**Documento de Pruebas**

**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

Este documento describe las pruebas realizadas al sistema de inteligencia artificial desarrollado para la resolución de PQR de Efigas. Se detallan los casos de prueba más importantes, los resultados obtenidos, y las limitaciones encontradas, con el propósito de garantizar la funcionalidad, usabilidad y conformidad del sistema con los requisitos iniciales.

## Objetivo

Validar el funcionamiento del sistema **Resolución de PQR con IA - Efigas** mediante la ejecución de casos de prueba diseñados para evaluar su desempeño en distintos escenarios, identificando posibles fallos y áreas de mejora para futuras iteraciones del proyecto.

## Alcance del Proyecto

El sistema desarrollado tiene como objetivo automatizar la resolución de preguntas, quejas y reclamos de los clientes de Efigas mediante el uso de inteligencia artificial integrada con WhatsApp.

#### **Limitaciones**

* El sistema no maneja consultas fuera del alcance de los documentos procesados (manuales, procedimientos y FAQs de Efigas).
* En proyectos futuros, se planea ampliar la capacidad del sistema para:
  + Procesar consultas sobre nuevos documentos no incluidos en la base actual.
  + Incorporar mejoras en el manejo de lenguaje ambiguo o consultas complejas.
  + Optimizar la experiencia del usuario en situaciones de inactividad prolongada.

## Metodología de Pruebas

Las pruebas se realizaron utilizando un enfoque estructurado que incluye:

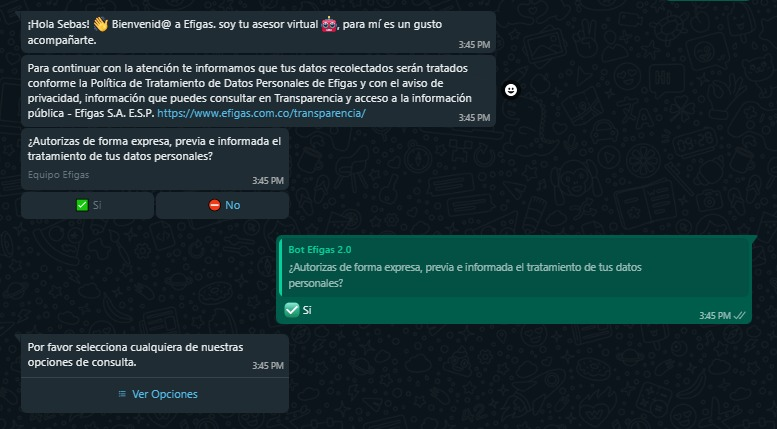
1. **Pruebas Funcionales**: Validación de que las funciones principales del sistema (recuperación de fragmentos, generación de respuestas, manejo de datos personales) operan según lo esperado.
2. **Pruebas de Usabilidad**: Evaluación de la experiencia del usuario final al interactuar con el sistema mediante WhatsApp.
3. **Pruebas de Rendimiento**: Medición de tiempos de respuesta y manejo de múltiples consultas simultáneas.

### **Casos de pruebas**

| **Caso de Prueba** | **Descripción** | **Resultado Esperado** | **Resultado Obtenido** |
| --- | --- | --- | --- |
| Inicio de conversación con aceptación de datos personales | Verificar que el sistema solicite aceptación del tratamiento de datos personales al iniciar. | Mensaje inicial claro y registro correcto. | Cumple |
| Consulta sobre documento "Brilla" | Validar que el sistema recupera fragmentos relevantes relacionados con el documento específico. | Respuesta precisa y relevante basada en el documento. | Cumple |
| No aceptación de tratamiento de datos personales | Confirmar que el sistema funcione a nivel general si no se aceptan los términos de datos. | Respuestas genéricas y funcionalidad básica. | Cumple |
| Encuesta de satisfacción | Asegurar que la encuesta se active al finalizar la consulta. | Activación automática de la encuesta final. | Cumple |
| Manejo de inactividad | Verificar que el sistema envíe mensajes de alerta tras inactividad y finalice la consulta. | Cuatro alertas antes de cerrar sesión. | Cumple |
| Manejo de fallos en respuesta | Verificar que las respuestas que se le dan al usuario sean las correctas y de no ser así, en el segundo error se le permite contactar con un asesor. | Al segundo fallo de respuesta le da opción de pasar con asesor | Cumple |

### **Resultados**

1. Inicio de conversación y aceptación del tratamiento de datos personales:

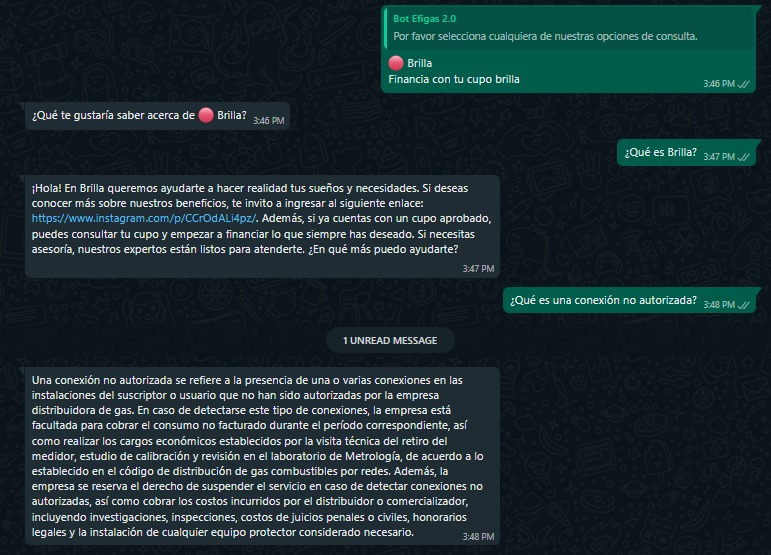


1. Ejemplo de consulta, para este caso sobre el documento de brilla.

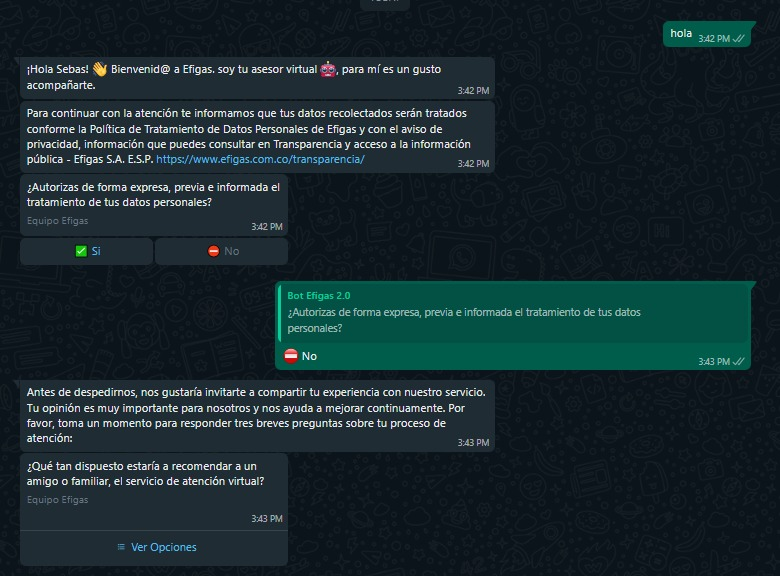
Se muestra unas opciones de documento sobre el cual se hará la consulta, porque de allí el sistema traerá el fragmento a responder.



Al seleccionar el documento correspondiente ya se puede hacer la pregunta para que nuestro asistente genere su respectiva respuesta.



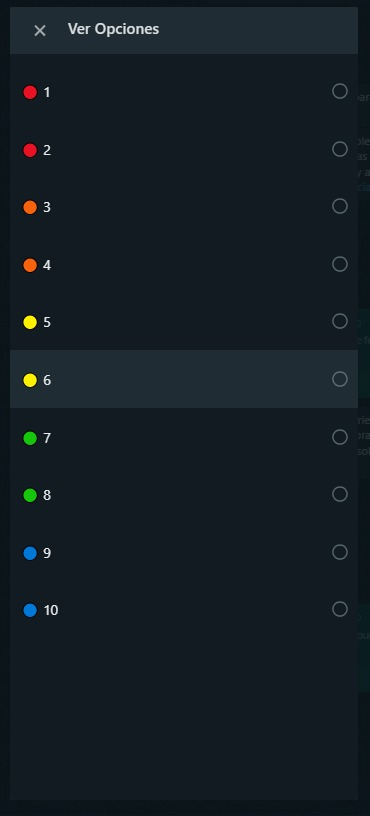
1. No se aceptan los términos y condiciones para tratamiento de datos personales. Lo que lleva a que el sistema inmediatamente te lleve a la encuesta de finalización de la consulta.



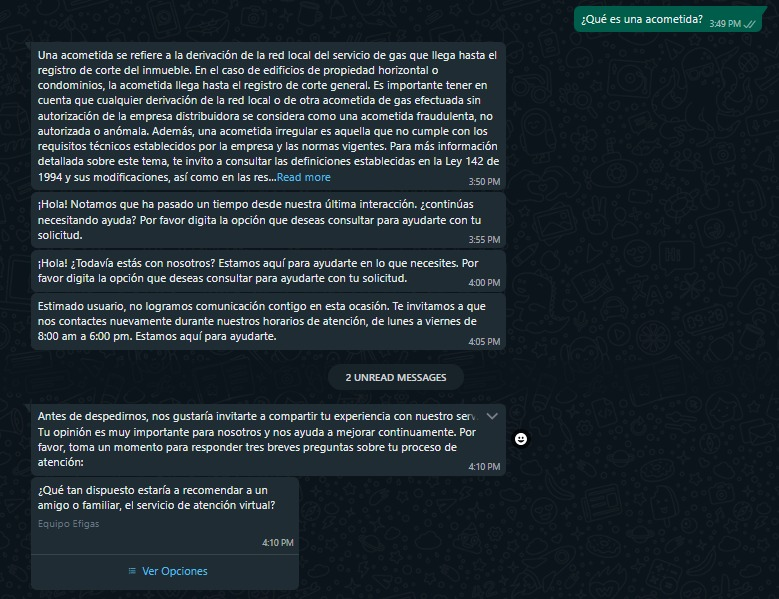
1. Encuesta de satisfacción. Se realiza una serie de preguntas sobre la atención que acabó de recibir



Opciones de la pregunta de calificación. Sobre qué tan dispuesto estaría a recomendar el servicio de atención al cliente.



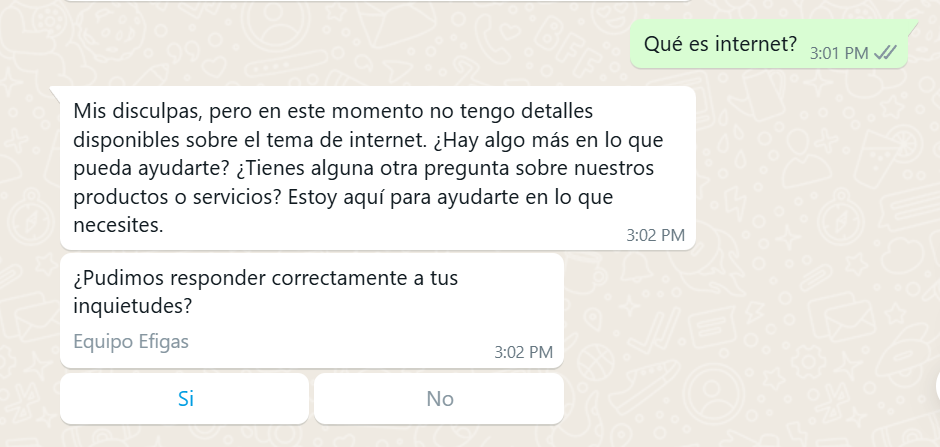
1. Manejo de inactividad



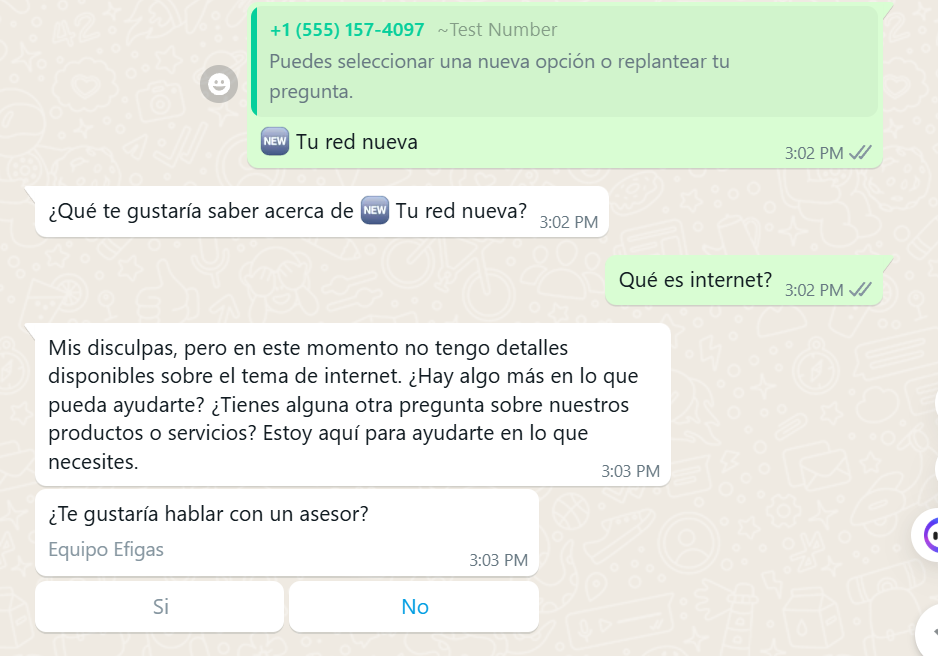
Sistema de salida y cierre de sesión cuando se deja de responder (envía cada 5 minutos un mensaje de alerta para decirle a usuario que si desea con la consulta, son 4 mensajes donde el último ya cuenta con la encuesta para finalizar)

1. Si falla 2 veces en la respuesta indicada le da la opción de pasar con un asesor.

En el primer error pide que se redacte de nuevo la pregunta.



Para el segundo error, ya da la posibilidad de pasar directamente con un asesor para buscar la satisfacción del cliente.



### **Conclusiones finales de los resultados**

De las pruebas realizadas, se observaron los siguientes resultados:

1. **Aspectos Funcionales Positivos**:
   * El sistema responde con precisión en la mayoría de los casos, recuperando fragmentos relevantes de los documentos.
   * La integración con WhatsApp permite una interacción fluida y sin interrupciones técnicas significativas.
2. **Áreas de Mejora Detectadas**:
   * En algunos casos de lenguaje ambiguo, el sistema genera respuestas menos precisas. Esto se debe a limitaciones en la base de datos vectorizada y el pipeline de RAG.
   * Se requiere una optimización adicional en el manejo de consultas simultáneas para evitar sobrecarga en horarios pico.

### **Conclusiones**

* Las pruebas confirmaron que el sistema cumple con la mayoría de los requisitos funcionales y de usabilidad establecidos en el proyecto inicial. Sin embargo, se identificaron oportunidades de mejora que pueden abordarse en futuras iteraciones para expandir el alcance y la capacidad del sistema, fortaleciendo su precisión y eficiencia.
* Debido a las restricciones de tiempo y alcance definidas inicialmente, quedaron pendientes varios ajustes e implementaciones, lo que ha limitado la ejecución de pruebas más allá de las implementaciones.