

MedTrust: Asistente Virtual Ético para Atención Médica Personalizada

Enrique Ulises Báez Gómez Tagle
Gabriela Shaooli Cassab

23 de noviembre de 2024

Resumen

Este proyecto presenta el desarrollo de "MedTrust", un Asistente Virtual Ético para la Atención Médica Personalizada, diseñado específicamente para su uso en farmacias. Mediante el uso de inteligencia artificial generativa y recuperación aumentada por generación (RAG), el asistente tiene como objetivo proporcionar recomendaciones médicas personalizadas, gestionar citas y monitorear tratamientos de manera continua. El proyecto incorpora principios éticos clave como la transparencia, privacidad, equidad y responsabilidad, garantizando que las recomendaciones emitidas sean confiables, seguras y alineadas con las normativas de protección de datos y la Ley de Salud en México. Además, se implementan mecanismos para garantizar que el sistema sea comprensible tanto para pacientes como para profesionales de la salud. A diferencia de herramientas genéricas, MedTrust se distingue por integrar principios éticos, garantizando transparencia y equidad en cada recomendación emitida. Este proyecto busca innovar en el ámbito de la atención médica en farmacias, brindando soluciones tecnológicas que respeten los derechos de los pacientes y promuevan la confianza en el uso de la inteligencia artificial en el sector salud.

Índice

1. Resumen del proyecto abordado	3
1.1. Antecedentes y contexto del proyecto	3
1.2. Definición y justificación del problema	3
1.3. Descripción de la propuesta de proyecto final	3
2. Preocupaciones éticas más importantes	4
3. Principios Éticos y Normativos que debe cumplir el proyecto	5
3.1. Leyes aplicables en México	5
3.2. Políticas internas de la empresa	6
3.3. Principios éticos	6

3.4.	Criterios recomendables	7
3.5.	Control Humano, Responsabilidad profesional Y Prevención de sesgos	7
3.6.	Privacidad de Datos y Manejo de Información Sensible	8
4.	Análisis de riesgos y beneficios	9
4.1.	Identificación de los involucrados y sus roles	9
4.2.	Análisis de impactos éticos	9
4.3.	Análisis Modo de Fallos y Análisis de Efectos (FMEA)	11
4.4.	Diseño ético durante todo el ciclo de vida	12
4.4.1.	Fase de planificación y análisis	12
4.4.2.	Fase de diseño	12
4.4.3.	Fase de desarrollo	13
4.4.4.	Fase de implementación	13
4.4.5.	Fase de mantenimiento y operación continua	13
4.4.6.	Fase de actualización y mejora continua	14
5.	Documentación de datasets y modelos	14
5.1.	Modelos utilizados	14
5.2.	Datasets utilizados	15
5.3.	Almacenamiento y Preparación de los Datos	16
5.3.1.	Origen y Propósito de los Datos	16
5.3.2.	Almacenamiento de los Datos	16
5.3.3.	Preparación de los Datos	16
5.3.4.	Almacenamiento Vectorial	17
5.3.5.	Políticas de Eliminación de Datos	17

1. Resumen del proyecto abordado

1.1. Antecedentes y contexto del proyecto

Hoy en día, estamos en una época donde la automatización y las recomendaciones generadas por IA están constantemente evolucionando.

Actualmente, las farmacias tienen un papel crucial en la atención médica primaria, siendo el primer punto de contacto para muchas personas que buscan orientación sobre sus medicamentos o tratamientos. La integración de esta herramienta busca ofrecer personalización de tratamientos, el monitoreo continuo de la salud y la optimización en la gestión de citas y consultas. Sin embargo, cada vez es más importante enfrentar desafíos éticos como la protección y manejo adecuado de los datos personales de los pacientes, así como la precisión de las recomendaciones emitidas por el sistema.

Este proyecto pretende garantizar la salud del paciente sin comprometer sus derechos fundamentales, permitiendo a las farmacias ofrecer un servicio más confiable y ético.

1.2. Definición y justificación del problema

En la actualidad, muchas farmacias enfrentan dificultades para ofrecer servicios médicos más completos debido a limitaciones en la interacción con los pacientes, así como la falta de herramientas tecnológicas, éticas y confiables, las herramientas de IA en el ámbito médico carecen de los principios fundamentales de transparencia y responsabilidad.

A pesar de los avances tecnológicos, gran parte de los productos disponibles no cuentan con mecanismos robustos que aseguren recomendaciones precisas y fundamentadas, lo que podría comprometer la calidad del cuidado ofrecido. Además, la protección de la privacidad de los pacientes sigue siendo una preocupación crítica, ya que no siempre se garantiza el manejo adecuado de datos sensibles. A esto se suma la falta de claridad en cuanto al funcionamiento interno de estas herramientas, lo que dificulta la comprensión de cómo y por qué se generan ciertas sugerencias, afectando la confianza en su uso.

La creación de un asistente virtual enfocado en farmacias podría resolver estos problemas, al proporcionar recomendaciones precisas y personalizadas mientras garantiza la seguridad de datos y la equidad en el acceso a los servicios.

1.3. Descripción de la propuesta de proyecto final

Este proyecto se centra en desarrollar un Asistente Virtual Ético para Atención Médica Personalizada, específicamente diseñado para su uso en farmacias.

El asistente ayudará a las farmacias a ofrecer un servicio más completo. Utilizará inteligencia artificial generativa, recuperación aumentada por generación (RAG) y agentes inteligentes para ofrecer recomendaciones médicas, gestionar citas y monitorear el progreso de los tratamientos.

A diferencia de otras herramientas, este asistente integrará principios éticos

desde su concepción, priorizando la precisión en las recomendaciones, la protección rigurosa de la privacidad de los pacientes y la total transparencia en la toma de decisiones. Además, se implementarán mecanismos que permitan a los pacientes y profesionales de la salud comprender claramente el proceso detrás de cada recomendación, fomentando así la confianza y el uso responsable de la tecnología.

Este asistente se distingue de otras herramientas por integrar principios éticos desde su diseño, priorizando:

1. **Transparencia:** El asistente explicará claramente cómo llega a sus recomendaciones, para así ofrecer una comprensión detallada de los factores y procesos que influyen en cada sugerencia, lo que generará confianza y certeza sobre lo que sugiere la herramienta.
2. **Privacidad:** Se implementarán mecanismos de protección de datos para garantizar la confidencialidad de la información de los pacientes. Estos mecanismos aseguran que los datos sensibles no sean guardados en ningún momento, preservando así la privacidad del paciente.
3. **Responsabilidad:** Se establecerán mecanismos de monitoreo continuo del desempeño del asistente, para identificar y corregir rápidamente cualquier error o desvío en su funcionamiento principal. De esta manera, se asegurará que el asistente actúe de forma confiable y correcta y que las recomendaciones se mantengan alineadas con los principios éticos del proyecto.
4. **Integración con el sistema de farmacias:** Facilitando la recomendación de productos disponibles de acuerdo a los síntomas y gestión de citas.

2. Preocupaciones éticas más importantes

En relación con los principios éticos abordados en clase, las preocupaciones más importantes para este proyecto son las siguientes:

1. **Privacidad de los datos:** Debido al manejo de información sensible de los pacientes, es esencial que el asistente cumpla con estrictos estándares de privacidad, como lo estipula la Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados (LGPDPPSO). Esto incluye la encriptación de los datos médicos y la implementación de controles de acceso para evitar filtraciones o accesos no autorizados.
2. **Transparencia en las recomendaciones:** Los usuarios deben poder entender cómo se generan las sugerencias y decisiones del asistente. La transparencia fomenta la confianza en el uso de la herramienta y asegura que las recomendaciones sean claras y justificadas. Esto se logrará

proporcionando explicaciones accesibles sobre el razonamiento detrás de cada sugerencia médica, permitiendo que tanto los pacientes como los farmacéuticos comprendan el proceso.

3. ***Responsabilidad y control humano:*** Aunque el asistente utiliza inteligencia artificial para automatizar recomendaciones, debe haber supervisión humana para garantizar que los resultados sean éticos y seguros. El equipo humano debe estar capacitado para intervenir si es necesario y auditar el sistema de manera regular para corregir cualquier desviación o error que pueda surgir en las recomendaciones.
4. ***Equidad en el acceso:*** El asistente debe garantizar que todas las personas, independientemente de su situación económica, ubicación geográfica, o acceso a recursos tecnológicos, tengan acceso a las recomendaciones y servicios médicos. La equidad es un principio fundamental para evitar la exclusión social o económica, y se deben evitar sesgos en la IA que puedan discriminar a ciertos grupos de la población.
5. ***Consentimiento informado:*** Antes de usar el asistente, los pacientes deben dar su consentimiento explícito y estar completamente informados de cómo se manejarán sus datos. Esto implica que el asistente debe explicar claramente los términos de uso, incluyendo cómo se almacenan los datos, por cuánto tiempo, y con qué propósito, garantizando la autonomía del paciente.
6. ***Minimización de riesgos:*** Es importante diseñar el sistema de manera que minimice cualquier riesgo asociado con la toma de decisiones automatizadas. El sistema debe identificar, evaluar y mitigar riesgos como errores en las recomendaciones médicas, fallos técnicos, o vulnerabilidades de seguridad que puedan comprometer la información sensible. Esto incluye auditorías frecuentes y protocolos de respuesta rápida ante cualquier problema detectado.

3. Principios Éticos y Normativos que debe cumplir el proyecto

El proyecto se alineará con ciertos criterios y principios éticos y normativos, especialmente en el contexto del manejo de información médica sensible y el uso de inteligencia artificial. A continuación, se detallan los principales puntos que el asistente debe cumplir en el contexto de México:

3.1. Leyes aplicables en México

El proyecto debe cumplir con las siguientes leyes y normativas:

1. ***Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados (LGPDPSO):*** Regula el tratamiento de datos

personales, especialmente en sectores sensibles como la salud, garantizando que los datos de los pacientes sean manejados de manera segura y confidencial.

2. ***Ley de Salud de México:*** Establece los principios bajo los cuales se puede utilizar tecnología en el ámbito médico, garantizando que la atención a la salud sea ética y cumpla con las normativas de transparencia, privacidad y responsabilidad.

3.2. Políticas internas de la empresa

Al implementar este sistema, se deben revisar las siguientes políticas para contar con ellas y aplicarlas:

1. ***Políticas de acceso y manejo de datos:*** Definir quién puede acceder a la información del asistente y cómo se manejarán los datos de los pacientes para garantizar su privacidad y seguridad.
2. ***Políticas de responsabilidad en el uso del asistente:*** Asegurarse de que las recomendaciones proporcionadas por el asistente estén siempre supervisadas por un profesional médico capacitado.
3. ***Actualización y mantenimiento continuo:*** La empresa debe asegurar que el asistente sea revisado y actualizado de manera constante, para evitar errores o fallos en las recomendaciones.

3.3. Principios éticos

La empresa debe asegurarse de que el asistente cumpla con los siguientes principios éticos:

1. ***Transparencia:*** Los usuarios deben entender cómo se generan las recomendaciones y poder acceder a una explicación clara y detallada de las decisiones del sistema.
2. ***Privacidad:*** Los datos de los pacientes deben ser protegidos de acuerdo con las leyes aplicables y no deben ser almacenados innecesariamente.
3. ***Equidad:*** Asegurar que todos los usuarios, independientemente de su condición socioeconómica o localización geográfica, tengan acceso a las recomendaciones y servicios del asistente.
4. ***Responsabilidad:*** La empresa debe ser responsable de cualquier recomendación emitida por el asistente, supervisando su correcto funcionamiento y corrigiendo posibles fallos.

3.4. Criterios recomendables

Se recomienda que la empresa establezca:

1. ***Equipo de revisión ética:*** Un equipo que se encargue de auditar constantemente el funcionamiento del asistente y asegurar que se respeten los principios éticos.
2. ***Mecanismos de retroalimentación de usuarios:*** Permitir que los usuarios puedan enviar quejas o sugerencias sobre el uso del asistente, lo que facilitará una mejora continua.
3. ***Capacitación continua:*** El personal que supervise y maneje el asistente debe estar capacitado para entender y actuar sobre las recomendaciones emitidas por el sistema.

3.5. Control Humano, Responsabilidad profesional Y Prevención de sesgos

En el desarrollo de MedTrust, se integran principios de control humano, responsabilidad profesional, y la identificación de posibles sesgos, con el objetivo de garantizar un sistema seguro, transparente y confiable en la atención médica.

1. ***Control humano:*** A pesar de la automatización avanzada mediante inteligencia artificial, es fundamental mantener la supervisión humana de las recomendaciones emitidas por MedTrust. Para ello, se establecerá la participación activa de profesionales médicos y personal farmacéutico, quienes deberán validar y ajustar las decisiones automatizadas del sistema cuando sea necesario. Este enfoque asegura que cualquier desvío o recomendación crítica sea revisada y aprobada por expertos, proporcionando una capa adicional de confianza y seguridad para los pacientes. Además, el personal médico actuará como responsable de auditar de forma periódica los resultados generados por la IA y de intervenir en decisiones complejas que requieran juicio clínico.
2. ***Responsabilidad profesional:*** Los médicos y farmacéuticos involucrados en el funcionamiento del asistente virtual asumen la responsabilidad directa de validar las recomendaciones antes de su implementación. Esta responsabilidad implica no solo la verificación continua de las sugerencias proporcionadas, sino también la capacidad de corregir rápidamente cualquier error o recomendación inapropiada que pueda surgir. Pues para ofrecer un servicio seguro, confiable y que alineado con las mejores prácticas clínicas necesitamos una supervisión constante y un compromiso activo para lograr el estándar de calidad deseado para la atención médica.
3. ***Prevención de sesgos:*** La presencia de sesgos implícitos en los modelos de inteligencia artificial puede llevar a resultados injustos y desiguales

en las recomendaciones médicas. Por ello, para MedTrust se buscan auditorías constantes del sistema, revisando tanto los datos utilizados como el modelo empleado para detectar y eliminar cualquier tipo de sesgo que pueda surgir debido a diferencias de género, etnia, condición socioeconómica, entre otros. Esta revisión constante se enfoca en asegurar la equidad, precisión y justicia en las recomendaciones para todos los pacientes, sin importar su contexto.

3.6. Privacidad de Datos y Manejo de Información Sensible

La privacidad es uno de los pilares éticos y normativos más importantes en el diseño y funcionamiento de MedTrust, ya que implica el manejo de información altamente sensible relacionada con la salud de los pacientes. En este proyecto, se implementan las siguientes medidas clave para garantizar la confidencialidad y seguridad de los datos personales:

1. **Minimización de Datos:** Se recolecta únicamente la información estrictamente necesaria para el funcionamiento del asistente, evitando la recopilación de datos irrelevantes o excesivos. Los datos sensibles no se almacenan localmente, únicamente se almacena la información mínima en la nube y la mayoría de los datos se usa únicamente durante la sesión activa.
2. **Encriptación Avanzada:** Todos los datos transmitidos entre el asistente y los usuarios están protegidos mediante protocolos de encriptación robustos, como TLS (Transport Layer Security), para prevenir accesos no autorizados durante la comunicación.
3. **Anonimización:** Los datos personales son anonimizados antes de cualquier análisis o procesamiento adicional, asegurando que no puedan ser vinculados a un individuo específico.
4. **Control de Acceso:** Se implementan sistemas de control de acceso estrictos para garantizar que solo personal autorizado tenga permisos para interactuar con la información sensible. El acceso está limitado a roles específicos y se registra mediante auditorías regulares.
5. **Cumplimiento Legal:** El proyecto se ajusta a las normativas de la *Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados (LGPDPPSO)* en México, y se alinearán con otros estándares internacionales, como el Reglamento General de Protección de Datos (GDPR) si es necesario en el futuro.
6. **Gestión de Consentimiento:** Se asegura que los usuarios den su consentimiento informado antes de proporcionar cualquier dato personal, explicando de forma clara cómo serán utilizados sus datos.

4. Análisis de riesgos y beneficios

4.1. Identificación de los involucrados y sus roles

En este proyecto, las principales personas involucradas y sus roles son los siguientes:

1. **Pacientes:** Usuarios finales que recibirán recomendaciones personalizadas y tendrán acceso a servicios médicos a través del asistente. Son los principales afectados en términos de privacidad, seguridad y calidad de las recomendaciones.
2. **Farmacéuticos:** Personal encargado de implementar y supervisar el uso del asistente en las farmacias. Tienen un rol crucial en la validación de las recomendaciones y la atención al paciente.
3. **Médicos:** Supervisores responsables de garantizar que las recomendaciones médicas del asistente sean seguras, precisas y acordes con los estándares médicos.
4. **Desarrolladores de IA:** Encargados del diseño, desarrollo e implementación del asistente, asegurando que el sistema cumpla con los estándares éticos y funcione correctamente.
5. **Reguladores de la salud:** Instituciones encargadas de garantizar que el asistente opere de acuerdo con las leyes y normativas mexicanas, como la Secretaría de Salud y el INAI.
6. **Especialistas en ética:** Aseguran que los principios éticos se mantengan a lo largo del desarrollo y uso del asistente, supervisando posibles dilemas éticos que puedan surgir.

4.2. Análisis de impactos éticos

Se consideran tanto los beneficios como los riesgos que el asistente podría generar:

1. **Beneficios:**
 - **Mejora en el acceso a servicios médicos:** El asistente podría aumentar la accesibilidad a servicios médicos personalizados en farmacias, especialmente en áreas donde el acceso a consultas médicas es limitado.
 - **Monitoreo de salud:** El asistente permitirá un monitoreo continuo de la salud de los pacientes, lo que mejorará la prevención de enfermedades y el seguimiento de tratamientos.
 - **Optimización de la gestión de citas:** La gestión automática de citas y consultas reducirá tiempos de espera y facilitará el acceso a servicios médicos.

2. **Riesgos Identificados y Estrategias de Mitigación** El desarrollo de MedTrust considera los posibles riesgos relacionados con la privacidad y seguridad, así como aquellos asociados con el uso de inteligencia artificial en un contexto médico. A continuación, se identifican los riesgos más relevantes y las estrategias adoptadas para mitigarlos:

- **Riesgo de Filtración de Datos:** *Impacto:* Exposición de datos sensibles podría comprometer la confianza de los usuarios y la integridad del sistema.
Mitigación: Implementación de auditorías de seguridad periódicas para identificar vulnerabilidades.
- **Riesgo de Recomendaciones Erróneas:** *Impacto:* Un error en las recomendaciones podría afectar la salud del paciente, generando consecuencias graves.
Mitigación: Validación constante por parte de médicos y farmacéuticos de las recomendaciones emitidas. Implementación de un sistema de retroalimentación para reportar recomendaciones imprecisas y ajustar el algoritmo.
- **Riesgo de Sesgos en la IA:** *Impacto:* Los modelos de IA podrían perpetuar sesgos existentes, afectando la equidad en las recomendaciones.
Mitigación: Realización de pruebas exhaustivas para identificar y corregir sesgos. Inclusión de información diversa de los medicamentos y validación de los mismos.
- **Dependencia Tecnológica:** *Impacto:* Los usuarios podrían confiar excesivamente en el asistente y reducir su interacción con profesionales de la salud.
Mitigación: Mensajes claros que indiquen que el asistente no reemplaza a un médico, sino que complementa la atención profesional. Supervisión constante de los farmacéuticos y médicos como intermediarios.
- **Vulnerabilidad a Ciberataques:** *Impacto:* Posibilidad de acceso no autorizado a los sistemas de MedTrust.
Mitigación: Implementación de sistemas de detección y respuesta rápida a ciberataques. Mantenimiento actualizado del software y componentes de la nube usados en el sistema..
- **Falta de Comprensión del Usuario:** *Impacto:* Los usuarios podrían no entender cómo se utilizan sus datos o cómo se generan las recomendaciones.
Mitigación: Interfaces accesibles y fáciles de entender que expliquen los procesos detrás de cada sugerencia médica. Guías informativas disponibles en todo momento dentro de la herramienta.

4.3. Análisis Modo de Fallos y Análisis de Efectos (FMEA)

Se realizará un análisis para identificar los posibles fallos en el funcionamiento del asistente y los efectos que estos fallos tendrían en los pacientes y el sistema médico en general. Se llevarán a cabo las siguientes acciones:

1. **Identificación de fallos:** Cada componente del sistema (recomendaciones, gestión de citas, monitoreo de salud) será evaluado para detectar posibles puntos de fallo.
2. **Evaluación del impacto:** Se analizará el impacto de cada fallo en términos de seguridad del paciente, pérdida de datos o mala recomendación médica.
3. **Plan de mitigación:** Para cada fallo identificado, se desarrollará un plan de contingencia que permita minimizar los riesgos, como auditorías continuas, pruebas automatizadas y revisión periódica del sistema por parte de médicos.

Fallo Detectado	Impacto	Mitigación
Filtración de datos sensibles	Pérdida de confianza por parte de los usuarios y posible incumplimiento legal.	Implementación de auditorías periódicas, encriptación avanzada, y uso de la nube.
Recomendaciones erróneas	Riesgo de afectaciones a la salud del paciente.	Validación constante por médicos y farmacéuticos, y sistema de retroalimentación para ajustar el algoritmo.
Sesgos en la IA	Inequidad en las recomendaciones médicas, discriminación hacia ciertos grupos.	Auditorías para detectar sesgos en los datos y en las recomendaciones.
Dependencia tecnológica	Menor interacción con profesionales de la salud.	Mensajes que aclaren que el asistente es complementario y no sustituye al médico. Supervisión constante de personal médico.
Vulnerabilidad a ciberataques	Posible acceso no autorizado a los datos del sistema.	Implementación de sistemas de detección y respuesta rápida, actualizaciones regulares.
Falta de comprensión del usuario	Mal uso del sistema o desconfianza en la herramienta.	Interfaces intuitivas y guías explicativas accesibles. Capacitación constante al personal.

Cuadro 1: Análisis Modo de Fallos y Análisis de Efectos (FMEA).

4.4. Diseño ético durante todo el ciclo de vida

El diseño ético no fue una etapa aislada del desarrollo, sino un enfoque que se integró en cada fase del ciclo de vida del “Asistente Virtual Ético para Atención Médica Personalizada”. Esto garantizó que el sistema no solo cumpliera con los requisitos funcionales, sino que también mantuviera los más altos estándares éticos desde la concepción hasta el mantenimiento continuo. A continuación se detallan las consideraciones éticas en cada etapa del ciclo de vida del proyecto:

4.4.1. Fase de planificación y análisis

Desde la fase inicial de planificación, se definieron los objetivos éticos de la herramienta, alineando estos con las normativas legales y los principios éticos fundamentales. En esta fase se:

- **Identificaron riesgos éticos:** Durante el análisis inicial, se previeron los posibles riesgos éticos, como la protección de datos personales, la transparencia en el funcionamiento del asistente y la posible dependencia excesiva de los usuarios en la IA.
- **Establecieron principios éticos:** Se definieron principios como la equidad, privacidad, responsabilidad y transparencia que guiaron todo el proceso de desarrollo del sistema.
- **Consultó con especialistas:** Se involucró a especialistas en ética, médicos y reguladores desde el principio para asegurarse de que los objetivos del proyecto fueran compatibles con los estándares éticos y normativos del sector.

4.4.2. Fase de diseño

En la fase de diseño del sistema, se implementaron las decisiones tomadas durante la planificación. Esto implicó:

- **Privacidad desde el diseño:** Se incorporaron medidas de privacidad desde el inicio, como el uso de técnicas de anonimización de datos y la minimización del almacenamiento de información sensible.
- **Transparencia en las decisiones:** El diseño del sistema garantizó que los usuarios pudieran comprender cómo se generaron las recomendaciones del asistente. Esto incluyó interfaces que explicaban los criterios utilizados por el sistema de IA para emitir recomendaciones.
- **Accesibilidad y equidad:** Se aseguró que el asistente estuviera diseñado para ser accesible a todos los usuarios, sin importar su nivel socioeconómico o localización geográfica. El diseño evitó sesgos que pudieran limitar el acceso de ciertos grupos a los servicios médicos.

4.4.3. Fase de desarrollo

Durante el desarrollo, se aplicaron los principios éticos en la implementación de cada componente del sistema:

- **Código ético:** Se siguió un código ético en el desarrollo del sistema, asegurando que los modelos de IA fueran imparciales, seguros y responsables. Esto incluyó la revisión constante del código para evitar que se introdujeran sesgos no deseados.
- **Pruebas de seguridad:** Se realizaron pruebas exhaustivas para garantizar que el sistema no fuera vulnerable a ataques que pudieran comprometer los datos personales de los pacientes.
- **Control de calidad:** Se implementaron controles de calidad que evaluaron no solo el desempeño técnico, sino también el cumplimiento ético, verificando que las recomendaciones emitidas fueran consistentes con los principios establecidos.

4.4.4. Fase de implementación

En la fase de implementación, se garantizó que las políticas y mecanismos éticos establecidos se cumplieran efectivamente:

- **Consentimiento informado:** Se aseguró que los pacientes comprendieran y aceptaran los términos de uso del asistente, incluidas las políticas de privacidad y el tratamiento de sus datos personales.
- **Auditorías éticas:** Se realizaron auditorías previas y posteriores a la implementación para verificar que el sistema operara conforme a los principios éticos y legales definidos. Estas auditorías fueron continuas a lo largo del ciclo de vida del sistema.
- **Capacitación del personal:** Se capacitaron a farmacéuticos, médicos y demás personal en el uso del asistente, asegurando que comprendieran las implicaciones éticas y técnicas del sistema.

4.4.5. Fase de mantenimiento y operación continua

Una vez que el asistente estuvo en funcionamiento, fue importante:

- **Revisión continua:** Se realizaron evaluaciones periódicas para asegurar que el sistema siguiera cumpliendo con los principios éticos, especialmente a medida que se incorporaron nuevas funcionalidades.
- **Mitigación de riesgos:** Se detectaron y corrigieron los fallos éticos que surgieron durante la operación del sistema ya en manos del usuario, específicamente recomendaciones inexactas.

- **Transparencia continua:** Se garantizó que los usuarios pudieran seguir comprendiendo el proceso detrás de las recomendaciones, implementando mejoras en las explicaciones o interfaces del sistema según fue necesario.
- **Involucramiento del equipo ético:** Se mantuvo la participación de especialistas en ética y reguladores para auditar de forma continua el funcionamiento del sistema, ajustándolo a nuevas normativas o principios que pudieron surgir.

4.4.6. Fase de actualización y mejora continua

En la fase de actualización, tanto para mejorar el sistema como para añadir nuevas personalidades, se mantuvo el enfoque ético:

- **Evaluación de nuevas funcionalidades:** Cada nueva funcionalidad que se incorporó al asistente fue evaluada desde una perspectiva ética, asegurando que no comprometiera los derechos de los pacientes o los principios del sistema.
- **Retroalimentación de usuarios:** Se incorporaron los comentarios de los usuarios en la mejora continua del asistente, prestando especial atención a cualquier preocupación ética que pudieron haber tenido.

5. Documentación de datasets y modelos

5.1. Modelos utilizados

Para el desarrollo de esta herramienta, se utilizó el modelo Claude 3.5 Sonnet, un modelo de lenguaje a gran escala (LLM) desarrollado por Anthropic. Claude 3.5 Sonnet es un modelo altamente capaz que ofrece mejoras en razonamiento, procesamiento visual y generación de texto en comparación con sus predecesores. A continuación, se detallan sus características principales:

- **Modelo:** Claude 3.5 Sonnet, desarrollado por Anthropic, es una evolución del modelo Claude 3 Sonnet. Este modelo se destaca por su capacidad mejorada en tareas de razonamiento y procesamiento, así como por su precisión en dominios especializados como la medicina y las ciencias.
- **Evaluaciones:** Claude 3.5 Sonnet ha sido evaluado en una serie de benchmarks estándar de la industria, como GPQA para preguntas y respuestas a nivel de posgrado y MMLU para razonamiento general. El modelo ha superado los estándares previos en la mayoría de las pruebas, lo que garantiza su capacidad para generar respuestas precisas en el contexto médico.
- **Seguridad:** El modelo ha sido sometido a rigurosas pruebas de seguridad, incluyendo evaluaciones en riesgos cibernéticos y autónomos, y ha sido clasificado con un nivel de seguridad ASL-2, lo que indica que no presenta riesgos catastróficos.

5.2. Datasets utilizados

Como el modelo Claude 3.5 Sonnet ya estaba pre-entrenado, no se realizaron entrenamientos adicionales en datasets personalizados. En cambio, se utilizó el modelo tal y como fue proporcionado por Anthropic, con sus capacidades previamente ajustadas para tareas de razonamiento y comprensión de lenguaje natural.

Para optimizar el rendimiento del modelo y adaptar sus respuestas al propósito específico del asistente, se desarrollaron e implementaron dos herramientas clave:

- **Guardrails personalizados:** Se diseñaron una serie de guardrails (límites y directrices de seguridad) específicos para garantizar que las respuestas del modelo fueran consistentes, éticas y alineadas con las normas médicas. Estos guardrails incluyeron:
 - Filtrado de información médica para garantizar que las respuestas estén alineadas con fuentes oficiales, proporcionando datos precisos y relevantes sin interpretaciones indebidas.
 - Respuestas ajustadas a la información contenida en los documentos de la CIMA, asegurando consistencia y fidelidad a las recomendaciones oficiales.
 - Restricciones para evitar sugerencias de diagnósticos o tratamientos que excedan el alcance de las recomendaciones autorizadas.
- **Prompts personalizados:** Se desarrollaron prompts cuidadosamente diseñados para optimizar las interacciones con el modelo. Estos prompts establecieron el contexto adecuado para que el modelo comprendiera mejor las preguntas de los usuarios y recuperara la información más relevante. Los prompts incluyeron:
 - Indicaciones detalladas para guiar al modelo a utilizar exclusivamente información proveniente de los documentos de la CIMA.
 - Plantillas contextuales específicas que facilitaron el análisis semántico de las consultas, asegurando respuestas más precisas y completas.
 - Adaptaciones dinámicas para ajustar el tono y el nivel de detalle de las respuestas, según las necesidades del usuario.

La implementación de estos guardrails y prompts personalizados permitió maximizar las capacidades del modelo Claude 3.5 Sonnet, garantizando una interacción segura, confiable y adaptada al dominio médico.

5.3. Almacenamiento y Preparación de los Datos

Para la implementación de este proyecto, se emplearon documentos en formato PDF extraídos directamente de la CIMA (Base de Datos de Medicamentos de la Agencia Española de Medicamentos y Productos Sanitarios). A continuación, se detallan los procesos de almacenamiento y preparación de los datos utilizados:

5.3.1. Origen y Propósito de los Datos

La información utilizada proviene exclusivamente de documentos PDF proporcionados por la CIMA, gestionada por la AEMPS (Agencia Española de Medicamentos y Productos Sanitarios). El propósito principal de utilizar estos documentos es garantizar que el asistente médico disponga de información actualizada y precisa sobre los medicamentos.

5.3.2. Almacenamiento de los Datos

Todos los documentos PDF extraídos de la CIMA se almacenan en Amazon S3, un servicio de almacenamiento escalable y seguro. Para asegurar la confidencialidad e integridad de los datos, los buckets de S3 están cifrados utilizando un servicio de gestión de claves criptográficas (AWS KMS). Esta clave garantiza que los datos almacenados estén protegidos contra accesos no autorizados.

- **Ubicación de S3:** Los datos se almacenan en un bucket configurado con cifrado en reposo y cifrado en tránsito.
- **Cifrado:** Se utiliza AWS Key Management Service (KMS) para gestionar las claves de cifrado, asegurando que todos los datos en S3 estén protegidos mediante cifrado AES-256.
- **Políticas de Seguridad y Red:** Los buckets de S3 están configurados con políticas estrictas de acceso, permitiendo únicamente a roles específicos dentro de la cuenta de AWS acceder a los datos. Además, se implementan políticas de red que restringen el acceso a los buckets desde direcciones IP no autorizadas.

5.3.3. Preparación de los Datos

La preparación de los datos se realizó mediante un proceso de **chunking** jerárquico, que permite dividir los documentos PDF en fragmentos más manejables para su procesamiento y análisis. Este enfoque facilita la indexación y recuperación eficiente de la información relevante.

- **Estrategia de Chunking:** Se utilizó una estrategia de *Hierarchical Chunking*, que organiza los datos en una estructura jerárquica, optimizando la búsqueda y el análisis de información específica dentro de los documentos.

- **Parsing del Modelo:** Para el procesamiento de los fragmentos de datos, se empleó el mismo modelo Claude 3.5 Sonnet utilizado en la generación de texto. Este modelo se encargó de interpretar y extraer la información relevante de cada chunk, asegurando una comprensión precisa del contenido.
- **Generación de Embeddings:** Una vez procesados los datos, se generaron embeddings utilizando el modelo Titan Embeddings G1 - Textv1.2, con dimensiones vectoriales de 1536. Estos embeddings permiten la representación numérica de los datos textuales, facilitando su integración en la base de datos vectorial.

5.3.4. Almacenamiento Vectorial

Los embeddings generados se almacenan en una base de datos vectorial serverless proporcionada por Amazon OpenSearch. Este almacenamiento está configurado con políticas de seguridad y red estrictas para garantizar la protección y disponibilidad de los datos.

- **Modelo de Embeddings:** Titan Embeddings G1 - Textv1.2.
- **Vector Store:** Se utilizó una base de datos vectorial en Amazon OpenSearch Serverless.
- **Seguridad del Vector Store:** La base de datos vectorial está protegida mediante políticas de acceso que restringen la interacción únicamente a servicios autorizados dentro de la red privada de la cuenta. Además, se habilitó el cifrado en tránsito y en reposo para todos los datos almacenados.

5.3.5. Políticas de Eliminación de Datos

Para cumplir con las normativas de privacidad y seguridad de datos, se estableció una política de eliminación de datos que asegura la eliminación permanente de la información cuando ya no es necesaria para el funcionamiento del asistente. Esta política se implementa tanto en los buckets de S3 como en la base de datos vectorial de OpenSearch.

- **Política de Eliminación:** Todos los datos almacenados se eliminan de manera segura tras un periodo de retención definido, utilizando mecanismos de borrado permanente proporcionados por AWS.

Aunque no se realizaron entrenamientos personalizados para este proyecto, el uso de Claude 3.5 Sonnet y la integración con la base de datos vectorial almacenada en Amazon OpenSearch Serverless permitieron desarrollar un asistente médico capaz de ofrecer recomendaciones éticas, precisas y confiables. La implementación de medidas de seguridad avanzadas y la gestión adecuada de los datos aseguran que la herramienta cumpla con los estándares más altos de protección y eficiencia.

Referencias

Normativas y Regulaciones

1. **Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados (LGPDPPSO):** Disponible en: <https://www.diputados.gob.mx/LeyesBiblio/pdf/L>
2. **Reglamento General de Protección de Datos (GDPR):** Disponible en: <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
3. **Ley Federal de Protección de Datos Personales en Posesión de los Particulares (México):** Disponible en: <https://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf>
4. **HIPAA (Health Insurance Portability and Accountability Act):** Normativa de EE.UU. sobre protección de datos médicos. Disponible en: <https://www.hhs.gov/hipaa/index.html>

Ética e IA en Salud

1. **Ethics Guidelines for Trustworthy AI (European Commission):** Disponible en: <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>
2. **IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems:** Disponible en: <https://ethicsinaction.ieee.org>
3. Floridi, L., & Cowls, J. (2019). .^A Unified Framework of Five Principles for AI in Society.” *Harvard Data Science Review*. DOI. Disponible en: <https://doi.org/10.1162/99608f92.8cd550d1>
4. Topol, E. J. (2019). ”High-performance medicine: the convergence of human and artificial intelligence.” *Nature Medicine*, 25(1), 44-56. DOI. Disponible en: <https://doi.org/10.1038/s41591-018-0300-7>

Técnicas y Herramientas de IA

1. Anthropic. Claude 3.5 Sonnet: Technical Overview.” Disponible en: <https://www.anthropic.com>
2. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer. ISBN: 978-0-387-31073-2.
3. Chollet, F. (2018). *Deep Learning with Python*. Manning Publications. ISBN: 978-1617294433.

Evaluación de Riesgos y FMEA

1. **”Failure Mode and Effects Analysis (FMEA): Principles and Best Practices”:** American Society for Quality. Disponible en: <https://asq.org>
2. Stamatis, D. H. (2003). *Failure Mode and Effect Analysis: FMEA from Theory to Execution*. ASQ Quality Press. ISBN: 978-0873895989.

Privacidad de Datos e Impactos Sociales

1. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). "The ethics of algorithms: Mapping the debate." *Big Data & Society*, 3(2), 2053951716679679. DOI: <https://doi.org/10.1177/2053951716679679>
2. Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs. ISBN: 978-1610395694.