**Bitácora del proyecto**

**PROYECTO INTEGRADOR**

**Resolución de PQR con IA - EFIGAS**

****

**Profesor:**

**CARLOS ALBERTO RUIZ VILLA**

**Natalia Dané Arboleda Quinchía**

**Nicolás Castro Giraldo**

**Juan Sebastián Giraldo Trujillo**

**Universidad de Caldas**

**Septiembre 2024**

## **Introducción**

Como parte del equipo de desarrollo de la aplicación Resolución de PQR con IA - Efigas, la creación de esta bitácora es fundamental para documentar de manera detallada el progreso del proyecto. Este documento servirá como un registro completo y ordenado de todas las actividades realizadas, decisiones tomadas, problemas encontrados, soluciones aplicadas y los hitos alcanzados a lo largo de la ejecución del proyecto.

La bitácora es una herramienta esencial para garantizar la trazabilidad de cada etapa del desarrollo, permitiendo identificar con claridad cómo se ha evolucionado desde la fase inicial de planificación hasta el despliegue final del sistema. También asegura la transparencia en los procesos y la alineación del equipo con los objetivos definidos para el proyecto.

El registro continuo de eventos en esta bitácora no solo facilita la gestión del proyecto, sino que también contribuye al aprendizaje colectivo, proporcionando información clave para mejorar futuras implementaciones. Además, es una fuente de referencia valiosa para responder a consultas o auditorías, ya que ofrece una visión clara de las decisiones técnicas y estratégicas que influyeron en el éxito del proyecto.

## **Objetivo de la Bitácora**

El objetivo de la bitácora es registrar de manera sistemática y detallada todas las actividades, decisiones y eventos relevantes durante el desarrollo del proyecto Resolución de PQR con IA - Efigas. Este documento permitirá:

* Garantizar la trazabilidad: Mantener un historial claro de las acciones realizadas, facilitando el seguimiento del proyecto y la resolución de problemas.
* Documentar aprendizajes clave: Registrar los desafíos enfrentados y las soluciones implementadas, aportando valor al conocimiento colectivo del equipo de desarrollo.
* Apoyar la toma de decisiones: Proveer un registro que ayude a evaluar el progreso y ajustar estrategias según sea necesario.
* Servir como herramienta de referencia: Ofrecer información detallada y precisa sobre las actividades del proyecto para auditorías, consultas, para proyectos similares en el futuro o la continuación del mismo.

En resumen, la bitácora es el eje central de la documentación operativa del proyecto, asegurando que cada acción realizada y cada decisión tomada queden registradas de manera clara, estructurada y accesible para todos los interesados.

Esto se hace para llevar un control de todo el proceso que se llevó a cabo, que nos pueda ayudar más adelante.

## 

## 

## 

## **Registro de Actividades**

#### **Semana 1: Inicio y Configuración Inicial**

Durante esta semana, el equipo centró sus esfuerzos en establecer las bases del proyecto y desarrollar los primeros elementos funcionales del sistema de IA. Las actividades realizadas incluyeron ajustes a los documentos base, implementación inicial del chatbot y el diseño de propuestas para el entregable final.

* **Actividades Realizadas**:
  + **Afinar la redacción de los documentos de insumo**:
    - Se redactaron manuales y procesos en formato PDF, optimizados para facilitar su tokenización. Natalia lideró la estructuración de cada documento, asegurando que el contenido tuviera una idea principal por página. Además, se eliminaron tablas e imágenes que podrían dificultar el procesamiento semántico.
  + **Agregar referencias a las respuestas del sistema**:
    - Nicolás desarrolló una función inicial para que las respuestas generadas incluyeran referencias a los documentos base, con metadatos como el nombre del documento y la página.
  + **Implementar filtro Adaptive RAG**:
    - Sebastián lideró la implementación de un filtro que permite seleccionar los fragmentos más relevantes de los documentos vectorizados en Chroma, mejorando la precisión de las respuestas.
  + **Propuesta inicial del entregable**:
    - Se discutió si el sistema debería funcionar en WhatsApp o en una interfaz web. La decisión final quedó pendiente tras evaluar ambas opciones.
  + **Desarrollo inicial del chatbot en Streamlit**:
    - El equipo desarrolló una primera versión funcional del chatbot con opciones de selector y capacidad para generar nuevas conversaciones.
* **Resultados Alcanzados**:
  + Documentos base optimizados.
  + Filtro inicial de Adaptive RAG funcional.
  + Primera versión del chatbot con interacción básica.

#### **Semana 2: Desarrollo de Funcionalidades Clave**

Esta semana se centró en ampliar la funcionalidad del sistema y avanzar en la construcción de bases de datos vectorizadas y flujos de trabajo clave para el algoritmo de RAG.

* **Actividades Realizadas**:
  + **Ampliación de los documentos base**:
    - Natalia añadió información complementaria a los documentos de procesos junto con Aleja (que es parte del equipo de atención al cliente), asegurando que el contenido abarcara más escenarios comunes para las PQR.
  + **Creación de bases de datos vectorizadas**:
    - Se generaron embeddings para cada proceso y se integraron en la base de datos vectorizada de Chroma, permitiendo búsquedas semánticas más eficientes.
  + **Función para referencias en las respuestas**:
    - Nicolás avanzó en la implementación de una función que conecta las respuestas generadas con sus referencias exactas, facilitando la trazabilidad.
  + **Diseño de diagramas de flujo**:
    - Sebastián elaboró un diagrama detallado que describe la lógica del algoritmo **Adaptive RAG**, incluyendo el flujo desde la selección del proceso hasta la generación de respuestas.
  + **Prototipo funcional con API de WhatsApp**:
    - El equipo de desarrollo, liderado por Sebastián, desarrolló un ejemplo inicial de conexión con la API de WhatsApp, evaluando su viabilidad técnica.
* **Resultados Alcanzados**:
  + Bases vectorizadas generadas e integradas.
  + Diagramas de flujo documentados.
  + Prototipo funcional en WhatsApp.

#### **Semana 3: Incorporación de Nuevas Capacidades**

El enfoque de esta semana fue la mejora en la recuperación de información y el análisis de costos para la implementación de WhatsApp Business.

* **Actividades Realizadas**:
  + **Memoria a corto plazo**:
    - Nicolás trabajó en la incorporación de memoria a corto plazo, permitiendo al sistema almacenar temporalmente el contexto de la conversación para ofrecer respuestas más coherentes.
  + **Rediseño del sistema de referencias**:
    - Se modificó la lógica de referencias para que estas se generarán a partir de las respuestas, mejorando la precisión y relevancia.
  + **Análisis de costos**:
    - Sebastián investigó los costos asociados con la implementación de WhatsApp Business, evaluando diferentes planes y capacidades.
* **Resultados Alcanzados**:
  + Sistema con memoria a corto plazo funcional.
  + Lógica de referencias mejorada.
  + Evaluación inicial de costos completada.

#### **Semana 4: Integración y Mejoras Operativas**

Durante esta semana, se realizaron actividades enfocadas en la integración del sistema y ajustes clave para garantizar la conformidad con las normativas de datos personales.

* **Actividades Realizadas**:
  + **Integración de rutas relativas y lógica completa**:
    - Sebastián lideró la integración de **FastAPI** con la lógica del sistema **RAG**, asegurando una conexión fluida entre la recuperación y la generación de respuestas.
  + **Tratamiento de datos personales**:
    - Se agregó un aviso al inicio de la conversación para solicitar confirmación sobre el tratamiento de datos personales, cumpliendo con las normativas legales aplicables.
  + **Base de datos para Fine-Tuning**:
    - Diego(parte del equipo de trabajo de Efigas) preparó una base de datos con pares de preguntas y respuestas para entrenar el modelo con ejemplos específicos, mejorando su precisión en consultas comunes.
* **Resultados Alcanzados**:
  + Sistema integrado con FastAPI.
  + Cumplimiento con normativas de datos personales.
  + Base de datos para Fine-Tuning preparada.

#### **Semana 5: Encuestas y Optimización del Chatbot**

El equipo centró sus esfuerzos en la implementación de encuestas de satisfacción y en el ajuste de tiempos e interacciones del chatbot.

* **Actividades Realizadas**:
  + **Formulario de calificación**:
    - Se diseñó un formulario que se activa al final de cada conversación en WhatsApp, almacenando los resultados en la base de datos para análisis posterior.
  + **Rediseño del menú principal**:
    - Se exploraron alternativas para facilitar el regreso al menú principal, basándose en la propuesta del equipo de atención al cliente de Efigas.
  + **Ajustes visuales y funcionales**:
    - El equipo modificó íconos y tiempos entre mensajes para optimizar la experiencia del usuario.
* **Resultados Alcanzados**:
  + Encuesta funcional integrada en WhatsApp.
  + Menú principal rediseñado.
  + Experiencia visual mejorada.

#### **Semana 6: Refinamiento y Gestión de Datos**

Esta semana se enfocó en refinar el flujo de interacciones y garantizar la seguridad en el manejo de datos personales.

* **Actividades Realizadas**:
  + **Rediseño del flujo de consultas no resueltas**:
    - Se creó un flujo que redirige automáticamente a un agente humano si el cliente reporta insatisfacción dos veces consecutivas.
  + **Gestión de datos personales**:
    - El sistema fue ajustado para solicitar datos personales solo en casos necesarios, asegurando su almacenamiento seguro.
* **Resultados Alcanzados**:
  + Flujo optimizado para consultas complejas.
  + Manejo de datos personales seguro y conforme a las normativas.

#### **Semana 7: Despliegue y Documentación**

Durante esta semana, el equipo centró sus esfuerzos en finalizar el sistema, realizar pruebas exhaustivas y preparar el despliegue en un entorno de producción. Además, se ajustaron elementos relacionados con el manejo de datos personales.

* **Actividades Realizadas**:
  + **Revisión de la cantidad de fragmentos recuperados del documento**:
    - Se verificó la calidad y cantidad de fragmentos que el sistema recupera en base a las consultas, asegurando que las respuestas incluyan información relevante y suficiente para los usuarios.
  + **Validación de Datos Personales (Celular y Correo)**:
    - Se implementó una solicitud de datos personales con un mecanismo para validar que el correo ingresado sea válido.
  + **Gestión de "No Tratamiento de Datos Personales":**
    - Se ajustó el sistema para que, en caso de que el cliente no autorice el tratamiento de sus datos personales, el chatbot funcione en un nivel general, limitando las funcionalidades dependientes de datos específicos del cliente.
  + **Despliegue, Pruebas y Ajustes de la Herramienta:**
    - Se desplegó la herramienta en el entorno de producción y se realizaron pruebas finales para identificar posibles errores o mejoras necesarias antes de abrirla al público general. Actividad realizada por parte del equipo de Efigas
  + **Documentación del Código y Repositorio en GitHub:**
    - El código del proyecto fue documentado detalladamente y subido a un repositorio privado en GitHub. Además, se configuró la cuenta suscrita de WhatsApp Business para el manejo de interacciones en producción.
  + **Revisión de Autoría del Código:**
    - Se realizó una revisión completa del código para garantizar que todo el contenido esté alineado con los estándares de calidad y las contribuciones individuales estén correctamente atribuidas. Esto por parte del equipo de Efigas.
* **Resultados alcanzados:**
  + Cumplimiento con normativas de datos personales:
  + Mejora en la recuperación de fragmentos
  + Documentación actualizada

## **Hitos más importantes del proyecto**

#### **1. Inicio del Proyecto**

* **Fecha**: Semana 1
* **Descripción Detallada**:  
  El proyecto comenzó con la organización del equipo y la definición del alcance inicial, marcando el inicio formal del desarrollo del sistema Resolución de PQR con IA - Efigas. Durante esta etapa se asignaron roles específicos a cada integrante y se estableció un plan preliminar que delineaba las actividades clave y los plazos para cada fase del proyecto. Además, se discutieron los objetivos del sistema, priorizando la capacidad de generar respuestas basadas en documentos vectorizados mediante el enfoque **RAG** (Retrieval-Augmented Generation).  
  En esta fase, también se identificaron los documentos esenciales de Efigas (manuales, procedimientos y FAQs) que servirían como base para entrenar el sistema de IA. La decisión de dividir estos documentos en fragmentos procesables sentó las bases para el almacenamiento en la base de datos vectorizada **Chroma**.
* **Resultados Clave**:
  + Planificación inicial consolidada con un cronograma preliminar.
  + Roles definidos para cada miembro del equipo.
  + Documentos clave identificados para procesamiento.

#### **2. Elección de la Plataforma de Interacción (WhatsApp o Web)**

* **Fecha**: Semana 3
* **Descripción Detallada**:  
  Este hito fue crucial para definir cómo los usuarios finales interactuarían con el sistema de IA. Inicialmente, el equipo evaluó dos opciones principales para el front-end: una plataforma web personalizada o el uso de **WhatsApp**.  
  La opción de WhatsApp fue seleccionada por varias razones técnicas y prácticas:
  + **Familiaridad para los usuarios finales**: WhatsApp es una plataforma ampliamente conocida y utilizada por los clientes de Efigas, lo que reduce la curva de aprendizaje.
  + **Facilidad de integración**: Las APIs de WhatsApp Business permiten manejar mensajes, sesiones y funcionalidades avanzadas como menús interactivos, manteniendo un entorno robusto.
  + **Escalabilidad y seguridad**: WhatsApp ofrece infraestructura confiable con encriptación de extremo a extremo, cumpliendo con las normativas de privacidad.
* Aunque la opción de desarrollar un front-end web ofrecía mayor personalización, la simplicidad y efectividad de WhatsApp hicieron que esta fuera la elección ideal para maximizar la accesibilidad del sistema.
* **Resultados Clave**:
  + WhatsApp seleccionado como plataforma de interacción.
  + Prototipo inicial de conexión con WhatsApp desarrollado para pruebas.

#### **3. Creación de Bases de Datos Vectorizadas e Integración con RAG**

* **Fecha**: Semana 2
* **Descripción Detallada**:  
  La construcción de la base de datos vectorizada fue un hito técnico esencial para habilitar el enfoque **RAG**. El equipo procesó los documentos base (manuales, procedimientos y FAQs) para generar embeddings utilizando un modelo de procesamiento de lenguaje natural (PLN).  
  Estos embeddings fueron almacenados en **Chroma**, una base de datos vectoriales diseñada para búsquedas semánticas rápidas y precisas. La integración con el pipeline de RAG permitió que el sistema recuperara fragmentos relevantes basados en la similitud semántica entre las preguntas del usuario y los documentos almacenados.  
  Esta fase también incluyó la implementación inicial del filtro **Adaptive RAG**, que mejoró la relevancia de los fragmentos seleccionados al priorizar aquellos con mayor proximidad semántica.
* **Resultados Clave**:
  + Base de datos vectorizada funcional con documentos de Efigas.
  + Integración del sistema RAG para recuperación y generación de respuestas.
  + Mejora en la precisión de las respuestas gracias al filtro Adaptive RAG.

#### **4. Implementación de Encuestas de Satisfacción**

* **Fecha**: Semana 5
* **Descripción Detallada**:  
  Como parte del compromiso de garantizar una experiencia de usuario positiva, se implementó un sistema automatizado de encuestas de satisfacción al final de cada conversación en WhatsApp. Este formulario permite recopilar datos sobre la percepción del cliente en aspectos como claridad, tiempo de respuesta y efectividad de las respuestas generadas por la IA.  
  El equipo desarrolló un formulario interactivo que se activa automáticamente una vez que el cliente finaliza su consulta. Los datos recopilados se almacenan en una base de datos para análisis posterior, permitiendo identificar patrones de satisfacción e insatisfacción. Además, se configuró una función que finaliza la conversación si el cliente no completa la encuesta después de 5 minutos, optimizando la gestión de recursos.
* **Resultados Clave**:
  + Encuestas de satisfacción integradas en el flujo de interacción.
  + Datos recopilados para análisis estadístico y mejoras futuras.
  + Mejora en la capacidad de monitorear la experiencia del usuario.

#### **5. Despliegue del Sistema en Producción**

* **Fecha**: Semana 7
* **Descripción Detallada**:  
  Este hito marcó la culminación de las fases de desarrollo, integración y pruebas, con el despliegue del sistema completo en un entorno de producción. El equipo realizó pruebas exhaustivas para garantizar la estabilidad y funcionalidad del sistema, evaluando tiempos de respuesta, precisión en la recuperación de información y compatibilidad con WhatsApp Business.  
  El despliegue también incluyó la configuración de la infraestructura en la nube para garantizar la escalabilidad y fiabilidad del sistema. Se documentó el código en un repositorio de GitHub, facilitando la gestión y mantenimiento del proyecto a largo plazo.
* **Resultados Clave**:
  + Sistema operativo y accesible para los clientes finales de Efigas.
  + Código documentado y repositorio organizado en GitHub.
  + Pruebas finales realizadas con éxito, garantizando una experiencia de usuario óptima.

## **Problemas y Soluciones del Proyecto**

Durante el desarrollo del proyecto **Resolución de PQR con IA - Efigas**, el equipo enfrentó diversos desafíos técnicos y estratégicos. A continuación, se describen los problemas más significativos encontrados y las soluciones implementadas para superarlos:

#### **1. Problema: Complejidad en la Tokenización y Generación de Embeddings**

* **Descripción del Problema**:  
  Los documentos base, como manuales y procedimientos, contenían información no estructurada, como tablas, imágenes y formatos inconsistentes. Esto dificulta el proceso de tokenización y generación de embeddings, ya que el contenido no estaba optimizado para su procesamiento por el modelo de lenguaje natural.
* **Solución Implementada**:
  + Se realizó un proceso de limpieza exhaustivo, eliminando elementos no procesables, como imágenes y formatos innecesarios, y dividiendo el texto en fragmentos manejables.
  + Se revisaron los documentos procesados para asegurar que cada fragmento capturara una idea principal, optimizando la calidad de los embeddings.
* **Resultados**:
  + Los documentos fueron tokenizados exitosamente y almacenados en la base de datos vectorizada **Chroma**, asegurando búsquedas semánticas precisas.

#### **2. Problema: Decisión entre WhatsApp o un Front-End Web**

* **Descripción del Problema**:  
  Durante las primeras semanas, surgió una discusión en el equipo sobre qué plataforma sería más adecuada como front-end para el sistema: un diseño web personalizado o WhatsApp. La opción de desarrollar un front-end ofrecía mayor personalización, pero requería tiempo y recursos significativos.
* **Solución Implementada**:
  + Se realizaron análisis técnicos y de usabilidad, destacando que WhatsApp era una plataforma conocida por los clientes de Efigas y requería menor tiempo de implementación.
  + Se creó un prototipo funcional en WhatsApp utilizando la **API de WhatsApp Business**, lo que permitió realizar pruebas iniciales y validar su viabilidad técnica.
* **Resultados**:
  + WhatsApp fue elegido como plataforma de interacción, facilitando la adopción por parte de los usuarios y optimizando los recursos del equipo.

#### **3. Problema: Baja Relevancia en la Recuperación de Fragmentos con RAG**

* **Descripción del Problema**:  
  Durante las pruebas iniciales, el sistema **RAG** no siempre seleccionaba fragmentos de texto que respondieran con precisión a las preguntas de los usuarios. Esto generaba respuestas poco útiles o ambiguas.
* **Solución Implementada**:
  + Se implementó el filtro **Adaptive RAG**, que analiza la similitud semántica de los fragmentos recuperados y prioriza aquellos con mayor relevancia para la consulta.
  + Además, se ajustaron los umbrales de similitud en la búsqueda semántica de Chroma para mejorar la precisión.
* **Resultados**:
  + La relevancia de los fragmentos recuperados mejoró significativamente, reduciendo las respuestas ambiguas y aumentando la precisión del sistema.

#### **4. Problema: Rechazo del Usuario al Compartir Datos Personales**

* **Descripción del Problema**:  
  Durante las pruebas con usuarios finales, se identificó que algunos clientes no se sentían cómodos compartiendo datos personales, lo que limitaba la personalización de las respuestas y la recopilación de métricas para análisis estadístico.
* **Solución Implementada**:
  + Se agregó un aviso al inicio de la interacción en WhatsApp, solicitando explícitamente el consentimiento para el tratamiento de datos personales.
  + El sistema fue ajustado para funcionar sin necesidad de datos personales si el usuario no daba su consentimiento.
* **Resultados**:
  + Los usuarios que no compartían sus datos aún podían utilizar el sistema, garantizando el cumplimiento de normativas de privacidad y aumentando la confianza del cliente.

#### **5. Problema: Escalabilidad del Sistema en WhatsApp Business**

* **Descripción del Problema**:  
  Durante las pruebas, el sistema enfrentó limitaciones en la gestión de múltiples usuarios simultáneamente en la API de WhatsApp Business, lo que generó demoras en las respuestas.
* **Solución Implementada**:
  + Se configuraron instancias adicionales en la nube para manejar sesiones simultáneas y evitar sobrecarga en los servidores.
  + Se optimizó el tiempo de respuesta reduciendo los pasos innecesarios en el pipeline del sistema.
* **Resultados**:
  + El sistema pudo manejar un mayor volumen de usuarios simultáneamente, garantizando una experiencia fluida incluso en horas pico.

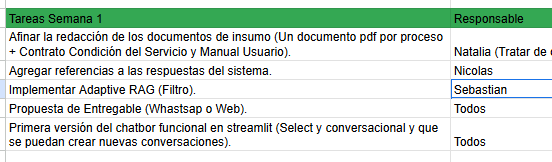
#### **6. Problema: Falta de Métricas para Medir la Satisfacción del Usuario**

* **Descripción del Problema**:  
  Al inicio del desarrollo, el sistema no tenía un mecanismo para medir la satisfacción del usuario ni identificar áreas de mejora basadas en la experiencia del cliente.
* **Solución Implementada**:
  + Se implementó una encuesta de satisfacción que se activa al finalizar cada interacción en WhatsApp.
  + Los resultados se almacenan en una base de datos para análisis posterior, permitiendo evaluar indicadores como claridad, tiempo de respuesta y efectividad.
* **Resultados**:
  + El equipo cuenta ahora con datos estadísticos que permiten realizar mejoras continuas en el sistema y ajustar las respuestas generadas según las necesidades del cliente.

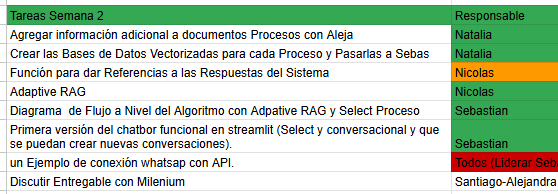
## **Anexos**

Los anexos contienen documentación adicional que respalda y complementa la información presentada en el cuerpo principal de la bitácora. Estos materiales ofrecen detalles técnicos, registros específicos y referencias clave utilizadas durante el desarrollo del proyecto.

### Actividades de la semana 1



### Actividades de la semana 2



Actividades de la semana 3

