https://pubmed.ncbi.nlm.nih.gov/34902803'), Document(page_content='Coronavirus disease 2019 COVID-19 has overwhelmed the healthcare and economy of the world with emerging new variants of severe acute respiratory syndrome coronavirus 2 SARS-CoV-2 posing an everlasting threat to humanity While most COVID-19 vaccines provide adequate protective immunological response against the original SARS-CoV-2 variant there is a pressing need to understand their biological and clinical responses Recent evidence suggests that some of the new variants of SARS-CoV-2 evade the protection conferred by the existing vaccines which may impede the ongoing efforts to expedite the vaccination programs worldwide These

concerns have also highlighted the importance of a pan-COVID-19 vaccine which is currently in the making. Thus it is imperative to have a better molecular and clinical understanding of the various COVID-19 vaccines and their immunological trajectory against any emerging variant of concern (VOCs) in particular to break this vicious cycle. Furthermore, other treatment regimens based on cellular therapies and monoclonal antibodies should be explored systematically as an alternative and readily available option considering the possibility of the emergence of more virulent SARS-CoV-2 mutants. In this review, we shed light on the various molecular mechanisms and clinical responses of COVID-19 vaccines. Importantly, we review the recent findings of their long-term immune protection and efficacy against emerging VOCs. Considering that other targeted and effective treatments will complement vaccine therapy, we provide a comprehensive understanding of the role of cell-based therapies, monoclonal antibodies, and immunomodulatory agents as alternative and readily available treatment modalities against any emerging SARS-CoV-2 variant.

, metadata='title': 'Molecular and Clinical Aspects of COVID-19 Vaccines and Other Therapeutic Interventions Apropos Emerging Variants of Concern', 'url': 'https://pubmed.ncbi.nlm.nih.gov/35002711'), Document(page_content='The persistent circulation of SARS-CoV-2 represents an ongoing global threat due to the emergence of new viral variants that can sometimes evade the immune system of previously exposed or vaccinated individuals. We conducted a follow-up study of adult individuals that had received an inactivated SARS-CoV-2 vaccine, evaluating antibody production and neutralizing activity over a period of 6 months. In addition, we performed mice immunization with inactivated SARS-CoV-2 and evaluated the immune response and pathological outcomes against Gamma and Zeta variant infection. Vaccinated individuals produced high levels of antibodies with robust neutralizing activity, which was significantly reduced against Gamma and Zeta variants. Production of IgG anti-S antibodies and neutralizing activity robustly reduced after 6 months of vaccination. Immunized mice demonstrated cellular response against Gamma and Zeta variants and after viral infection reduced viral loads, IL-6 expression, and histopathological outcome in the lungs. TNF levels were unchanged in immunized or not immunized mice after infection with the Gamma variant. Furthermore, serum neutralization activity rapidly increases after infection with the Gamma and Zeta variants. Our data suggest that immunization with inactivated WT SARS-CoV-2 induces a promptly responsive cross-reactive immunity response against the Gamma and Zeta variants, reducing COVID-19 pathological outcomes', metadata='title': 'Protective Immunity against Gamma and Zeta Variants after Inactivated SARS-CoV-2 Virus Immunization', 'url': 'https://pubmed.ncbi.nlm.nih.gov/34960708')]

Respuesta: According to the provided documents, several studies indicate that newly emerged SARS- CoV-2 variants possess certain characteristics enabling them to partially escape the protective effects of current vaccines (PubMed document links: 34970509, 34902803). Although these variations do not completely negate the benefits offered by present vaccinations - particularly regarding serious illnesses – they pose significant challenges when attempting to develop truly universal vaccines. Here’s what you need to know about the key factors contributing to such escapes and potential implications for upcoming vaccines:

Antibody Evasions through Mutations: One mechanism behind the ability of newer strains like Delta, Omicron etc., to circumvent previous immunity lies within specific genetic changes occurring primarily in spike proteins found on virus particles. This results in altered structures recognized differently than those encountered during prior exposures via earlier waves or successful vaccinations. For instance, one notable example includes Epsilon sublineage B.1.429 carrying K417N substitution near RBD region; another involves Alpha lineages harboring N501Y change close to the same area. Such modifications hinder binding between circulating antibodies generated upon exposure to ancestral forms and subsequent encounters with novel ones. Therefore, keeping track of continually shifting patterns becomes crucial since even small shifts could potentially render large portions of populations vulnerable if left undetected. To mitigate risks associated with rapid evolutionary adaptability exhibited by SARS-CoV-2, scientists continue exploring multiple avenues aimed towards creating next generation vaccines capable of addressing diverse threats effectively while minimizing reliance solely on single targets susceptible to frequent modification. Strategically speaking, focusing research around broadening coverage beyond just dominant protein regions would help ensure broader spectrum defenses against both established and incoming variants alike. Additionally, incorporating heterologous prime boost strategies involving different types of adjuvant systems alongside traditional messenger ribonucleic acid (RNA), vector-, live attenuated-, or whole-virion approaches offers promise in enhancing overall breadth and longevity of induced immune responses. Moreover, combining multivalent designs encompassing combinations of relevant conserved structural components across distinct clades ensures greater resilience amidst continuous variation among outbreak clusters. Last but certainly not least, integrating advanced computational modeling techniques facilitating real time monitoring and prediction capabilities enables proactively anticipating likely trends before widespread dissemination occurs. By employing all mentioned tactics simultaneously, scientific communities stand poised to tackle

complex issues surrounding dynamic nature of SARS-CoV-2 and minimize negative impacts arising from recurring pandemics caused by similar entities going forward.

Puntuación: **G.3**

Validación por terceros

-1º Ejemplo de validación por terceros

Prompt: Is covid relevant nowadays in Madrid?

Contexto: [Document(page_content='The intrusion of infectious diseases in everyday life forces humans to reassess their attitudes Indeed pandemics are able catalyze rapid transitions in scientific knowledge politics social behaviors culture and arts The current Coronavirus disease-19 COVID-19 outbreak has driven an unprecedented interest toward the influenza pandemic of 1918 The issue is whether history can shed light on the best preventive response and future scenarios The aim of this review is to highlight the parallelism between the two pandemics Starting from epidemiology and clinical features but further focusing on social and cultural issues it is possible to unveil great similarities Their outbreak pattern lead to hypothesize a similar duration and death burden in absence of effective vaccines or innovative treatments for COVID-19 Thus then as now preventive medicine represents the first and most effective tool to contain the course of the pandemic being treatments available only supportive At the same time both pandemics shared the same pattern of narration e g scapegoating and the same impact on minorities in high-income countries Furthermore visual art responded to pandemic issues in 2020 in the form of Graffiti art while similar role was ruled by Expressionism movement during the Spanish flu Photography also was capable to document both catastrophic scenarios Thus it is possible to find a lot of clinical and social similarities between the two pandemics Nevertheless if the Spanish flu was not unforeseen COVID-19 spillover was partially predictable and its global impact will hopefully not be overshadowed by a major crisis such as World War I', metadata={'title': 'COVID-19 and Spanish flu-18 review of medical and social parallelisms between two global pandemics', 'url': 'https://pubmed.ncbi.nlm.nih.gov/34909487'}), Document(page_content='Measuring mortality has been a challenge during the COVID-19 pandemic Here we compared the results from the Spanish daily mortality surveillance system MoMo of excess mortality estimates using a time series analysis with those obtained for the confirmed COVID-19 deaths reported to the National Epidemiological Surveillance Network RENAVE The excess mortality estimated at the beginning of March 2020 was much greater than what has been observed in previous years and clustered in a

very short time The cumulated excess mortality increased with age In the first epidemic wave the excess mortality estimated by MoMo was 1.5 times higher than the confirmed COVID-19 deaths reported to RENAVE but both estimates were similar in the following pandemic waves Estimated excess mortality and confirmed COVID-19 mortality rates were geographically distributed in a very heterogeneous way The greatest increase in mortality that has taken place in Spain in recent years was detected early by MoMo coinciding with the spread of the COVID-19 pandemic MoMo is able to identify risk situations for public health in a timely manner relying on mortality in general as an indirect indicator of various important public health problems', metadata='title': 'The Impact of COVID-19 on Mortality in Spain Monitoring Excess Mortality MoMo and the Surveillance of Confirmed COVID-19 Deaths', 'url': 'https://pubmed.ncbi.nlm.nih.gov/34960692'), Document(page_content='BACKGROUND All events in Europe including EURO2020 football championship were postponed due to Covid-19 in 2020 Instead Euro2020 took place in 2021 as mitigation measures were relaxed cross-country mobility increased and the Delta variant was spreading across Europe This study explored the possibility of an increased Covid-19 spread across Europe following EURO2020 matches METHODS Covid-19 data on cases vaccination and delta variant for participating countries host cities regions and neighboring countries for May till July 2021 were obtained from European Centre for Disease Prevention and Control Our World in Data Johns Hopkins COVID-19 Dashboard and the UK Government website EURO2020 data was obtained from the UEFA official website RESULTS A general increase in Covid-19 positivity trend in Europe was observed following a week of EURO2020 matches across most countries and host cities A similar trend was observed for the Delta variant sample positivity rate The increased incidence was mostly among the young generation <49 years A decline in positive cases was observed on a national level for most countries following the Finals match CONCLUSION The EURO2020 was an anticipated mass sports event and it was the first-time spectators were allowed to enter stadiums in Europe Stadiums instituted several mitigations to safeguard the spectators although reports of transmission were still present The major challenges were the gatherings outside the stadiums that might have contributed to these observations Targeted restrictions might be required during mass sport events especially in the presence of highly transmissible variants and low vaccination rates among the young generation', metadata='title': 'An observational study of the Covid-19 situation following the first pan-European mass sports event', 'url': 'https://pubmed.ncbi.nlm.nih.gov/35000189'), Document(page_content='In this article the authors argue that the response to the COVID-19 pandemic is primarily focused exclusively on the virus

and would benefit from a broadening of perspective to a more systemic approach After presenting in a deliberately caricatural way the opposition between the Pasteurian and holistic paradigms they present some limitations of the Pasteurian approach which is still dominant in the medical community They put forward two explanations for the perpetuation of this dominance relating to research methods and medical training They then apply their analysis to the COVID-19 response strategy arguing that the neglect of social determinants and of the role of co-morbidities - which play an essential role in the severity of the disease - constitutes a clear failure of public health It is thus deprived of a whole range of prevention and health promotion tools and even of early holistic care They conclude by suggesting that the COVID-19 crisis reveals the limits of the Pasteurian model and calls for a paradigm shift allowing for a more complex systemic and coherent approach to COVID-19 - and all other health problems', metadata='title': 'COVID-19 pandemic have we not reached the limits of the Pasteurian paradigm', 'url': 'https://pubmed.ncbi.nlm.nih.gov/34881826']]

Respuesta: Based on the provided documents, there have been significant impacts of COVID-19 on mortality in Madrid (Spain), as evidenced through excessive mortality estimation systems like "Momo." These studies suggest large increases in overall mortality related to COVID-19, particularly within older populations. Additionally, ongoing monitoring helps detect potential risks for public health concerns associated with the pandemic. Therefore, yes, COVID-19 remains relevant today in terms of affecting morbidity and mortality patterns in Madrid.

Puntuación: tabla **G.4**

-2º Ejemplo de validación por terceros

Prompt: which are the common syntoms of the covid?

Contexto: [INST] Instruction: You are an expert in COVID-19. Please answer the question based on your knowledge of abstracts from a series of articles published in PubMed: [Document(page_content='Since its onset the new coronavirus SARS-CoV-2 has manifested multiple paradoxes related to both complications of the infection and treatment resistance The experiences in therapy started from the use of neutralizing monoclonal antibodies and anti-cytokine therapy anti IL-6 anti-TNF-alpha and CCR5 up to experiments of immunomodulation with glucocorticoids and proposal of very powerful immunomodulatory drugs like cyclophosphamide in optimized diagrams in order to induce the depression of the immune hyperergic reaction with the maintenance of TH and TC lymphocyte resistance to keep the anti-viral activity in the absence of an etiological treatment In this context also the

second major immunologic mechanism could function the molecular mimicry - cross-reactivity between viral epitopes and various antigenic determinants from the body other than the tissue structures of the respiratory system. In a similar manner in the past genital staphylococcal infections among women wearing tampons sometimes generated a toxic-septic shock with an unexpected evolution to death. It is also about an atypical cytokine storm generated by the staphylococcal toxins super-antigenic but probably the aggressiveness was also generated by the cross reactivity with chemical antigens from the cotton fibers because this disorder generally appeared among women in the menstrual period sometimes the local genital infection being minor. Moreover with the lapse of time the reaction of shock no longer appeared probable through the increase of immunologic tolerance to cotton fibers. In a similar manner in COVID-19 a reaction of hyper-sensitivity to a gaseous antigen could overlap considering the decrease of tolerance at the invasive administration of oxygen the contact with BALT - the lymphatic system associated to the bronchia where the resident macrophage is the siderophage and partial maintenance at the non-invasive administration through mask or extracorporeal which induces the first contact with the lymphatic elements associated to the oropharyngeal cavity or blood where the siderophage can no longer be observed. In this context some adult patients with COVID-19 could manifest an immunologic genetic defect of TS at BALT level as women with staphylococcal toxic-septic shock had a local defect at the VALT level lymphatic tissue associated with the genital tract. In children the syndrome is rare and more multisystem which rather suggests the breaking of central tolerance through the lack of intra-thymus or intra-medullary deletion of the auto-reactive clones of Th TC while among adults the defect seems to be more peripheral local with the involvement of TS and resident macrophages siderophages. This could explain not only the higher level of severity in elder patients and those with cardiac disorders but also the rarity of the pathology in patients aged under nine years', metadata='title': 'COVID-19 Infection - the Possibility to Involve Oxygen as Immunogen in the ARDS Determinism Acute Respiratory Distress Syndrome', 'url': 'https://pubmed.ncbi.nlm.nih.gov/34925610'), Document(page_content='Before eliciting an adaptive immune response SARS-CoV-2 must overcome seven constitutive respiratory defense barriers. The first is the mucus covering the respiratory tract's luminal surface which entraps inhaled particles including infectious agents and eliminates them by mucociliary clearance. The second barrier comprises various components present in the airway lining fluid: the surfactants. Besides providing low surface tension that allows efficient gas exchange at the alveoli, surfactants inhibit the invasion of epithelial cells by respiratory viruses, enhance pathogen uptake by

phagocytes and regulate immune cells functions The respiratory tract microbiota constitutes the third defense barrier against SARS-CoV-2 It activates the innate and adaptive immune cells and elicits anti-infectious molecules such as secretory IgA antibodies defensins and interferons The fourth defense barrier comprises the antimicrobial peptides defensins and lactoferrin They show direct antiviral activity inhibit viral fusion and modulate the innate and adaptive immune responses Secretory IgA antibodies the fifth defense barrier besides protecting the local microbiota against noxious agents also inhibit SARS-CoV-2 cell invasion If the virus overcomes this barrier it reaches its target the respiratory epithelial cells However these cells also act as a defense barrier the sixth one since they hinder the virus access to receptors and produce antiviral and immunomodulatory molecules such as interferons lactoferrin and defensins Finally the sensing of the virus by the cells of innate immunity the last constitutive defense barrier elicits a cascade of signals that activate adaptive immune cells and may inhibit the development of productive infection The subject of the present essay is discussing these mechanisms', metadata='title': 'The seven constitutive respiratory defense barriers against SARS-CoV-2 infection', 'url': 'https://pubmed.ncbi.nlm.nih.gov/34932765'), Document(page_content='COVID-19 was caused by the original coronavirus severe acute respiratory syndrome associated coronavirus-2 SARS CoV2 which originated in Wuhan China COVID-19 had a large breakout of cases in early 2020 resulting in an epidemic that turned into a pandemic This quickly enveloped the global healthcare system The principal testing method for COVID-19 detection according to the WHO is reverse transcription polymerase chain reaction RT-PCR Isolation of patients quarantine masking social distancing sanitizer use and complete lockdown were all vital health-care procedures for everyone With the new normal and vaccination programmes the number of cases and recovered patients began to rise months later The easing of restrictions during the plateau phase resulted in a rebound of instances which hit the people with more ferocity and vengeance towards the start of April 2021 Coronaviruses have evolved to cause respiratory enteric hepatic and neurologic diseases resulting in a wide range of diseases and symptoms such as fever cough myalgia or fatigue shortness of breath muscle ache headache sore throat rhinorrhea hemoptysis chest pain nausea vomiting diarrhoea anosmia and ageusia Coronavirus infections can be mild moderate or severe in intensity COVID-19 pulmonary dysfunction includes lung edema ground-glass opacities surfactant depletion and alveolar collapse Patients who presented with gastrointestinal GI symptoms such as anorexia nausea vomiting or diarrhoea had a higher risk of negative outcomes COVID-19 s influence on cognitive function is one of COVID-19 s long-term effects More clinical situations need to be reviewed by healthcare professionals

so that an appropriate management protocol may be developed to reduce morbidity and death in future coming third fourth wave cases', metadata='title': 'Synopsis of symptoms of COVID-19 during second wave of the pandemic in India', 'url': 'https://pubmed.ncbi.nlm.nih.gov/34881534'), Document(page_content='The emergence of severe acute respiratory syndrome type 2 coronavirus SARS-CoV-2 and its complications have demonstrated the devastating impact of a new infectious pathogen The organisational change promulgated by the isolation of affected communities is of extreme importance to achieve effective containment of the contagion and good patient care The epidemiological study of the population of a small rural community in the North East of Italy revealed how much the virus had circulated during Spring 2020 and how contagion has evolved after a prolonged lockdown In the 1st phase NAAT Nucleic Acid Amplification Testing was performed in cases with more or less severe symptoms and a study was performed to trace the infection of family members Only 0 2 of the population tested positive on NAAT via nasopharyngeal swab during this 1st phase In the 2nd phase a random sample of the general population were tested for circulating anti-Sars-Cov-2 immunoglobulins This showed that approximately 97 9 of the population were negative while 2 1 with positive IgG at a distance of the population had contracted the virus in a mildly symptomatic or asymptomatic form The main symptom in subjects who developed immunity was fever Antibodies were found in subjects with forced coexistence with quarantined or infected subjects The mutual spatial distance by categories has shown higher relative prevalence of IgG positive and IgM negative cases in close proximity but also far from the infected with respect to an intermediate distance This suggests that subjects living in thinly populated areas could come in contact with the virus more likely due to intentional relational proximity while those living nearby could also be infected through random proximity', metadata='title': 'Epidemiological study in a small rural area of Veneto Italian region during Sars-Cov-2 Pandemia', 'url': 'https://pubmed.ncbi.nlm.nih.gov/34853349')]

Respuesta: According to the provided documents, COVID-19 can exhibit a variety of symptoms ranging from mild to severe. Some commonly reported symptoms include fever, cough, myalgia (muscle pain), fatigue, shortness of breath, loss of taste and smell (anosmia and ageusia respectively), headaches, sore throats, rhinosinusitis, conjunctival congestion, hoarse voice, abdominal pain, dizziness, joint pains, chills, sweats, skin discoloration, red eyes, confusion, delirium, seizures, difficulty waking up, bluish lips or face, breathing difficulties, high temperature above 38°C / 100.4 °F, persistent pain or pressure in the chest, heart palpitations, rapid pulse, anxiety, depression,

and even mental status changes leading to comas. Additionally, individuals experiencing gastrointestinal issues such as anorexia, nausea, vomiting, or diarrhea might experience increased risks of unfavorable outcomes. Long term impacts of Covid-19 include damage to organs like lungs, liver, kidneys, brain etc., affecting their functionality causing chronic conditions.

Puntuación: tabla **G.5**

-3º Ejemplo de validación por terceros

Prompt: How can we determine the costs generated by the COVID-19?

Contexto: [Document(page_content='BACKGROUND The COVID-19 pandemic has prompted hospitals to respond with stringent measures Accurate estimates of costs and resources used in outbreaks can guide evaluations of responses We report on the financial expenditure associated with COVID-19 the bed-days used for COVID-19 patients and hospital services displaced due to COVID-19 in a Singapore tertiary hospital METHODS We conducted a retrospective cost analysis from January to December 2020 in the largest public hospital in Singapore Costs were estimated from the hospital perspective We examined financial expenditures made in direct response to COVID-19 hospital admissions data related to COVID-19 inpatients and the number of outpatient and emergency department visits non-emergency surgeries inpatient days in 2020 compared with preceding years of 2018 and 2019 Bayesian time-series was used to estimate the magnitude of displaced services RESULTS USD 41.96 million was incurred in the hospital for COVID-19-related expenses Facilities set-up and capital assets accounted for 51.6% of the expenditure patient-care supplies comprised 35.1% Of the 19,611 inpatients tested for COVID-19 in 2020 727.3% had COVID-19 The total inpatient- and intensive care unit (ICU) -days for COVID-19 patients in 2020 were 8009 and 8 days respectively A decline in all hospital services was observed from February following a raised disease outbreak alert level most services quickly resumed when the lockdown was lifted in June CONCLUSION COVID-19 led to an increase in healthcare expenses and a displacement in hospital services Our findings are useful for informing economic evaluations of COVID-19 response and provide some information about the expected costs of future outbreaks', metadata={'title': 'Impact of the COVID-19 pandemic on a tertiary care public hospital in Singapore resources and economic costs', 'url': 'https://pubmed.ncbi.nlm.nih.gov/34902499'})], Document(page_content='OBJECTIVES COVID-19 has altered health sector capacity in low-income and middle-income countries LMICs Cost data to inform evidence-based priority setting are urgently needed Consequently in this paper we calculate the full economic health sector costs of COVID-19

clinical management in 79 LMICs under different epidemiological scenarios

METHODS We used country-specific epidemiological projections from a dynamic transmission model to determine number of cases hospitalisations and deaths over 1 year under four mitigation scenarios We defined the health sector response for three base LMICs through guidelines and expert opinion We calculated costs through local resource use and price data and extrapolated costs across 79 LMICs Lastly we compared cost estimates against gross domestic product GDP and total annual health expenditure in 76 LMICs

RESULTS COVID-19 clinical management costs vary greatly by country ranging between <0.1-12% of GDP and 0.4-223% of total annual health expenditure excluding out-of-pocket payments Without mitigation policies COVID-19 clinical management costs per capita range from US \$43.39 to US \$75.57 in 22 of 76 LMICs these costs would surpass total annual health expenditure In a scenario of stringent social distancing costs per capita fall to US \$1.10-US \$1.32

CONCLUSIONS We present the first dataset of COVID-19 clinical management costs across LMICs These costs can be used to inform decision-making on priority setting Our results show that COVID-19 clinical management costs in LMICs are substantial even in scenarios of moderate social distancing Low-income countries are particularly vulnerable and some will struggle to cope with almost any epidemiological scenario The choices facing LMICs are likely to remain stark and emergency financial support will be needed', metadata={'title': 'Stark choices exploring health sector costs of policy responses to COVID-19 in low-income and middle-income countries', 'url': 'https://pubmed.ncbi.nlm.nih.gov/34857521'}, Document(page_content='AIMS The COVID-19 pandemic has claimed the lives of more than 800 000 people in the United States US and has been estimated to carry a societal cost of 16 trillion over the next decade The availability of COVID-19 vaccines has had a profound effect on the trajectory of the pandemic with wide-ranging benefits We aimed to estimate the total societal economic value generated in the US from COVID-19 vaccines

METHODS We developed a population-based economic model informed by existing data and literature to estimate the total societal value generated from COVID-19 vaccines by avoiding COVID-19 infections as well as resuming social and economic activity more quickly To do this we separately estimated the value generated from life years saved healthcare costs avoided quality of life gained and US gross domestic product GDP gained under a range of plausible assumptions

RESULTS Findings from our base-case analysis suggest that from their launch in December 2020 COVID-19 vaccines were projected to generate 5.0 trillion in societal economic value for the US from avoided COVID-19 infections and resuming unrestricted social and economic activity more quickly Our scenario analyses suggest that the value could range between 1

8 and 9.9 trillion Our model indicates that the most substantial sources of value are derived from reduction in prevalence of depression 1.9 trillion gains to US GDP 1.4 trillion and lives saved from fewer COVID-19 infections 1.0 trillion LIMITATIONS Constructed as a projection from December 2020 our model does not account for the Delta or future variants nor does it account for improvements in COVID-19 treatment CONCLUSIONS The magnitude of economic benefit from vaccination highlights the need for coordinated policy decisions to support continued widespread vaccine uptake in the US', metadata='title': 'The societal economic value of COVID-19 vaccines in the United States', 'url': 'https://pubmed.ncbi.nlm.nih.gov/34989654'), Document(page_content='OBJECTIVE This study aims to estimate the cost of clinical management of COVID-19 infected patients based on their severity by exploring the resources used in health care provision in Myanmar METHODS A multicenter retrospective cost analysis of COVID-19 patients was performed using the micro-costing approach from the perspective of the health system It covered two cost components namely direct and indirect cost of treating a patient Input data and their quantities were obtained from COVID-19 Standard Treatment Guidelines of Ministry of Health and Sports and administrative and financial records of resource utilization of three designated health facilities in Yangon Region Valuation of these resources was based on the price list from the Procurement Section of the Ministry RESULTS This study estimated the unit cost of clinical management of COVID-19 infected patients with no symptom to be 953 552 MMK 717 USD with mild-moderate symptoms to be 1 155 222 MMK 869 USD and with severe-critically ill conditions to be 5 705 052 MMK 4290 USD Average cost for a patient per day was 86 687 MMK 65 USD for asymptomatic patients 105 020 MMK 79 USD for mild-moderate patients and 283 252 MMK 214 USD for severe-critically ill patients Since the first case detected till December 31 2020 COVID-19 clinical management cost was accounted for 139 Billion MMK 104 Million USD for total 124 630 confirmed cases CONCLUSIONS COVID-19 pandemic has caused health systems to incur the significant health care expenses Timely implementation of the sustainable affordable and efficient policy for COVID-19 responses is of utmost important for every nation especially in the face of a pandemic This study provides the fundamental inputs for strategic planning for future economic evaluations of different policy interventions and policy recommendations for health systems to remain resilient during and after the COVID-19 pandemic in Myanmar', metadata='title': 'Cost estimates of COVID-19 clinical management in Myanmar', 'url': 'https://pubmed.ncbi.nlm.nih.gov/34961536')]

Respuesta: There have been several studies, such as those cited above from publications in PubMed, which aim to quantify the costs generated by the COVID-19 pandemic. Methodologies include conducting retrospective cost analyses (from either the hospital or society perspectives), estimating costs based on epidemiologic models, and valuing resources utilized directly in providing medical care to affected individuals. Data collection involves obtaining input prices and quantity usage figures typically sourced from government agencies, standardized treatments protocols, and facility records. Ultimately, determining accurate costs requires careful consideration of various factors including infection rates, mortality levels, required medical equipment and personnel allocation, lengthy stays at hospitals, and potential disruptions to other essential healthcare procedures.

Puntuación: tabla **G.6**

Para las tres validaciones, las puntuaciones son las siguientes:

- **Primera validación:** 5, 5, 5, 4, 4, 3, 3, 2
- **Segunda validación:** 5, 3, 4, 3, 5, 4, 5, 5
- **Tercera validación:** 5, 3, 4, 3, 5, 4, 5, 5

A continuación, se presentan los cálculos de la media para cada métrica:

1. Relevancia del Contexto:

$$\text{Media} = \frac{5 + 5 + 5}{3} = 5,0 \quad (\text{G.1})$$

2. Exactitud de la Información:

$$\text{Media} = \frac{5 + 3 + 3}{3} = 3,67 \quad (\text{G.2})$$

3. Claridad de la Respuesta:

$$\text{Media} = \frac{5 + 4 + 4}{3} = 4,33 \quad (\text{G.3})$$

4. Exhaustividad:

$$\text{Media} = \frac{4 + 3 + 3}{3} = 3,33 \quad (\text{G.4})$$

5. Pertinencia:

$$\text{Media} = \frac{4 + 5 + 5}{3} = 4,67 \quad (\text{G.5})$$

6. Fluidez del Lenguaje:

$$\text{Media} = \frac{3 + 4 + 4}{3} = 3,67 \quad (\text{G.6})$$

7. Consistencia:

$$\text{Media} = \frac{3 + 5 + 5}{3} = 4,33 \quad (\text{G.7})$$

8. Referencias a Fuentes:

$$\text{Media} = \frac{2 + 5 + 5}{3} = 4,0 \quad (\text{G.8})$$

Métrica	Descripción	Escala de Puntuación
Relevancia del Contexto	Mide qué tan relevante es el contexto recuperado para el prompt dado.	1-5 (1: Muy irrelevante, 5: Muy relevante)
Exactitud de la Información	Evalúa la exactitud factual de la respuesta generada respecto a la información sobre COVID-19.	1-5 (1: Incorrecta, 5: Totalmente correcta)
Claridad de la Respuesta	Mide la claridad y coherencia de la respuesta proporcionada.	1-5 (1: Muy confusa, 5: Muy clara)
Exhaustividad	Evalúa si la respuesta cubre adecuadamente el tema planteado en el prompt.	1-5 (1: Muy incompleta, 5: Muy completa)
Pertinencia	Mide la pertinencia de la respuesta en el contexto de la pregunta.	1-5 (1: Irrelevante, 5: Muy pertinente)
Fluidez del Lenguaje	Evalúa la fluidez y naturalidad del lenguaje utilizado en la respuesta.	1-5 (1: Muy forzado, 5: Muy natural)
Consistencia	Mide la consistencia interna de la respuesta.	1-5 (1: Inconsistente, 5: Muy consistente)
Referencias a Fuentes	Evalúa si la respuesta incluye o se basa en referencias a fuentes confiables y actualizadas.	1-5 (1: Sin referencias, 5: Muy bien referenciada)

Tabla G.1: Tabla de puntuaciones para validar un modelo de lenguaje especializado en COVID-19.

Métrica	Nota
Relevancia del Contexto	5
Exactitud de la Información	5
Claridad de la Respuesta	3
Exhaustividad	5
Pertinencia	5
Fluidez del Lenguaje	2
Consistencia	4
Referencias a Fuentes	5

Tabla G.2: Notas de validación del modelo de lenguaje especializado en COVID-19.

Métrica	Nota
Relevancia del Contexto	5
Exactitud de la Información	5
Claridad de la Respuesta	3
Exhaustividad	5
Pertinencia	5
Fluidez del Lenguaje	2
Consistencia	4
Referencias a Fuentes	5

Tabla G.3: Notas de validación del modelo de lenguaje especializado en COVID-19.

Métrica	Nota
Relevancia del Contexto	3
Exactitud de la Información	4
Claridad de la Respuesta	5
Exhaustividad	2
Pertinencia	5
Fluidez del Lenguaje	5
Consistencia	4
Referencias a Fuentes	2

Tabla G.4: Notas de validación del modelo de lenguaje especializado en COVID-19.

Métrica	Nota
Relevancia del Contexto	5
Exactitud de la Información	5
Claridad de la Respuesta	5
Exhaustividad	4
Pertinencia	4
Fluidez del Lenguaje	3
Consistencia	3
Referencias a Fuentes	2

Tabla G.5: Notas de validación del modelo de lenguaje especializado en COVID-19.

Métrica	Nota
Relevancia del Contexto	5
Exactitud de la Información	3
Claridad de la Respuesta	4
Exhaustividad	3
Pertinencia	5
Fluidez del Lenguaje	4
Consistencia	5
Referencias a Fuentes	5

Tabla G.6: Notas de validación del modelo de lenguaje especializado en COVID-19.

Apéndice H

Anexo de sostenibilización curricular

H.1. Introducción

A día de hoy, para que un proyecto cumpla con unos mínimos de calidad, no sólo vale con que la parte técnica esté correctamente desarrollada sino que es necesario que cumpla ciertas pautas de sostenibilidad para que, a parte de ser un proyecto que cumpla un propósito tecnológico también sea respetuoso con el medio ambiente y promueva el desarrollo de la sociedad promoviendo valores tan importantes como la justicia social y la economía equitativa y viable a largo plazo.

A lo largo de este apéndice se defenderá este proyecto como uno que impulsa, no sólo a mejorar la salud pública sino que también tiene cómo objetivo entender cómo situaciones como la pandemia vivida en 2020 afecta de manera distinta a diferentes grupos sociales con el fin de romper desigualdades y concienciar sobre aquellos grupos más necesitados.

H.2. Competencias de sostenibilidad

Principio Ético

Las guías del principio ético dictan que el proyecto "debe esforzarse por educar a la ciudadanía reconociendo el valor intrínseco de cada persona, situando la libertad y la protección de la vida como objetivos de las políticas públicas y los comportamientos individuales". Esto se puede reflejar en este

proyecto ya que su ultimo fin es facilitar la labor a profesionales de la salud que luchan por la vida humana.

Principio Holístico

Este principio dice que "La universidad, en todas sus facetas, debe actuar desde una concepción integral e interdependiente de los componentes de la realidad social, económica y ambiental". Este proyecto busca informar del distinto impacto que la pandemia tuvo en gente que posee pocos recursos, dificultad de acceso al sistema sanitario y otras dificultades frente a gente más agraciada en ese aspecto intentando, mediante la información y divulgación, cerrar poco a poco las brechas entre ambos grupos. En el ámbito medioambiental el proyecto busca ahorrar en los grandes costes eléctricos que conllevan entrenar un modelo, reduciendo la huella energética producida sin que ello conlleve la pérdida de resultados.

Principio de Complejidad

El documento indica que este principio trata de "La adopción de enfoques sistémicos y transdisciplinares que permitan una mejor comprensión de la complejidad de las problemáticas sociales, económicas y ambientales, así como de la implicación en las mismas de todas las actividades ciudadanas y profesionales". Este principio es fácil de justificar debido a la yuxtaposición de campos, tanto de la informática como el de la salud, empleando técnicas de ambos ámbitos para mejorar la información y la salud pública.

Principio de Globalización

La guía indica que "La adopción de enfoques que establezcan relaciones entre los contenidos curriculares y las realidades locales y globales". Los abstracts recuperados en este proyecto, abarcan todo tipo de realidades y entornos afectados por esta brutal enfermedad, varios artículos hacen hincapié en los efectos que tiene el entorno social y económico a la hora de afrontar la infección.

Principio de Transversalidad

Se define este principio como: integración de los contenidos dirigidos a la formación de competencias para la sostenibilidad en las diversas áreas de conocimiento, asignaturas y titulaciones. Al igual que en el principio de Complejidad, se puede ver el impacto que distintas áreas tienen sobre el

proyecto por lo que, fácilmente, una persona con conocimientos en alguna de los ámbitos relacionados podría mejorar el proyecto ya sea actualizando y manteniendo la aplicación o mejorando y revisando los datos de entrenamiento. También sería de gran utilidad la ayuda de otras titulaciones, por ejemplo, en el marco legal.

Principio de Responsabilidad Social Universitaria

La definición que aporta la guía sobre este principio es la siguiente: Contribución de la Universidad a la sostenibilidad de la Comunidad. Se reflejará en la gestión interna y en la colaboración con entidades y organismos en proyectos de investigación y acciones que contribuyan a mejorar la calidad de la formación universitaria y el avance en la resolución de los problemas sociales, económicos y ambientales". Este proyecto puede ayudar a estudiantes que tengan problemas a la hora de entender algún aspecto relevante del Covid-19 aún cuando los profesores estén fuera del horario lectivo. Además y debido a la flexibilidad que aportan los prompts a la hora de interactuar con el modelo la respuesta se puede adecuar a las necesidades del alumno.[[CRUE , 2012](#)]

Bibliografía

- [Bil, 2014] (2014). ¿cuánto cuesta contratar a un trabajador? <https://www.billin.net/calculadora-contratar-trabajador/>.
- [Hug, 2016] (2016). Spaces. <https://huggingface.co/spaces>.
- [tal, 2024] (2024). Salario para Programador Junior en España - Salario Medio. <https://es.talent.com/salary>.
- [Chung et al., 2000] Chung, L., Nixon, B., Yu, E., and Mylopoulos, J. (2000). Non-functional requirements. *Software Engineering*, page 78.
- [CRUE , 2012] CRUE (2012). Directrices_{sostenibilidad}. (*originalworkpublished2005*).
- [Europea, 2024] Europea, C. (2024). Ley de IA | Configurar el futuro digital de Europa.
- [H, 2023] H, A. (2023). Abstracts of 10,000 Covid Research Papers.
- [Hayat et al., 2019] Hayat, F., Rehman, A. U., Arif, K. S., Wahab, K., and Abbas, M. (2019). The influence of agile methodology (scrum) on software project management. In *2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, pages 145–149.
- [Katz, 1985] Katz, M. J. (1985). *Elements of the Scientific Paper*. Yale University Press. Google-Books-ID: umYHoNJnygUC.
- [Ken Schwaber and Jeff Sutherland, 2020] Ken Schwaber and Jeff Sutherland (2020). Scrum Guide | Scrum Guides.

- [Malan et al., 2001] Malan, R., Bredemeyer, D., and Consulting, B. (2001). Functional requirements capture the intended behavior of the system. This behavior may be expressed as services, tasks or functions the system is required to perform. This white paper lays out important concepts and discusses capturing functional requirements in such a way that they can drive architectural decisions and be used to validate the architecture.
- [Meo, 2018] Meo, S. A. (2018). Anatomy and physiology of a scientific paper. *Saudi Journal of Biological Sciences*, 25(7):1278–1283.
- [Schwaber and Sutherland, 2020] Schwaber, K. and Sutherland, J. (2020). La guía definitiva de scrum: las reglas del juego.