



|  |
| --- |

Caso de Estudio EcoMarket

**Profesor:** Alexis Andrés Duran Contreras. **Integrantes:** Santiago López, Harold Peralta.  
**Asignatura:** Desarrollo de soluciones con inteligencia artificial  
**Fecha de entrega:** 22/09/2025.

# Índice

[**Índice 1**](#_btabwiblswtb)

[1 Introducción 2](#_ag1gc8xzv750)

[2.2 Necesidades 2](#_t473edyw02b2)

[2.3 Justificación del Uso de LLM y RAG 3](#_o68dzo7w09c)

[3. Formulación de Prompts 3](#_xd2ww1isfnw)

[3.1 Estrategias de Optimización 3](#_r6q816y55fdn)

[3.2 Ejemplos de Prompts 4](#_1lglvoce5f6c)

[3.3 Justificación 4](#_1k4y5a4q2alo)

[4. Diseño del Pipeline RAG 5](#_71bp8fienrl5)

[4.1 Diagrama del Flujo de Trabajo 5](#_9wgxlghgwwv8)

[4.2 Componentes del Pipeline 6](#_rkbyvs7vz27z)

[4.3 Beneficios 6](#_th3drforkmqt)

[7. Arquitectura de la Solución 7](#_wmmazee0gxdu)

[7.1 Esquema General 7](#_ejuo77klctt7)

[7.2 Selección de LLM 7](#_7tkdybiplqqf)

[7.3 Escalabilidad y Seguridad 7](#_knbg39ya0mdq)

[8. Implementación 7](#_495do7fb2y12)

[8.1 Herramientas Usadas 7](#_2r1apywna4md)

[8.2 Ejemplos de Implementación 7](#_ro5xbmuvz7o1)

[8.3 Retos y Soluciones 8](#_5xtx1ktn8rqp)

[9. Evaluación y Resultados 8](#_o418ix9qvnh7)

[9.1 Métricas 8](#_3yyp76vbwb0u)

[9.2 Comparativa con Sistema Anterior 8](#_g80xaq4w8him)

[10. Conclusiones 8](#_9fxmrj3ndfri)

[11. Referencias 9](#_o27ak8nr0nzu)

[12. Anexos 9](#_qasrjbgftpb4)

## 

## 

## 

## 

## 

## 

## 

## 

## 1 Introducción

Contexto:EcoMarket SPA es una empresa chilena emergente dedicada a la venta de productos ecológicos y sostenibles. Comenzó con una tienda en el Barrio Lastarria, Santiago, y gracias a su éxito y compromiso con la sostenibilidad, ha expandido su presencia a Valdivia y Antofagasta.

El crecimiento de la empresa ha generado desafíos operativos: el sistema monolítico actual presenta problemas de rendimiento y disponibilidad, afectando la eficiencia y la experiencia del cliente. La atención al cliente se ve especialmente afectada, ya que gran parte del tiempo se dedica a orientar a los usuarios hacia el departamento correcto.

Para solucionar esto, se propone implementar un **bot de atención al cliente**, que permita responder consultas frecuentes de manera automática, guiar a los clientes al área adecuada y optimizar los procesos internos, mejorando tanto la experiencia del usuario como la productividad del equipo.

* Motivación: la necesidad de modernizar la atención al cliente y mejorar eficiencia operativa.
* Alcance: diseño e implementación de un pipeline RAG con LLM para un bot que gestione consultas de clientes y soporte interno.

### 1.2 Necesidades

La empresa requiere soluciones que permitan optimizar la atención al cliente y mejorar la eficiencia operativa. Entre las necesidades principales se encuentran:

* **Responder consultas frecuentes de manera rápida y precisa:** Actualmente, muchas preguntas de clientes son repetitivas y podrían ser atendidas automáticamente mediante un sistema inteligente.
* **Derivar automáticamente las consultas al departamento correcto:** Reducir la dependencia del personal para redirigir consultas y asegurar que los usuarios obtengan la información correcta de manera inmediata.
* **Reducir tiempo de trabajo repetitivo del personal:** Liberar al equipo de atención al cliente de tareas rutinarias, permitiéndoles enfocarse en consultas más complejas y estratégicas.
* **Mejorar la experiencia del cliente:** Ofrecer respuestas consistentes, rápidas y disponibles 24/7, aumentando la satisfacción y fidelización de los clientes.

### 1.3 Justificación del Uso de LLM y RAG

Para abordar los desafíos mencionados, se propone la implementación de un **bot de atención al cliente** basado en un pipeline **RAG (Retrieval-Augmented Generation) con LLM (Large Language Model)**.

* **LLM (Large Language Model):** Permite generar respuestas naturales, coherentes y contextuales ante las consultas de los clientes. Su capacidad de comprensión del lenguaje facilita la interacción fluida, simulando una atención humana.
* **RAG (Retrieval-Augmented Generation):** Permite recuperar información precisa de fuentes internas como FAQ, manuales de productos y bases de datos. Esto asegura que las respuestas del bot estén siempre actualizadas y sean exactas, incluso ante preguntas complejas o poco frecuentes.

**Beneficios esperados:**

* Reducción del tiempo de respuesta y mejora en la satisfacción del cliente.
* Optimización de la productividad del personal, liberándose de tareas repetitivas.
* Incremento de la eficiencia operativa y soporte para el crecimiento de la empresa sin comprometer la calidad del servicio.

## 2 Formulación de Prompts

### 2.1 Estrategias de Optimización

La efectividad de un bot basado en LLM depende en gran medida de la **calidad de los prompts** utilizados. Para EcoMarket SPA, se han definido las siguientes estrategias:

* **Definición de prompts claros, concisos y específicos:** Los prompts deben guiar al modelo para generar respuestas precisas, evitando ambigüedades que puedan generar información incorrecta o confusa.
* **Adaptación del lenguaje al estilo de la empresa:** El bot debe comunicarse de manera **amigable y profesional**, reflejando la identidad de EcoMarket SPA y generando confianza en los clientes.
* **Segmentación de prompts por función:** Se diseñan prompts específicos para distintos objetivos, como responder preguntas frecuentes, derivar consultas al departamento adecuado y guiar a los usuarios en procesos internos.
* **Incorporación de contexto relevante:** Los prompts incluyen información sobre productos, políticas de la empresa y procedimientos internos para asegurar que las respuestas sean precisas y útiles.

### 2.2 Ejemplos de Prompts

A continuación, se presentan algunos ejemplos representativos de prompts para el bot de EcoMarket SPA:

1. **Respuesta a consultas de productos:** *“Como bot de EcoMarket SPA, responde preguntas sobre nuestros productos ecológicos de forma clara, amigable y profesional. Incluye información sobre disponibilidad, beneficios ecológicos y formas de uso.”*
2. **Derivación de consultas a departamentos:** *“Analiza la consulta del cliente y determina el departamento más adecuado para atenderla (ventas, soporte técnico, logística, devoluciones). Indica la acción que debe seguir el usuario para contactarse con el departamento correcto.”*
3. **Guía de atención general:** *“Si un cliente realiza una pregunta general sobre EcoMarket SPA o nuestros servicios, proporciona una respuesta cordial, resumida y útil, e invita al usuario a consultar más información en nuestra web o contacto directo.”*
4. **Manejo de preguntas complejas:** *“Si la consulta del cliente es compleja o no está en la base de datos, sugiere pasos claros para que pueda resolver su problema o derivarlo al equipo humano correspondiente.”*

### 2.3 Justificación

Cada prompt cumple un objetivo estratégico:

* **Precisión en la respuesta:** Los prompts claros y específicos permiten al bot generar información correcta, evitando errores o malentendidos.
* **Coherencia en el lenguaje:** La adaptación al estilo de EcoMarket SPA asegura que la comunicación sea consistente, profesional y cercana al cliente.
* **Correcta derivación de consultas:** Los prompts que orientan la derivación de consultas reducen la carga del personal y mejoran la eficiencia operativa.
* **Flexibilidad y escalabilidad:** Una formulación cuidadosa permite actualizar los prompts fácilmente conforme la empresa crece o se agregan nuevos productos y servicios.

Con esta estrategia de prompts, el bot se convierte en una herramienta confiable, capaz de mantener la calidad del servicio incluso frente a un alto volumen de consultas y clientes.

## 3. Diseño del Pipeline RAG

### 3.1 Diagrama del Flujo de Trabajo

El pipeline RAG (Retrieval-Augmented Generation) permite combinar la **recuperación de información relevante** con la **generación de respuestas naturales** por un LLM. El flujo de trabajo propuesto para EcoMarket SPA es el siguiente:

1. **Recepción de la consulta del cliente:** El cliente envía su consulta a través de la web, la app móvil o canales de mensajería integrados.
2. **Recuperación de información relevante:** El sistema busca en la base de datos, FAQ, manuales de productos y documentación interna para obtener información pertinente que ayude a contextualizar la respuesta.
3. **Procesamiento por LLM con contexto recuperado:** El LLM recibe la consulta del cliente junto con la información recuperada, generando una respuesta coherente, precisa y alineada al estilo de EcoMarket SPA.
4. **Generación de respuesta y/o derivación al departamento adecuado:** El bot entrega la respuesta al cliente o, si la consulta requiere intervención humana, indica el departamento correcto para derivarla, asegurando eficiencia y continuidad en la atención.

### 3.2 Componentes del Pipeline

El pipeline RAG se compone de varias tecnologías y procesos que trabajan de manera integrada:

* **Indexación de documentos:** Se crean embeddings semánticos de los documentos internos (FAQ, manuales, políticas), permitiendo que la información sea buscada de forma eficiente según el contenido, no solo palabras clave.
* **Búsqueda semántica:** Se utilizan motores como **FAISS, Pinecone o Weaviate** para realizar búsquedas rápidas y relevantes en grandes volúmenes de información, asegurando que el LLM reciba únicamente datos útiles para generar la respuesta.
* **Integración con LLM:** Modelos como **GPT-4 o LLaMA2** procesan la consulta junto con la información recuperada, generando respuestas naturales, claras y adaptadas al estilo de la empresa.
* **Módulo de derivación:** Determina, con base en la consulta, si se requiere intervención humana y a qué departamento se debe redirigir, optimizando la eficiencia del equipo.

### 3.3 Beneficios

La implementación de un pipeline RAG con LLM para EcoMarket SPA ofrece ventajas significativas:

* **Respuestas precisas y contextualizadas:** La combinación de recuperación de información y generación de lenguaje natural asegura que los clientes reciban información correcta y coherente.
* **Reducción de tiempo de atención y carga del personal:** El bot se encarga de preguntas frecuentes y derivaciones automáticas, liberando al personal para tareas más complejas.
* **Mejora de la experiencia del cliente:** Atención rápida, personalizada y disponible 24/7, aumentando la satisfacción y fidelización.
* **Escalabilidad y adaptabilidad:** El sistema puede crecer junto con la empresa, incorporando nuevas fuentes de información y ajustando los prompts según sea necesario.

## 4. Arquitectura de la Solución

### 4.1 Esquema General

La arquitectura de la solución propone un **bot de atención al cliente** integrado directamente en el sitio web y la aplicación móvil de EcoMarket SPA. Este bot se conecta a un **pipeline RAG**, que combina la recuperación de información relevante con un **LLM** para generar respuestas naturales y contextuales.

**Flujo de interacción:** Cliente → Bot (Web/App) → Pipeline RAG → LLM → Respuesta o derivación → Cliente.

Esto asegura que la atención sea **rápida, precisa y coherente con la identidad de la empresa**, y que el personal humano solo intervenga cuando sea estrictamente necesario.

### 4.2 Selección de LLM

Se recomienda el uso de **GPT-4** debido a:

* Alta capacidad de comprensión del lenguaje natural.
* Generación de respuestas coherentes y adaptadas al estilo amigable y profesional de EcoMarket SPA.
* Capacidad para integrarse con RAG y manejar consultas complejas con información contextual.

El LLM se configura para priorizar **precisión, claridad y coherencia**, alineándose con los objetivos de eficiencia operativa y satisfacción del cliente.

### 4.3 Escalabilidad y Seguridad

Para garantizar un sistema confiable y seguro, se consideran los siguientes aspectos:

* **Manejo seguro de datos de clientes:** Cumplimiento de normativas locales de protección de datos (por ejemplo, Ley de Protección de la Vida Privada en Chile).
* **Caching de consultas frecuentes:** Almacenar temporalmente respuestas recurrentes para reducir tiempo de procesamiento y carga en el LLM.
* **Monitoreo de performance y registro de interacciones:** Control de métricas como tiempo de respuesta, consultas atendidas y derivaciones realizadas, para optimizar continuamente la eficiencia del bot.

## 5. Implementación

### 5.1 Herramientas Usadas

Para la implementación se emplean tecnologías especializadas que permiten integrar el LLM con RAG y optimizar la experiencia del cliente:

* **LangChain:** Facilita la integración del LLM con pipelines RAG, gestión de prompts y recuperación de información.
* **FAISS / Pinecone:** Plataformas de búsqueda vectorial para realizar consultas semánticas rápidas y precisas en grandes volúmenes de información.
* **Backend:** Python o Node.js, según la infraestructura existente en EcoMarket SPA, para manejar la lógica del bot y la comunicación con el LLM y las bases de datos.

### 5.2 Ejemplos de Implementación

La implementación incluye:

* **Fragmentos de código para prompts, embeddings y recuperación de información:** Permiten configurar cómo el bot procesa las consultas y accede a los datos relevantes.
* **Simulaciones de interacción cliente-bot:** Ensayos para validar que el bot responde correctamente, deriva consultas cuando es necesario y mantiene el estilo comunicacional de EcoMarket SPA.

Ejemplo conceptual:

* Consulta del cliente → Se genera embedding → Búsqueda semántica en FAISS → Contexto enviado al LLM → Respuesta generada → Bot entrega respuesta o deriva al departamento.

### 5.3 Retos y Soluciones

Durante la implementación pueden surgir desafíos técnicos, con soluciones específicas:

* **Indexación de documentos no estructurados:** Se realiza preprocesamiento y limpieza de datos para asegurar que la información pueda ser buscada y utilizada correctamente por el LLM.
* **Derivación de consultas ambiguas:** Se diseñan **prompts iterativos** que validan contexto antes de derivar la consulta, minimizando errores y mejorando la experiencia del cliente.
* **Optimización de rendimiento:** Uso de caching y búsqueda semántica para reducir latencia y mejorar la eficiencia del sistema.

## 6. Evaluación y Resultados

### 6.1 Métricas

Para evaluar la efectividad del bot de EcoMarket SPA, se establecen métricas clave que permiten cuantificar mejoras operativas y de atención al cliente:

* **Tiempo promedio de respuesta:** Se mide cuánto tarda el bot en entregar respuestas a las consultas de los clientes. El objetivo es reducir significativamente los tiempos respecto al sistema anterior.
* **Precisión en derivación de consultas:** Se registra el porcentaje de consultas correctamente dirigidas al departamento adecuado, minimizando errores y la necesidad de intervención humana.
* **Reducción de carga de trabajo del personal:** Se calcula el porcentaje de consultas atendidas automáticamente por el bot frente al total de consultas, evidenciando cuánto se ha liberado al equipo de atención al cliente.
* **Satisfacción del cliente:** A través de encuestas breves o sistemas de rating tras cada interacción, se evalúa la percepción del cliente sobre la rapidez, claridad y utilidad de las respuestas.

Estas métricas permiten realizar un seguimiento continuo del desempeño del sistema, identificar oportunidades de mejora y garantizar que los objetivos de eficiencia y calidad de servicio se cumplan.

### 6.2 Comparativa con Sistema Anterior

Al comparar el nuevo sistema con el sistema monolítico anterior, se esperan las siguientes mejoras:

* **Velocidad:** Reducción significativa en los tiempos de respuesta, gracias al procesamiento automático y a la recuperación de información semántica.
* **Exactitud:** Mayor precisión en la información entregada y en la derivación de consultas, gracias a la integración de RAG con LLM y prompts bien diseñados.
* **Satisfacción del cliente:** Experiencia más fluida y consistente, con atención disponible 24/7 y respuestas adaptadas al estilo de la empresa.
* **Eficiencia operativa:** Disminución del trabajo repetitivo para el personal, permitiendo enfocarse en consultas complejas o estratégicas, aumentando la productividad general.

En conjunto, estas mejoras permiten a EcoMarket SPA mantener la calidad del servicio mientras se adapta al crecimiento de clientes y al aumento de consultas, asegurando tanto la eficiencia operativa como la fidelización de los usuarios.

## 7 Conclusiones

### 7.1 Resumen

La implementación del bot de atención al cliente basado en **pipeline RAG con LLM** ha permitido identificar y confirmar varios beneficios clave para EcoMarket SPA:

* **Atención más rápida y eficiente:** Las consultas frecuentes son respondidas automáticamente, reduciendo los tiempos de espera de los clientes.
* **Reducción de errores humanos:** La derivación automática de consultas al departamento correcto disminuye equivocaciones y mejora la consistencia de la información entregada.
* **Mejora en la experiencia del cliente:** La interacción es más fluida, coherente y alineada al estilo amigable y profesional de la empresa, aumentando la satisfacción y fidelización.
* **Optimización del trabajo del personal:** El equipo de atención al cliente se libera de tareas repetitivas, pudiendo enfocarse en consultas complejas y tareas estratégicas.

Estos hallazgos confirman que la solución tecnológica no solo mejora la eficiencia operativa, sino que también fortalece la relación con los clientes y la capacidad de EcoMarket SPA de escalar su servicio.

### 7.2 Recomendaciones

Para maximizar el impacto del proyecto y garantizar su sostenibilidad, se proponen las siguientes acciones:

1. **Expansión a nuevos canales de atención:** Integrar el bot en plataformas como WhatsApp, redes sociales y chat en aplicaciones móviles para alcanzar más clientes y ofrecer atención omnicanal.
2. **Actualización continua de la base de conocimiento:** Mantener la información de FAQ, manuales y documentación interna actualizada, asegurando que las respuestas del bot sigan siendo precisas y relevantes.
3. **Monitoreo y ajuste constante de prompts:** Revisar y optimizar periódicamente los prompts para mejorar la precisión, coherencia y capacidad de derivación del bot.
4. **Análisis de métricas de desempeño:** Evaluar constantemente tiempo de respuesta, precisión y satisfacción del cliente para detectar oportunidades de mejora y garantizar la eficiencia del sistema.

Con estas recomendaciones, EcoMarket SPA puede consolidar un servicio de atención automatizado robusto, escalable y alineado con su crecimiento y compromiso con la sostenibilidad y calidad de atención.

## 8. Anexos

**Link Github:**

https://github.com/HAPEMA/Evaluacion-IA

**Link Presentacion:** https://www.canva.com/design/DAGzvF5\_PC0/FWGZM27-kWLJR\_1MOhIK1A/edit?utm\_content=DAGzvF5\_PC0&utm\_campaign=designshare&utm\_medium=link2&utm\_source=sharebutton