



Máster en Data Science & Big Data Analytics (MDS+8)

“Creación de un chatbot para seguros de autos de Mapfre mediante generación aumentada de recuperación (RAG)”

Grupo 3 - MAPFRE:

Enrique Bonet Bailén

David Garrido Jiménez

Miguel Molina Álvarez

Andrea Vásquez Sánchez

0. ÍNDICE

1.	Punto de partida	4
2.	Estado de la cuestión	5
2.1.	Soluciones actuales y sus limitaciones.....	5
2.2.	Avances recientes en la tecnología.....	5
2.3.	Propuesta de este proyecto.....	6
2.4.	Impacto esperado	6
3.	El problema	6
3.1.	¿Cómo afecta a mapfre y por qué es un problema?.....	6
3.2.	¿Por qué es importante resolverlo?	7
4.	La solución.....	7
5.	Recolección y tratamiento de datos	7
5.1.	Identificación de las fuentes de datos	8
5.2.	Carga de documentos	8
5.3.	Limpieza y segmentación de la información.....	8
5.4.	Generación y persistencia de embeddings	8
5.5.	Almacenamiento y recuperación del índice.....	8
5.6.	Verificación de existencia del índice	9
5.7.	Integración con modelos externos	9
6.	Desarrollo del proyecto	9
6.1.	Procesamiento de imágenes con “geminimultimodal”	9
6.2.	Configuración y carga de la api key.....	9
6.3.	Modelo de datos para información vehicular.....	9
6.4.	Generación de embeddings con geminimultimodal.....	10
6.5.	Integración en la aplicación	10
6.6.	Configuración de la aplicación	10
6.7.	Procesamiento del lenguaje natural en la aplicación	10
6.8.	Tratamiento del prompt en la aplicación.....	11
6.9.	Procesamiento de imágenes en la conversación	11
6.10.	Evaluación del modelo	12
6.10.1.	Generación de datos para evaluación	12
6.10.2.	Evaluación de recuperación	12
6.10.3.	Evaluación en lotes	12
7.	Presentación del producto.....	12
7.1.	¿Qué es?.....	12
7.2.	Características principales	12

7.3.	Cómo funciona	13
7.4.	Retos y limitaciones	14
7.5.	Beneficios clave.....	14
7.6.	Demo.....	15
8.	Conclusiones	15
9.	Aplicación en la empresa	16
10.	Próximos pasos	17
11.	Consideraciones financieras	17
11.1.	Presupuesto estimado	17
11.2.	Estimación de beneficios	18
11.3.	Análisis de coste-beneficio.....	19
11.3.1.	Costes del proyecto.....	19
11.3.2.	Beneficio neto	19
11.4.	Rentabilidad del proyecto (roi, van y payback)	19
11.4.1.	Conclusión.....	20
12.	Riesgos y mitigaciones	21
?	información inexacta	21
?	sincronización y actualización de información	21
?	falta de personalización	21
?	seguridad y privacidad de datos	21
?	interrupciones del servicio.....	22
?	cumplimiento normativo	22
?	baja adopción por parte de los usuarios.....	22
?	sesgo en las respuestas de ia	22
?	falta de comprensión del lenguaje natural	22
?	escalabilidad limitada.....	23
?	integración con sistemas de mapfre.....	23
?	falta de mantenimiento y monitoreo	23
13.	Q&A.....	23
14.	Bibliografía y referencias	25
14.2.	Repositorios y ejemplos de código en github	25
14.3.	Documentación técnica y artículos.....	25
14.4.	Fuentes audiovisuales	25

1. PUNTO DE PARTIDA

El presente proyecto tiene como objetivo desarrollar un chatbot avanzado para la empresa Mapfre, utilizando técnicas de Generación Aumentada por Recuperación (RAG). Este chatbot está diseñado para optimizar la experiencia del usuario al interactuar con los servicios de seguros, facilitando consultas personalizadas y la contratación de seguros de auto.

El chatbot tiene como finalidad principal **incrementar la eficiencia en el servicio al cliente**, mejorando la accesibilidad a la información y reduciendo el tiempo de respuesta a consultas frecuentes. Además, busca potenciar la conversión de usuarios interesados en productos de Mapfre en clientes activos, al proporcionar respuestas precisas, contextualmente relevantes y adaptadas a las necesidades individuales.

El chatbot se apoya en una arquitectura basada en embeddings, combinando las capacidades de LlamaIndex y Gemini para procesar y gestionar la información contenida en documentos (.pdf) relevantes sobre productos y servicios de Mapfre.

A continuación, se detalla el **flujo funcional y las principales funcionalidades del sistema**. Para gestionar los **embeddings** de manera eficiente, se procesan documentos (.pdf) que contienen datos relevantes sobre los productos y servicios de Mapfre, incluyendo imágenes y textos.

Utilizando la combinación de **Llama y Gemini**, se generan embeddings que capturan la **esencia semántica** de cada documento o fragmento de información. Estos embeddings se integran en **índices gestionados por LlamaIndex**, permitiendo una recuperación eficiente de la información basada en **similitud semántica**.

Para responder eficazmente a las consultas, el **chatbot** inicia la interacción recibiendo preguntas o solicitudes relacionadas con productos o procesos de seguros. Estas pueden abarcar desde **consultas generales** hasta requerimientos específicos, como la **cotización de seguros** o la **recopilación de documentos necesarios**.

Al recibir la consulta, el sistema **convierte la pregunta en un embedding**, lo compara con los índices previamente generados y recupera los textos o fragmentos de información que mejor responden a la consulta, basándose en la **similitud semántica**.

Una vez recuperada la información relevante, un **modelo de lenguaje (LLM)** procesa estos datos para integrar y sintetizar la información, generando una respuesta en **lenguaje natural**, que sea clara, precisa y útil para el usuario.

Para complementar sus capacidades, el **chatbot no solo responde consultas**, sino que también está diseñado para ofrecer un servicio integral, incluyendo:

1. **Asistencia en la cotización de seguros:** Guía al usuario a través del proceso de cotización, proporcionando información detallada sobre los productos y requisitos para contratar un seguro.
2. **Facilitación de la recopilación y verificación de documentos:** Ayuda al usuario a identificar y reunir los documentos necesarios para formalizar procesos relacionados con seguros, asegurando que se cumplan todos los requisitos.

3. **Recomendaciones personalizadas:** Basándose en el perfil del usuario y en sus consultas previas, el sistema puede sugerir **seguros de automóvil** que más se ajusten a sus necesidades, mejorando la experiencia y aumentando la satisfacción del cliente.

La evaluación del modelo se realizó en varias etapas. Se utilizó un conjunto de preguntas y respuestas creadas artificialmente para evaluar la capacidad de recuperación del sistema mediante métricas como MRR y HIT RATE. Para garantizar la calidad de las respuestas, se aplicaron evaluadores de fidelidad (faithfulness) y relevancia, comparando las respuestas generadas con los datos de referencia.

Con este sistema, **Mapfre podrá ofrecer una experiencia de usuario mejorada**, reducir costos operativos asociados con la atención al cliente y aumentar la tasa de conversión de potenciales clientes. Este enfoque, además, posiciona a la empresa como un referente en la innovación tecnológica aplicada al sector de seguros, alineando la solución con los objetivos estratégicos de Mapfre en la era digital.

2. ESTADO DE LA CUESTIÓN

La atención al cliente es un aspecto clave en el sector asegurador, pero la gestión de grandes volúmenes de consultas repetitivas y la necesidad de personalización suponen un desafío. Muchas empresas han implementado chatbots tradicionales basados en reglas, los cuales funcionan bien para preguntas simples, pero fallan en consultas más complejas y carecen de integración con bases de datos no estructuradas, como documentos PDF.

2.1.1. Soluciones actuales y sus limitaciones

Los chatbots basados en reglas son eficaces para consultas simples y procesos predecibles, pero presentan importantes limitaciones. Su **falta de flexibilidad** les impide responder preguntas fuera de su programación, **no pueden procesar información no estructurada** como documentos PDF o bases de datos internas, y **carecen de personalización**, lo que afecta la experiencia del usuario.

Si bien algunas empresas han comenzado a usar **modelos de lenguaje avanzados (LLMs)**, su eficacia se ve limitada cuando no están correctamente integrados con bases de datos especializadas, reduciendo su precisión en sectores como el asegurador.

2.1.2. Avances recientes en la tecnología

La evolución de los **modelos de lenguaje de gran tamaño (LLMs)**, como **Gemini, GPT-3 y GPT-4**, ha transformado la capacidad de los sistemas para **comprender, procesar y generar lenguaje natural** con mayor precisión y contexto. Cuando estos modelos se combinan con **técnicas de Recuperación Aumentada por Generación (RAG)**, logran:

- Manejar consultas complejas al integrar información contextual en las respuestas.
- Procesar grandes volúmenes de datos no estructurados, como documentos y bases de datos textuales.
- Ofrecer respuestas adaptadas a las necesidades específicas del usuario.

Además, herramientas como **LlamaIndex y LangChain** han facilitado la implementación de estas capacidades, permitiendo la creación de índices semánticos y la integración eficiente de LLMs en sistemas conversacionales avanzados.

2.1.3. Propuesta de este proyecto

En este contexto, este proyecto tiene como objetivo el desarrollo de un **chatbot inteligente para Mapfre**, basado en **RAG y modelos de lenguaje avanzados (LLMs)**. Esta solución busca:

- **Superar las limitaciones de los chatbots tradicionales**, proporcionando respuestas precisas y personalizadas, respaldadas por información actualizada y específica de la empresa.
- **Optimizar la contratación de seguros**, ofreciendo una experiencia de usuario más fluida, intuitiva y eficiente.
- **Aprovechar información no estructurada**, integrando documentos PDF y bases de datos internas para enriquecer las respuestas y mejorar la precisión del sistema.

2.1.4. Impacto esperado

La implementación de este chatbot mejorará significativamente la experiencia del cliente en Mapfre al ofrecer respuestas rápidas, precisas y personalizadas. Al automatizar la atención de consultas frecuentes, reducirá la carga operativa de los agentes humanos, permitiéndoles concentrarse en gestiones más complejas y estratégicas.

Además, su capacidad para **adaptarse a cada usuario y proporcionar información contextualizada** aumentará la tasa de conversión, impulsando la contratación de seguros de manera más eficiente. Con esta innovación, Mapfre fortalecerá su liderazgo en el sector asegurador, consolidándose como una empresa pionera en la transformación digital y la aplicación de inteligencia artificial en la atención al cliente.

3. EL PROBLEMA

Los chatbots tradicionales carecen de flexibilidad y precisión al manejar consultas complejas, lo que limita su utilidad en sectores como el asegurador. Su incapacidad para acceder a información no estructurada y ofrecer respuestas personalizadas afecta la experiencia del usuario y reduce la conversión de clientes. Para abordar estas deficiencias, se requiere una solución más avanzada que integre información contextual de manera dinámica y optimice la interacción con el cliente.

3.1.1. ¿Cómo afecta a Mapfre y por qué es un problema?

El uso de chatbots tradicionales en la atención al cliente de Mapfre afecta la experiencia del usuario y la eficiencia operativa. **Respuestas imprecisas y falta de personalización generan frustración**, reduciendo la confianza en la marca y favoreciendo la fuga de clientes hacia la competencia.

Además, **la falta de información clara y rápida impacta la tasa de conversión**, ya que los usuarios pueden abandonar el proceso de contratación. Operativamente, **la derivación de consultas a agentes humanos aumenta costos y tiempos de respuesta**, afectando la eficiencia.

El **manejo ineficiente de información no estructurada**, como documentos en PDF, limita el acceso a datos clave y genera respuestas incompletas. En un mercado competitivo, **no adoptar tecnologías avanzadas como RAG y LLMs** podría dejar a Mapfre en desventaja. Optimizar la interacción con los clientes es clave para fortalecer su competitividad en el sector asegurador.

3.1.2. ¿Por qué es importante resolverlo?

Un **chatbot avanzado** mejorará la **experiencia del cliente**, proporcionando respuestas más rápidas, precisas y personalizadas. Además, reducirá la **carga operativa** en los agentes, optimizando la eficiencia del servicio.

A nivel estratégico, fortalecerá la **competitividad de Mapfre**, aumentando la conversión de clientes y facilitando la contratación de seguros. También permitirá un mejor aprovechamiento de **información no estructurada**, asegurando respuestas más completas y relevantes.

4. LA SOLUCIÓN

En respuesta a los desafíos identificados, esta solución se basa en **Recuperación Aumentada por Generación (RAG)** y **modelos de lenguaje avanzados (LLMs)** para transformar grandes volúmenes de información en una base de conocimiento estructurada y accesible. Esto permite que el chatbot **recupere y procese información no estructurada**, como documentos PDF y bases de datos internas, ofreciendo respuestas **precisas, relevantes y personalizadas**.

A diferencia de los chatbots tradicionales y los LLMs sin integración con datos especializados, esta arquitectura optimiza la gestión de consultas, permitiendo **acceso en tiempo real a información específica de Mapfre**, lo que facilita procesos como la **cotización de seguros** y la **recopilación de documentos**.

A continuación, se compara esta solución con otros enfoques existentes en el mercado:

Característica	Chatbots Basados en Reglas	LLMs sin RAG	Nuestro Chatbot con RAG
Manejo de consultas complejas	?	Parcialmente	?
Recuperación de documentos PDF	?	Parcialmente	?
Integración con imágenes	?	No	?
Personalización de respuestas	?	Parcialmente	?

5. RECOLECCIÓN Y TRATAMIENTO DE DATOS

Tomando en cuenta la problemática y lo que queremos resolver, hemos diseñado un proceso de tratamiento de datos enfocado en garantizar que el sistema sea capaz de manejar grandes volúmenes de información no estructurada y proporcionar respuestas precisas y relevantes. Este enfoque asegura que

el chatbot cuente con una base de conocimiento sólida, accesible y bien organizada para responder a las consultas de los usuarios. Para ello, se definieron las siguientes etapas clave:

5.1.1. Identificación de las fuentes de datos

Se seleccionaron **documentos internos** relevantes de Mapfre, como términos y condiciones, políticas de seguros, preguntas frecuentes y guías de contratación.

Estos datos están principalmente en formatos no estructurados PDF lo que requiere procesos específicos para su transformación.

5.1.2. Carga de Documentos

Se emplea la clase “SimpleDirectoryReader” del módulo “llama_index.core” para leer todos los archivos de texto contenidos en un directorio definido. Esta función, implementada en “load_and_clean_documents”, recorre de manera recursiva el directorio y carga la información de cada documento. Esto garantiza que se capturen todos los archivos disponibles sin necesidad de especificar cada uno manualmente.

5.1.3. Limpieza y Segmentación de la Información

Una vez cargados, los documentos pasan por un proceso de segmentación utilizando el “SemanticSplitterNodeParser”. Este componente divide cada documento en fragmentos o **“nodos” semánticamente coherentes**. SemanticSplitterNodeParser utiliza un método de segmentación automática para dividir un documento en nodos, agrupando oraciones que están semánticamente relacionadas. A diferencia de los métodos tradicionales que dividen el texto en fragmentos de tamaño fijo. Además, el “NODE_PARSER” utiliza el modelo de embeddings “(GeminiEmbedding(model_name=“models/text-embedding-004”))” para evaluar y mantener la coherencia semántica de cada nodo.

5.1.4. Generación y Persistencia de Embeddings

Con los nodos obtenidos, se crea un índice vectorial mediante la clase “VectorStoreIndex”. Este índice utiliza los embeddings generados por el modelo “(GeminiEmbedding(model_name=“models/text-embedding-004”))” para transformar los nodos en vectores que representan su significado semántico. El resultado es una **estructura de datos optimizada para la búsqueda y recuperación de información basada en similitud semántica**.

5.1.5. Almacenamiento y Recuperación del Índice

Para optimizar el proceso, los nodos generados se almacenan en un archivo pickle “nodes.pkl”, permitiendo su reutilización sin necesidad de volver a procesar los documentos desde cero. Asimismo, el índice vectorial se persiste en un directorio específico “EMBEDDINGS_STORAGE” a través del método “persist” en formato JSON lo que posibilita la **carga del índice** en ejecuciones posteriores sin tener que reconstruirlo.

5.1.6. Verificación de Existencia del Índice

Antes de iniciar una nueva recolección, el sistema verifica la existencia de un índice previamente persistido (comprobando la presencia del archivo “docstore.json”). Si el índice existe, se carga directamente; de lo contrario, se procede a crear uno nuevo mediante la función “create_index”.

5.1.7. Integración con Modelos Externos

Tanto el modelo de embedding“(GeminiEmbedding(model_name="models/text-embedding-004"))” como el modelo de lenguaje **GEMINI** se configuran utilizando una **API Key** cargada desde un archivo externo “google_api_key.txt”. Esto garantiza que las operaciones de embedding y generación de lenguaje se realicen de forma segura y acorde con las políticas del proveedor del servicio.

El **tratamiento y la recolección de datos** constituyen un pilar fundamental de este proyecto, ya que la calidad de los datos procesados tiene un impacto directo en la capacidad del chatbot para responder de manera **precisa y contextual**. Este enfoque asegura que Mapfre pueda ofrecer una experiencia de usuario de **alta calidad** y mantener su **posición competitiva** en el mercado asegurador.

6. DESARROLLO DEL PROYECTO

6.1.1. Procesamiento de imágenes con “GeminiMultiModal”

Tras completar el preprocesamiento de la información descrito en el apartado 5, se implementó una función multimodal para captar información necesaria para la contratación del seguro directamente desde la captura de una imagen del vehículo a asegurar.

6.1.2. Configuración y carga de la API Key

Para procesar imágenes y extraer información estructurada, se establece una configuración inicial en la que se define la ubicación base del proyecto y la ruta del archivo que contiene la API Key de Google.

- **Ruta base:** Se obtiene de manera dinámica usando “os.path.abspath”.
- **Carga de la API Key:** Se utiliza la función “load_api_key” para extraer la clave desde un archivo “google_api_key.txt”. Si la clave no se puede cargar, se lanza una excepción para evitar que el sistema continúe sin acceso a la API.

6.1.3. Modelo de datos para información vehicular

Se define una estructura de datos utilizando “pydantic”, lo que permite validar y estructurar la información extraída de las imágenes de vehículos. La clase “VehicleInfo” especifica los siguientes campos:

- “matricula”: Número de matrícula del vehículo.
- “ano_matricula”: Año de matriculación del vehículo.
- “modelo”: Marca y modelo del vehículo.
- “color”: Color del vehículo.

6.1.4. Generación de embeddings con GeminiMultiModal

Se implementa la función “pydantic_gemini”, que emplea el modelo “GeminiMultiModal” para analizar imágenes y extraer información vehicular en formato JSON.

- Definición del modelo: Se instancia “GeminiMultiModal” con el modelo “gemini-2.0-flash-exp”.
- Procesamiento de imágenes Se usa “MultiModalLLMCompletionProgram” para estructurar la salida en formato “PydanticOutputParser”, basado en la clase “VehicleInfo”.
- Respuesta: Se obtiene un “response_text” en formato JSON con la información extraída de la imagen.

6.1.5. Integración en la aplicación

Tras desarrollar el modelo de procesamiento de imágenes y extracción de información vehicular, se integró esta funcionalidad dentro de la aplicación para garantizar un flujo continuo de datos en la asistencia a la contratación de seguros. La integración permite que el asistente virtual combine datos textuales y visuales en tiempo real para generar respuestas más precisas y contextualizadas.

6.1.6. Configuración de la aplicación

Para la integración en la aplicación, se define la estructura principal con las siguientes configuraciones:

- **Carga de la API Key:** Se utiliza “load_api_key” para obtener la clave de autenticación de Google.
- **Carga del Índice de Embeddings:** Se usa “check_and_load_index” para recuperar o crear un índice de recuperación de documentos, lo que permite realizar consultas eficientes en la base de conocimientos almacenada.
- **Configuración de la IA:** Se establece “Gemini” como modelo de lenguaje y “(GeminiEmbedding(model_name=“models/text-embedding-004”))” como modelo de embeddings para la generación de respuestas basadas en la consulta del usuario.
- **Memoria del Chat:** Se implementa “ChatMemoryBuffer” con un límite de 20,000 tokens para mantener el contexto de la conversación y permitir interacciones más fluidas y personalizadas.

6.1.7. Procesamiento del lenguaje natural en la aplicación

La aplicación utiliza **Chainlit** como framework para manejar la interacción con los usuarios en tiempo real. La lógica principal de recuperación y generación de respuestas se estructura de la siguiente manera:

- **Inicialización del Modelo de Lenguaje:**
 - Se configura “Gemini” como el modelo principal para procesar las consultas de los usuarios.
 - Se establece “GeminiEmbedding” para calcular la similitud semántica entre la consulta del usuario y los documentos almacenados en el índice.
- Manejo de **Sesiones** con Chainlit:
 - Se utiliza “cl.on_chat_start” para inicializar un “chat_engine” al comienzo de la conversación, configurando la recuperación de documentos y la memoria de contexto.

- Las interacciones del usuario son gestionadas en “cl.on_message”, donde las consultas de texto son procesadas directamente por el “chat_engine”, y las imágenes son analizadas con “pydantic_gemini”.
- Proceso de **Recuperación de Información**:
 - Cada consulta del usuario es transformada en un vector de embeddings utilizando “GeminiEmbedding”.
 - Se busca en el índice de embeddings mediante “RetrieverQueryEngine”, recuperando los documentos más relevantes.
 - La respuesta final es generada combinando la información recuperada con el contexto del chat.

6.1.8. Tratamiento del Prompt en la Aplicación

El sistema utiliza un “system_prompt” predefinido que establece las **directrices** de comportamiento del asistente virtual. Este prompt incluye:

- **Identidad del asistente**: Se presenta como un profesional cualificado en seguros de coche de Mapfre, denominado "Mappi".
- **Restricciones de respuesta**: Solo responde sobre seguros de coche de Mapfre y no sobre otras empresas.
- **Formato de conversación**: Mantiene respuestas cortas, educadas y en español, utilizando emojis para mejorar la experiencia de usuario.
- **Flujo de conversación**:
 - En caso de preguntas sobre contratación, sigue un flujo guiado para recopilar los datos necesarios y calcular el precio del seguro.
 - Si el usuario ya es cliente, proporciona asistencia para trámites, reclamaciones o modificación de pólizas.
 - Procesa imágenes enviadas por el usuario para obtener información adicional sobre el vehículo antes de calcular el precio del seguro.

6.1.9. Procesamiento de imágenes en la conversación

Dentro de la aplicación, se implementa una función para manejar el procesamiento de imágenes enviadas por los usuarios:

- **Recepción de imágenes**: Se identifican los archivos de tipo imagen dentro del mensaje del usuario.
- **Almacenamiento temporal**: Se guardan en el directorio configurado (“IMAGE_PATH”).
- **Carga de imágenes**: Se utiliza “SimpleDirectoryReader” para extraer datos de las imágenes.
- **Procesamiento** con “pydantic_gemini”: Se extrae información vehicular y se devuelve en formato JSON.
- **Respuesta al usuario**: Se envía la información procesada al usuario a través del chat.

Esta integración garantiza que la aplicación pueda recuperar información de manera eficiente, responder preguntas de los usuarios utilizando el modelo de lenguaje Gemini y procesar imágenes de manera contextual, proporcionando respuestas personalizadas y mejorando la experiencia de usuario en la contratación y consulta de seguros.

6.1.10. Evaluación del Modelo

Una vez completada la integración de la funcionalidad de **procesamiento de imágenes y recuperación de información basada en embeddings**, se llevó a cabo una **evaluación del rendimiento del modelo**. Esta evaluación tiene como objetivo validar la **precisión de la recuperación de información**, la **fidelidad de las respuestas generadas** y la **relevancia**. Se implementaron **métricas clave** para garantizar que la solución funcione de manera óptima en un entorno de producción.

6.1.11. Generación de Datos para Evaluación

Para evaluar el rendimiento del modelo, se generan pares de **preguntas y contextos** basados en los datos procesados. Si el archivo de preguntas y respuestas ("**q&a.pkl**") no existe, se crea y guarda en formato **pickle** para su posterior uso.

6.1.12. Evaluación de Recuperación

Se utiliza **RetrieverEvaluator** para evaluar métricas como **MRR** y **Hit Rate**, asegurando la **precisión en la recuperación de información**. Para manejar grandes volúmenes de consultas, se divide el conjunto de datos en **fragmentos** antes de la evaluación.

6.1.13. Evaluación en Lotes

Se usa **BatchEvalRunner** para evaluar **múltiples consultas en paralelo**. Con ello se evalúa la Fidelidad y Relevancia: Se implementan evaluaciones de **fidelidad** ("**FaithfulnessEvaluator**") y **relevancia** ("**RelevancyEvaluator**") utilizando múltiples versiones del modelo "Gemini". Se comparan los resultados de los modelos "pro1_5", "flash1_5" y "**flash2**" para determinar cuál ofrece las respuestas más precisas y relevantes.

7. PRESENTACIÓN DEL PRODUCTO

7.1.1. ¿Qué es?

Hemos desarrollado un **chatbot avanzado basado en RAG** para optimizar la atención al cliente en **seguros de auto**, permitiendo obtener información precisa, cotizar pólizas y gestionar documentos de forma eficiente. Para una interacción fluida, el chatbot está implementado en **Chainlit**, una plataforma especializada en aplicaciones conversacionales con IA.

Esta solución permite a los usuarios obtener información precisa, realizar cotizaciones y gestionar documentación de manera rápida y eficiente. Está diseñado para operar de manera autónoma, reduciendo la carga operativa de los agentes humanos y mejorando la conversión de clientes potenciales.

7.1.2. Características Principales

Nuestra solución implementa tecnologías avanzadas que optimizan la interacción con los usuarios y mejoran la eficiencia operativa de Mapfre.

- **Interacción multimodal:** El chatbot no solo procesa texto, sino que también puede analizar imágenes para facilitar la recopilación de datos:
 - Texto: Disponible en plataformas web y móviles, permitiendo interacción en lenguaje natural.
 - Imágenes: Capaz de analizar fotos de vehículos para extraer datos como matrícula, modelo y color, agilizando la cotización del seguro.
- **Recuperación Aumentada por Generación (RAG):**
 - Combina modelos de lenguaje (LLMs) con recuperación de información estructurada.
 - Permite consultar bases de datos internas y documentos en PDF para ofrecer respuestas precisas y basadas en datos oficiales de Mapfre.
- **Automatización de procesos:**
 - Cotización automatizada: Basada en los datos del usuario y la información del vehículo.
 - Recopilación de documentos: Para facilitar la contratación o gestión de siniestros.
 - Generación y envío de contratos provisionales: Reduciendo tiempos de espera para los clientes.
- **Personalización avanzada:** Respuestas adaptadas según el historial del usuario y los productos contratados previamente.
- **Integración con sistemas digitales:** Conexión con el CRM, ERP y otras aplicaciones internas de Mapfre, asegurando una experiencia fluida para los clientes y los agentes de seguros.

7.1.3. Cómo Funciona

El chatbot opera en un flujo estructurado para optimizar la interacción:

1. **Consulta del Usuario:** El usuario puede interactuar por texto o enviar una imagen del vehículo.
2. **Recuperación de Información:** Se generan embeddings a partir de la consulta y se buscan coincidencias en la base de datos de Mapfre.
3. **Generación de Respuesta:** El modelo de lenguaje procesa la información recuperada y construye una respuesta en lenguaje natural (LLM).
4. **Entrega de Respuesta:** Se muestra la respuesta en el formato más adecuado según la consulta del usuario.

7.1.4. Retos y Limitaciones

- **Retos Técnicos:**
 - Garantizar la actualización constante de los datos en la base de conocimiento.
 - **Optimizar la latencia** en la recuperación y generación de respuestas para consultas complejas.
- **Retos de Negocio:**
 - **Integración** con los sistemas internos sin generar fricciones.
 - Garantizar la **adopción** del chatbot por parte de los **usuarios**, asegurando una experiencia intuitiva y confiable.
- **Limitaciones Actuales:**
 - **Escalabilidad** a otros seguros: Actualmente, el chatbot solo está enfocado en seguros de autos.
 - **Consultas fuera del dominio de Mapfre:** Preguntas sobre otras aseguradoras deben ser derivadas a agentes humanos.
 - **Dependencia de los datos** iniciales: La calidad del chatbot depende directamente de la información cargada en la base de conocimientos.

7.1.5. Beneficios Clave

La implementación del **chatbot en Mapfre** representa una **mejora sustancial** tanto para los usuarios como para la compañía.

Para los clientes, el chatbot garantiza una **disponibilidad 24/7**, permitiendo **acceso inmediato** a información y asistencia sin depender de horarios de atención. Además, proporciona **respuestas rápidas y precisas**, reduciendo tiempos de espera y facilitando la navegación en procesos complejos como la **cotización y contratación de seguros**. Su capacidad de **procesar imágenes** agiliza aún más la gestión, eliminando errores manuales y simplificando la experiencia del usuario.

Desde la perspectiva de Mapfre, el chatbot contribuye a una **reducción de costos operativos**, aliviando la carga de trabajo de los agentes humanos y optimizando el uso de recursos. También impulsa el **aumento en la tasa de conversión**, al guiar de manera efectiva a los usuarios a través del proceso de contratación. Además, mejora la **gestión de datos no estructurados**, permitiendo un acceso más eficiente a la información contenida en documentos internos.

En conjunto, esta solución refuerza la innovación de Mapfre y su liderazgo en la **digitalización del sector asegurador**.

7.1.6. Demo

🗣️ ¡Hola! 🤖 Soy MAPPI, tu asistente virtual de Mapfre. ¿En qué puedo ayudarte hoy con tu seguro de auto? 🚗

📄

Me gustaría tener un seguro a todo riesgo para mi coche


🗣️ ¡Perfecto! Un seguro a todo riesgo te ofrece la mayor protección para tu vehículo, cubriendo daños propios y a terceros, además de otras coberturas adicionales como robo, incendio y asistencia en viaje. 🌟

Para poder ofrecerte la mejor opción, necesito algunos datos. Si quieres, puedes subir una foto de tu coche para agilizar el proceso. 📷

Empecemos con el tipo de seguro que deseas, que en este caso es a todo riesgo. ¿Es correcto? ✅

📄

Este es mi coche



🗣️ ¡Entendido! Perfecto, ahora necesito algunos datos adicionales para poder calcular el precio de tu seguro a todo

chat.input.placeholder

📷

Chainlit

8. CONCLUSIONES

El desarrollo de este proyecto como parte de nuestro Trabajo de Fin de Máster nos ha permitido aplicar en un entorno real los conocimientos adquiridos durante nuestra formación. La creación de un chatbot con RAG para Mapfre nos ha brindado la oportunidad de profundizar en tecnologías avanzadas de procesamiento del lenguaje natural, recuperación de información y automatización de procesos en el sector asegurador.

A lo largo de este proyecto, nos hemos enfrentado a desafíos tanto técnicos como de negocio. En el ámbito técnico, trabajar con modelos de lenguaje y sistemas de recuperación de información nos ha llevado a optimizar la eficiencia y precisión de las respuestas del chatbot. Además, la integración con bases de datos y la gestión de grandes volúmenes de información no estructurada han sido aspectos clave en nuestro aprendizaje. Desde el punto de vista del negocio, entendimos la importancia de la experiencia del usuario y cómo la automatización puede mejorar la eficiencia operativa y la conversión de clientes.

Uno de los aprendizajes más valiosos ha sido la importancia de la actualización constante de la base de conocimiento y la necesidad de garantizar que el chatbot responda de manera precisa y relevante. También hemos identificado que, aunque la tecnología puede automatizar muchas tareas, siempre es necesario un punto de contacto humano en casos específicos.

Este trabajo nos ha permitido no solo mejorar nuestras habilidades técnicas en el desarrollo de soluciones con inteligencia artificial, sino también adquirir una visión más amplia sobre su impacto en el mundo empresarial. Consideramos que este proyecto puede sentar las bases para futuras mejoras y una posible implementación en otras áreas dentro de Mapfre.

En conclusión, este TFM ha sido una experiencia enriquecedora que nos ha permitido aplicar nuestro conocimiento en un caso práctico con impacto real. Hemos consolidado nuestras habilidades en inteligencia artificial, procesamiento de lenguaje natural y automatización, además de desarrollar competencias clave en trabajo en equipo, toma de decisiones y resolución de problemas en entornos complejos. Estamos convencidos de que los aprendizajes adquiridos durante este proceso nos serán de gran utilidad en nuestra futura carrera profesional.

9. APLICACIÓN EN LA EMPRESA

La implementación de un **chatbot multimodal avanzado** supone un **hito estratégico** en la transformación digital del sector asegurador. Su integración optimiza la **interacción con los clientes**, automatiza procesos internos y refuerza el liderazgo de **Mapfre** en innovación tecnológica.

Beneficios para Mapfre

- **Mejora en la atención al cliente:** El chatbot opera **24/7**, ofreciendo respuestas inmediatas y precisas. Su capacidad para reducir tiempos de espera minimiza la frustración del usuario, mejorando su experiencia y fortaleciendo la fidelización.
- **Eficiencia operativa y reducción de costes:** La automatización de consultas frecuentes **libera a los agentes humanos** para tareas más complejas, reduciendo costos operativos y aumentando la productividad. Además, el chatbot maneja interacciones en **varios idiomas**, mejorando la capacidad de respuesta.
- **Personalización y precisión:** Gracias a la **Recuperación Aumentada por Generación (RAG)**, el chatbot accede a bases de datos internas, proporcionando respuestas **específicas y actualizadas** según el perfil del usuario, lo que fortalece la confianza en la aseguradora.
- **Captación de nuevos clientes y cross-selling:** El chatbot facilita la contratación de seguros y detecta **oportunidades de venta cruzada**, sugiriendo productos complementarios y mejorando la conversión.
- **Análisis de datos y mejora continua:** Cada interacción genera **datos valiosos** sobre las necesidades y comportamientos de los clientes, permitiendo la optimización del servicio y la adaptación de estrategias en función de la demanda del mercado.
- **Cumplimiento de normativas y seguridad:** Garantiza la **protección de datos** y la confidencialidad de la información del cliente, reforzando la confianza y asegurando el cumplimiento de la normativa vigente.

- **Diferenciación competitiva:** La adopción de IA coloca a **Mapfre** en la vanguardia del sector, alineándose con las tendencias tecnológicas y mejorando la eficiencia operativa.
- **Escalabilidad y evolución:** El chatbot está diseñado para **adaptarse a nuevas tendencias**, incorporando mejoras continuas basadas en análisis de datos y feedback de los clientes.

En conjunto, esta solución **mejora la eficiencia, reduce costos y refuerza la posición de Mapfre como líder en digitalización** dentro del sector asegurador.

10. PRÓXIMOS PASOS

Se considera que hay varias líneas a cubrir que surgen como necesidad o de manera complementaria a este trabajo, se pasa a enumerar ciertas de ellas:

- **Personalización:** Manejar la posibilidad de enlace con una cuenta personal para ofrecer información aún más personalizada a cada usuario y de esta manera, paralelamente, conseguir ampliar la base de datos de clientes transformados y no transformados.
- **Cierre de ventas:** Introducir las fórmulas exactas para calcular el precio de los productos y generar un envío automático de un contrato provisional.
- **Seguimiento en mediante un análisis de la recopilación de los datos y perfeccionamiento:** Se recopilan datos sobre el uso y las cuestiones más frecuentes tratadas con el chatbot para potenciar su rendimiento, seguimiento de KPIs del propio chatbot (observando, por ejemplo, en qué casos hay descontento en los usuarios) y agregar nuevas funcionalidades.
- **Ampliación a otros productos:** Dentro de Mapfre hay seguros de coche, salud, hogar, moto, viajes, mascotas, vida, decesos... Con un buen planteamiento de los datos preprocesados del RAG se podría intentar cubrir estos productos.
- **Integración con otros sistemas:** Se integra el chatbot con otros sistemas de Mapfre, como el CRM y en las aplicaciones móviles.

11. CONSIDERACIONES FINANCIERAS

11.1.1. Presupuesto estimado

El presupuesto estimado para la elaboración del chatbot incluye todos los costes asociados con el desarrollo, implementación y mantenimiento del mismo. A continuación se encuentra el desglose estimado de los mismos:

- **Costes de desarrollo:**
 - Infraestructura: (GCP Cloud Run)
 - API y modelos de IA (Gemini 2.0)
 - Desarrollo y pruebas
 - Personal técnico externo (Ingeniero IA + Backend/DevOps)

Categoría	Coste aproximado
Infraestructura	3.630€
API y modelos de IA (mensual)	4.098€
Desarrollo y pruebas	10.000€ - 30.000€
Personal externo	15.000€ - 40.000€
TOTAL	32.728 - 77.728 €/año

- **Costes de implementación**

- Integración con sistemas
- Optimización y escalabilidad
- Seguridad y cumplimiento

Categoría	Coste aproximado
Integración con sistemas	15.000€ - 40.000€
Optimización y escalabilidad	10.000€ - 25.000€
Seguridad y cumplimiento	10.000€ - 30.000€
TOTAL	35.000 - 95.000 €/año

- **Costes de mantenimiento**

- Operación Continua
- Actualizaciones y mejoras
- Soporte Técnico

Categoría	Coste aproximado
Operación Continua	20.000€ - 50.000€
Actualizaciones y Mejoras	15.000€ - 40.000€
Soporte Técnico	10.000€ - 30.000€
TOTAL	45.000 - 120.000 €/año

11.1.2. Estimación de beneficios

Para la estimación de los beneficios esperados con la implementación del chatbot, contamos con la siguiente información:

- **Situación actual Mapfre (sin Chatbot):**
 - Leads anuales: 1.200.000
 - Tasa de conversión actual: 0,5%
 - Pólizas anuales: $1.200.000 \times 0,5\% = 6.000$ pólizas/año
 - Precio medio por póliza = 500€
 - Ingresos anuales estimados = $6.000 \text{ pólizas} \times 500\text{€} = 3.000.000\text{€}$
- **Escenario esperado con Chatbot:**
 - Aumento de 200.000 leads anuales = 1.400.000 leads anuales
 - Aumento del 0,2% en tasa de conversión = Tasa de conversión del 0,7%
 - Pólizas anuales esperadas: $1.400.000 \times 0,7\% = 9.800$ pólizas/año
 - Ingresos anuales esperados: $9.800 \text{ pólizas} \times 500\text{€} = 4.900.000\text{€}$
- **Incremento de Ingresos Generados tras la implementación del Chatbot:**
 - Aumento de pólizas contratadas: $9.800 - 6.000 = 3.800$ pólizas adicionales
 - Incremento de ingresos: $4.900.000\text{€} - 3.000.000\text{€} = 1.900.000\text{€}$ adicionales

Según las estimaciones realizadas, el chatbot permitirá aumentar las pólizas en un 63,33%, pasando de 6.000 a 9.800 pólizas contratadas al año.

Esto reduce el coste de adquisición por cliente, ya que más leads se convierten en pólizas sin aumentar en la misma proporción la inversión en marketing.

11.1.3. Análisis de coste-beneficio

Tras lo desglosado en los puntos anteriores, conocemos los costes y beneficios estimados del proyecto:

11.1.4. Costes del proyecto

- Coste de Desarrollo: 32.728 - 77.728 €/año
- Costes de Mantenimiento 30.000 - 80.000 €/año
- Costes de Implementación: 35.000 - 95.000 €/año
- Costes Total del Proyecto:
 - Escenario de Coste Mínimo: $32.728\text{€} + 30.000\text{€} + 35.000\text{€} = 97.728 \text{ €/año}$
 - Escenario de Coste Máximo: $77.728\text{€} + 80.000\text{€} + 95.000\text{€} = 252.728 \text{ €/año}$

11.1.5. Beneficio Neto

- Escenario de Coste Máximo (más conservador):
 $1.900.000 \text{ €} - 252.728 \text{ €} = 1.647.272 \text{ €/año}$
- Escenario de Coste Mínimo (más optimista):
 $1.900.000 \text{ €} - 97.728 \text{ €} = 1.802.272 \text{ €/año}$

11.1.6. Rentabilidad del proyecto (ROI, VAN y Payback)

- Ingresos adicionales esperados: 1.900.000€/año
- Coste total estimado: 97.728 - 252.728 €/año

- Beneficio neto esperado: 1.647.272 - 1.802.272 €/año
- Retorno de la inversión (ROI):
 - Escenario Conservador (Beneficio mínimo / Coste máximo):
 $(1.647.272 / 252.728) \times 100 = 651,80\%$
 - Escenario Intermedio 1 (Beneficio máximo / Coste máximo):
 $(1.802.272 / 252.728) \times 100 = 713,130\%$
 - Escenario Intermedio 2 (Beneficio mínimo / Coste mínimo):
 $(1.627.272 / 97.728) \times 100 = 1665,10\%$
 - Escenario Optimista (Beneficio máximo / Coste mínimo)
 $(1.802.272 / 97.728) \times 100 = 1844,17\%$
- Valor Actual Neto(VAN):
 - VAN - Escenario Mínimo (5 años, 5% de descuento): 7.034.097 €
 - VAN - Escenario Máximo (5 años, 5% de descuento): 7.550.166 €
 - VAN - Escenario Mínimo (5 años, 10% de descuento): 6.146.728 €
 - VAN - Escenario Máximo (5 años, 10% de descuento): 6.579.300 €
- Payback period (Tiempo de recuperación):
 - Escenario Mínimo:
 $(97.728 / 1.647.272) = 0,06$ años (22 días aprox)
 - Escenario Máximo:
 $(252.728 / 1.802.272) = 0,14$ años (2 meses aprox)

Sujeto a pruebas, el proyecto asegura un ROI positivo, con un tiempo de recuperación estimado entre 1 y 2 meses, consolidando una solución estratégica para Mapfre.

11.1.7. Conclusión:

- En todos los escenarios calculados, el ROI está por encima del 100%, lo que indica que la inversión en el Chatbot es rentable y los beneficios obtenidos significativos.
- Todos los escenarios tienen un VAN positivo, lo que indica que el proyecto genera más valor del que cuesta.
- En cualquiera de los casos, la inversión se recupera en menos de 2 meses.
- La empresa decide autofinanciarse para el desarrollo del chatbot debido a que los números calculados muestran un retorno de la inversión muy positivo y en menos de un año, por lo que no es necesario salir en busca de financiación.

12. RIESGOS Y MITIGACIONES

El desarrollo e implementación del chatbot en Mapfre conlleva una serie de riesgos que pueden afectar a su rendimiento, precisión y aceptación por parte de los usuarios. A continuación, se detallan los principales riesgos identificados y las estrategias para mitigarlos:

- **Información inexacta**

Riesgo: La integración de las distintas fuentes de datos, incluyendo documentos oficiales de los seguros de Mapfre, con modelos de IA Generativa, puede generar respuestas incorrectas o inconsistentes, afectando la confianza del usuario.

Mitigación: Para ello, se propone implementar un proceso de validación de respuestas utilizando herramientas de control de calidad y pruebas automatizadas. Además, se podría incorporar un mecanismo de feedback que permita a los usuarios reportar respuestas erróneas para su corrección.

- **Sincronización y actualización de información**

Riesgo: Aunque el chatbot utiliza una arquitectura RAG que permite acceder a información en tiempo real, pueden existir problemas de sincronización si los documentos fuente no están correctamente indexados o si hay cambios en la estructura de los datos. Esto podría ocasionar respuestas basadas en versiones desactualizadas de las pólizas.

Mitigación: Establecer un sistema de indexación automatizada que garantice que las actualizaciones en las pólizas y documentos sean reflejadas de inmediato en el chatbot. Implementar un mecanismo de control de versiones y alertas que notifique sobre cambios en la estructura de datos para ajustar el proceso de recuperación en tiempo real.

- **Falta de personalización**

Riesgo: Si el chatbot no puede adaptar lo suficiente sus respuestas a cada cliente, la experiencia de usuario puede verse afectada, reduciendo la conversión de leads.

Mitigación: Implementar algoritmos de aprendizaje automático que permitan al chatbot responder preguntas específicas y ofrecer cotizaciones personalizadas utilizando la información del cliente. Estos algoritmos pueden aprender de las interacciones previas con los usuarios y ofrecer una experiencia más personalizada y atractiva.

- **Seguridad y privacidad de datos**

Riesgo: El chatbot maneja información personal sensible de los clientes, lo que expone a Mapfre a riesgos de brecha de seguridad como filtraciones y ataques cibernéticos.

Mitigación: Aplicar encriptación de extremo a extremo para la transmisión y almacenamiento de datos. Implementar controles de acceso estrictos y auditorías de seguridad regulares para prevenir brechas de datos.

- **Interrupciones del servicio**

Riesgo: Fallos en la infraestructura podrían dejar al chatbot inoperativo, afectando la experiencia del usuario y la imagen de Mapfre.

Mitigación: Utilizar una infraestructura en la nube con balanceo de carga y redundancia geográfica. Implementar planes de recuperación ante desastres y pruebas de carga para garantizar la estabilidad del sistema.

- **Cumplimiento normativo**

Riesgos: El chatbot debe cumplir con el RGPD y otras regulaciones sobre protección de datos y privacidad.

Mitigación: Implementar un equipo de cumplimiento regulatorio que monitorice cambios en las leyes y normativas, asegurando que el chatbot se ajuste rápidamente a estos cambios. Realizar auditorías periódicas de cumplimiento para verificar que el chatbot cumple con las regulaciones vigentes.

- **Baja adopción por parte de los usuarios**

Riesgos: Algunos clientes pueden preferir la interacción humana en lugar de utilizar un chatbot.

Mitigación: Implementar una estrategia de comunicación y formación para los usuarios, destacando la facilidad y rapidez del chatbot. Además, recoger feedback y realizar mejoras continuas basadas en las sugerencias y preocupaciones del usuario. Sería conveniente habilitar la opción de transferir la conversación a un agente humano en caso de ser necesario.

- **Sesgo en las respuestas de IA**

Descripción: Los modelos de IA pueden generar respuestas sesgadas si el entrenamiento de datos no es representativo.

Mitigación: Aplicar técnicas de eliminación de sesgos en el entrenamiento del modelo y realizar revisiones periódicas para detectar y corregir desviaciones en las respuestas.

- **Falta de comprensión del lenguaje natural**

Riesgos: El chatbot podría tener dificultades para comprender el lenguaje natural del usuario, especialmente si utiliza jerga, frases complejas o errores gramaticales, lo que dificulta una comunicación fluida y eficiente.

Mitigación: Utilizar modelos avanzados de NLP con capacidad de detección de errores y aprendizaje continuo para mejorar la comprensión del usuario.

- **Escalabilidad limitada**

Riesgo: Un aumento en la demanda podría saturar la infraestructura, reduciendo la capacidad de respuesta del chatbot.

Mitigación: Diseñar una arquitectura escalable basada en microservicios que permita la asignación dinámica de recursos en función del tráfico de usuarios.

- **Integración con sistemas de Mapfre**

Riesgos: Si el chatbot no está bien integrado con los sistemas internos de Mapfre, su funcionalidad puede verse limitada.

Mitigación: Utilizar APIs estandarizadas para conectar el chatbot con los sistemas de gestión de seguros y garantizar un acceso fluido a la información.

- **Falta de mantenimiento y monitoreo**

Riesgos: La ausencia de actualizaciones y monitoreo del chatbot puede llevar a una degradación en su rendimiento y precisión.

Mitigación: Establecer un plan de mantenimiento con revisiones periódicas, implementación de mejoras basadas en el feedback de los usuarios y un sistema de monitoreo en tiempo real para identificar problemas rápidamente.

13. Q&A

Esta sección busca responder las preguntas clave sobre nuestro proyecto, proporcionando una visión clara de los desafíos, soluciones y aprendizajes obtenidos durante el desarrollo del chatbot.

¿Cuáles fueron los principales desafíos al integrar la API de Gemini con el chatbot?

La principal dificultad fue optimizar la latencia de respuesta del chatbot al utilizar modelos de generación y embeddings de Gemini. También fue necesario garantizar que la API manejara eficientemente múltiples consultas simultáneas sin afectar el rendimiento.

¿Cómo afecta el ChatMemoryBuffer de 20,000 tokens al rendimiento del chatbot?

Este límite permite que el chatbot mantenga el contexto de conversaciones prolongadas sin perder información relevante. Sin embargo, requiere una optimización constante para equilibrar la capacidad de almacenamiento de contexto con la velocidad de respuesta.

¿Cómo decidieron qué restricciones incluir en el system_prompt del chatbot?

Se diseñaron reglas claras para evitar respuestas fuera del ámbito de los seguros de Mapfre, garantizar la cortesía en el trato y mantener el lenguaje claro y conciso. También se incluyó un protocolo de derivación a agentes humanos para consultas fuera del alcance del chatbot. La metodología fue ensayo-error.

¿Cómo validaron que la función `calcular_seguro()` proporciona estimaciones realistas?

No tenemos acceso a datos reales por lo que se hizo una simulación que para su implementación real debería ser modificada. **Sugerimos una simulación basada en datos del sector**, con variables como vehículo, zona y perfil del conductor. Para producción, debe integrarse con la base de datos de Mapfre.

¿Qué dificultades encontraron al implementar el procesamiento de imágenes con `pydantic_gemini`?

Uno de los principales retos fue garantizar que las imágenes analizadas realmente correspondieran a vehículos y accidentes, filtrar imágenes irrelevantes. También fue clave optimizar la calidad de las imágenes para obtener resultados precisos sin afectar el tiempo de procesamiento.

¿Cómo midieron el impacto del chatbot en la tasa de conversión de clientes?

Se establecieron métricas clave como la cantidad de cotizaciones generadas, la tasa de conversión de leads y la reducción del tiempo medio de respuesta. Estas métricas fueron comparadas con datos previos a la implementación del chatbot.

¿Cuánto podría ahorrar Mapfre con la implementación del chatbot?

Aunque aún no hay cifras definitivas, se estima que la automatización de consultas y cotizaciones podría reducir significativamente la carga operativa, permitiendo ahorrar en costos de atención al cliente y optimizar recursos humanos.

¿Cómo se asegura la personalización de las respuestas del chatbot?

El chatbot utiliza datos de clientes, historial de interacciones y modelos de recomendación para adaptar sus respuestas a cada usuario. Además, permite la validación de información con preguntas de seguimiento para mejorar la personalización.

¿Cómo se previene el fraude en el uso de imágenes para evaluación de siniestros?

Se implementan algoritmos de detección de inconsistencias en imágenes y se comparan con bases de datos de incidentes previos. También se pueden solicitar imágenes adicionales o verificaciones manuales en casos sospechosos.

¿Cómo maneja el chatbot consultas ambiguas o preguntas fuera de su alcance?

Buscamos detectar intenciones con NLP, respuestas de aclaración para poder derivar a agentes humanos cuando es necesario.

¿Qué estrategias se implementan para seguir optimizando el chatbot después de su lanzamiento?

El chatbot irá mejorando a través de aprendizaje continuo con datos reales, mejora de prompts y refinamiento de modelos basados en interacciones previas.

14. BIBLIOGRAFÍA Y REFERENCIAS

14.1. Documentación Oficial de Librerías y Frameworks

- Chainlit. (2024). Chainlit: Open-source framework for LLM apps. Disponible en <https://docs.chainlit.io>
- Gemini API. (2024). Google AI: Gemini models for text and multimodal tasks. Disponible en <https://ai.google.dev>
- LlamaIndex. (2024). LlamaIndex: A data framework for LLMs. Disponible en <https://gpt-index.readthedocs.io/en/latest/>
- Pydantic. (2024). Pydantic: Data validation and settings management using Python. Disponible en <https://docs.pydantic.dev>
- Sentence-Transformers. (2024). Sentence-Transformers: Pretrained models for text embeddings. Disponible en <https://www.sbert.net>
- Python Software Foundation. (2024). Pickle Library (Python): Object serialization in Python. Disponible en <https://docs.python.org/3/library/pickle.html>

14.1.1. Repositorios y Ejemplos de Código en GitHub

- Google. (2024). Gemini LlamaIndex QA Chroma WebPageReader [Repositorio de GitHub]. Disponible en https://github.com/google/generative-ai-docs/blob/main/examples/gemini/python/llamaindex/Gemini_LlamaIndex_QA_Chroma_WebPageReader.ipynb

14.1.2. Documentación Técnica y Artículos

- Google Cloud. (2024). Quickstart for Retrieval-Augmented Generation (RAG) with Ray, LangChain, and Hugging Face. Disponible en <https://cloud.google.com/blog/products/ai-machine-learning/rag-quickstart-with-ray-langchain-and-huggingface>
- LlamaIndex. (2024). Using Gemini Embeddings with LlamaIndex. Disponible en <https://docs.llamaindex.ai/en/stable/examples/embeddings/gemini/>
- LlamaIndex. (2024). Storing Vector Data with PostgreSQL in LlamaIndex. Disponible en https://docs.llamaindex.ai/en/v0.10.19/examples/vector_xstores/postgres.html

14.1.3. Fuentes Audiovisuales

- AI with Tarum (2024). End-to-End LlamaIndex and Gemini Project [Video]. YouTube. Disponible en <https://www.youtube.com/watch?v=IxMZSweBYPw>