

**INTEGRACIÓN DEL MODELO DE INTELIGENCIA ARTIFICIAL GPT-4o mini
CON EL SOFTWARE DE COMPRAS HANDELBAY**

**JUAN CAMILO VALENCIA HENAO
MAIKOL RICO PÉREZ**



**UNIVERSIDAD DE
MANIZALES**

**UNIVERSIDAD DE MANIZALES
FACULTAD DE CIENCIAS E INGENIERÍA
PROGRAMA
MANIZALES
2024**

**INTEGRACIÓN DEL MODELO DE INTELIGENCIA ARTIFICIAL GPT-4o mini
CON EL SOFTWARE DE COMPRAS HANDELBAY**

**JUAN CAMILO VALENCIA HENAO
MAIKOL RICO PÉREZ**

Trabajo de Grado presentado como opción parcial para optar
al título de Ingeniero en Sistemas y Telecomunicaciones

Presidente / Asesor temático
CARLOS BETANCOURT CORREA

**UNIVERSIDAD DE MANIZALES
FACULTAD DE CIENCIAS E INGENIERÍA
PROGRAMA
MANIZALES
2024**

CONTENIDO

	Pág.
1. ÁREA PROBLEMÁTICA	13
1.1 DESCRIPCIÓN	13
1.2 DELIMITACIÓN	13
1.3 FORMULACIÓN	14
2. OBJETIVOS	15
2.1 OBJETIVO GENERAL	15
2.2 OBJETIVOS ESPECÍFICOS	15
3. JUSTIFICACIÓN	16
3.1 NOVEDAD	16
3.2 INTERÉS	16
3.3 UTILIDAD	16
4. MARCO TEÓRICO	17
4.1 INTELIGENCIA ARTIFICIAL (IA)	17
4.1.1 Inteligencia Artificial Generativa.	19
4.1.2 Modelos de Lenguaje Grande (LLM).	21
4.1.3 Procesamiento del Lenguaje Natural (PLN).	22
4.1.4 GPT (Generative Pre-trained Transformer).	23
4.2 SISTEMAS DE GESTIÓN DE COMPRAS	27
4.3 TECNOLOGÍA EN COMPRAS EMPRESARIALES	28
4.4 REFERENTE NORMATIVO Y LEGAL	29
4.5 ANTECEDENTES	31
5. METODOLOGÍA	34
5.1 TIPO DE TRABAJO	34
5.2 PROCEDIMIENTO	34
5.2.1 Fase 1. Revisión de la necesidad.	34
5.2.2 Fase 2. Diseño y Arquitectura.	35
5.2.3 Fase 3. Desarrollo.	35
5.2.4. Fase 4. Pruebas.	36
5.2.5 Fase 5. Documentación.	36
6. RESULTADOS	37
6.1 DESCRIPCIÓN DE RESULTADOS	37

6.2 DISCUSIÓN DE RESULTADOS	62
7. CONCLUSIONES	64
8. RECOMENDACIONES	65
REFERENCIAS BIBLIOGRÁFICAS	66
ANEXO A - Casos de uso	69
ANEXO B - Diagramas	75
ANEXO C - Pruebas	78
ANEXO D - Manual de uso	87
ANEXO E - Manual de uso (video explicativo)	92
ANEXO F - Manual técnico	93
ANEXO G - Manual técnico (video explicativo)	102
ANEXO H - Comparativo de indicadores de eficiencia	103
ANEXO I - Demo de la integración en producción	106
ANEXO J - Resumen Analítico	107

LISTA DE FIGURAS

	Pág.
Figura 1. Ramas de la Inteligencia Artificial.	18
Figura 2. Aplicaciones de la Inteligencia Artificial Generativa.	20
Figura 3. Precio de modelo GPT-4o mini	39
Figura 4. Precio de modelo GPT-4o mini (continuación)	40
Figura 5. Tabla de límites de cada modelo en Tier 1	41
Figura 6. Mockup de botón en header	43
Figura 7. Mockup de modal	43
Figura 8. Mockup de modal con respuesta	44
Figura 9. Dashboard de OpenAI	45
Figura 10. Sección de facturación	45
Figura 11. Sección de facturación - Agregar método de pago	46
Figura 12. Sección de crear credenciales (API Keys)	46
Figura 13. Documentación de OpenAI	47
Figura 14. Servicio de ChatGPT - Prompt por defecto parte 1	48
Figura 15. Servicio de ChatGPT - Prompt por defecto parte 2	48
Figura 16. Servicio de ChatGPT - Envía solicitudes	49
Figura 17. Servicio de ChatGPT - Instrucciones para preguntas tipo 2	49
Figura 18. Archivo secrets.yml	50
Figura 19. Ruta del controlador	50
Figura 20. Controlador de ChatGPT - Recibe pregunta	51
Figura 21. Controlador de ChatGPT - Procesa respuesta de ChatGPT	51
Figura 22. Controlador de ChatGPT - Llamado de consultas parte 1	52
Figura 23. Controlador de ChatGPT - Llamado de consultas parte 2	53
Figura 24. Interfaz gráfica del botón	54
Figura 25. Interfaz gráfica modal	55
Figura 26. Interfaz gráfica modal con loader	55
Figura 27. Interfaz gráfica con respuesta - parte 1	56
Figura 28. Interfaz gráfica con respuesta - parte 2	57
Figura 29. Consultas SQL con restricción de data - parte 1	58
Figura 30. Consultas SQL con restricción de data - parte 2	58

LISTA DE CUADROS

	Pág.
Cuadro 1. Evolución de los modelos de GPT	25
Cuadro 2. Aspectos relevantes del modelo GPT-4o mini	26
Cuadro 3. Indicadores de eficiencia sin la integración con GPT-4o mini	38
Cuadro 4. Tipos de consultas al asistente virtual	42
Cuadro 5. Protocolo de pruebas funcionales y de rendimiento	59
Cuadro 6. Indicadores de eficiencia con la integración con GPT-4o mini	62

LISTA DE ANEXOS

	Pág.
ANEXO A. Casos de uso	67
ANEXO B. Diagramas	73
ANEXO C. Pruebas	76
ANEXO D. Manual de uso	82
ANEXO E. Manual de uso (video explicativo)	87
ANEXO F. Manual técnico	88
ANEXO G. Manual técnico (video explicativo)	97
ANEXO H. Comparativo de indicadores de eficiencia	98
ANEXO I. Resumen Analítico	101

GLOSARIO

API (Interfaz de Programación de Aplicaciones): Conjunto de reglas y protocolos que permiten que un software se comunique con otro.

Cotización: Documento o información proporcionada por un proveedor que detalla el costo de un bien o servicio.

GPT (Generative Pre-trained Transformer): Es un modelo de inteligencia artificial basado en transformadores, entrenado en grandes cantidades de texto para realizar tareas relacionadas con el lenguaje, como generar texto coherente, responder preguntas y traducir lenguajes.

GPT-4o mini: Versión optimizada y más ligera del modelo GPT-4o, creada para ofrecer un alto rendimiento con menores costos operativos.

HandelBay: Software colombiano para gestión de compras. En este caso es donde se desarrolla el proyecto específico.

Inteligencia Artificial (IA): Rama de la informática que se ocupa del desarrollo de sistemas capaces de realizar tareas que normalmente requieren inteligencia humana, como el reconocimiento de voz, la toma de decisiones y el procesamiento de lenguaje natural.

Mockup: Representación visual de un diseño de interfaz de usuario que simula cómo se verá y funcionará el sistema final. Los mockups son utilizados para planificar la estructura y funcionalidad antes del desarrollo.

Proceso de compra: Conjunto de actividades que realiza una empresa para adquirir bienes o servicios.

Tokens: Unidades mínimas de información utilizadas para medir tanto la entrada como la salida en modelos de inteligencia artificial. Un token puede representar una palabra corta o parte de una palabra más larga.

SQL (Structured Query Language): Lenguaje de programación utilizado para gestionar y manipular bases de datos relacionales.

RESUMEN

En este proyecto se propone el diseño y desarrollo de una integración entre el modelo de inteligencia artificial GPT-4o mini y el software de compras colombiano HandelBay, con el objetivo de optimizar las tareas de sus usuarios, permitiéndoles interactuar con una interfaz para realizar preguntas comunes sobre su gestión de compras y proveedores, y obteniendo respuestas precisas y rápidas a través de consultas a la base de datos del sistema.

La investigación busca mejorar la eficiencia en las operaciones del área de compras, minimizando el tiempo y el esfuerzo requeridos para analizar múltiples cotizaciones y seleccionar proveedores. Además, se espera que la automatización de estos procesos reduzca los costos asociados a la contratación de analistas y practicantes para realizar estas tareas.

Asimismo, este desarrollo pretende establecer un precedente en el uso de tecnologías avanzadas como la inteligencia artificial en el ámbito empresarial colombiano, promoviendo la adopción de soluciones tecnológicas que optimicen los procesos internos y mejoren la toma de decisiones. Aunque esta integración beneficia inicialmente a la empresa específica que utilice HandelBay, la expectativa a futuro es que sirva de referencia para otras organizaciones interesadas en la modernización y eficiencia de sus áreas de compras.

PALABRAS CLAVES: Inteligencia Artificial, GPT-4o mini, Gestión de Compras, Automatización, HandelBay.

ABSTRACT

This project proposes the design and development of an integration between the GPT-4o mini artificial intelligence model and the Colombian purchasing software HandelBay, in order to optimize the tasks of its users, allowing them to interact with an interface to ask common questions about your purchasing and supplier management, and obtaining accurate and fast responses through queries to the system's database.

This investigation seeks to improve efficiency in purchasing area operations, minimizing the time and effort required to analyze multiple quotes and select suppliers. Additionally, automation of these processes is expected to reduce the costs associated with hiring analysts and interns to perform these tasks.

Likewise, this development aims to establish a precedent in the use of advanced technologies such as artificial intelligence in the Colombian business environment, promoting the adoption of technological solutions that optimize internal processes and improve decision making. Although this integration initially benefits the specific company that uses HandelBay, the expectation in the future is that it will serve as a reference for other organizations interested in the modernization and efficiency of their purchasing areas.

KEYWORDS: Artificial Intelligence, GPT-4o mini, Purchasing Management, Automation, HandelBay.

INTRODUCCIÓN

En las últimas décadas, el avance de la inteligencia artificial (IA) y otras tecnologías emergentes ha transformado numerosos sectores alrededor del mundo. Desde la automatización de procesos industriales hasta el análisis de datos de forma avanzada, la IA ha demostrado ser una herramienta poderosa para mejorar la eficiencia y la toma de decisiones en diversos ámbitos. Podría decirse que estos desarrollos tecnológicos están redefiniendo la manera en que operan las empresas, así como también están impulsando nuevas formas de innovación y competitividad en el mercado.

Particularmente en Colombia, las organizaciones enfrentan desafíos cada vez más competitivos y dinámicos, lo que les exige una constante optimización de sus procesos internos. Por ejemplo, las áreas de Compras, se encuentran bajo mucha presión para mejorar la eficiencia y reducir sus costos operativos, lo que les lleva a adoptar nuevas tecnologías. La evaluación y selección de proveedores, una tarea crítica para este equipo de trabajo, a menudo se ve enlentecida por ejecutarlas a cabo manualmente, sin estandarizar los procesos y además con tendencia a la subjetividad en la toma de decisiones, conllevando a presentar ineficiencias y por tanto mayores costos.

Ante esta situación, la necesidad de implementar soluciones tecnológicas que automaticen los procesos asociados al área de Compras se hace evidente. La integración de la inteligencia artificial en estas actividades puede ofrecer mejoras significativas, proporcionando análisis más precisos y rápidos, y facilitando decisiones más objetivas al estar basadas en datos. Todo esto con el fin de contribuir a la reducción de costos, tiempos, operatividad y finalmente mejorar la eficiencia en la gestión de compras y proveedores.

El presente proyecto de grado propone la integración del modelo de inteligencia artificial GPT-4o mini con el software de compras colombiano llamado HandelBay. Esta iniciativa tiene como objetivo facilitar a los usuarios del área de compras una herramienta para interactuar con GPT-4o mini y realizar consultas comunes, obtener respuestas precisas y rápidas basadas en la información almacenada en la base de datos de HandelBay para su selección de proveedores, gestión de compras, generación de informes y otras actividades inherentes al área, ofreciendo una solución tecnológica avanzada y efectiva. cuya relevancia radica en su potencial para transformar un área crítica de la operación empresarial,

demostrando el impacto positivo de la IA en la optimización y agilidad de ciertos procesos. Asimismo, este proyecto refuerza el compromiso de la Facultad de Ciencias e Ingeniería con la innovación y el desarrollo tecnológico, formando profesionales capaces de enfrentar los desafíos actuales y futuros del entorno empresarial.

1. ÁREA PROBLEMÁTICA

1.1 DESCRIPCIÓN

En la era digital actual, el rápido avance de la inteligencia artificial y otras tecnologías innovadoras ha generado un cambio paradigmático en la manera en que las organizaciones abordan sus procesos internos. Sin embargo, es crucial que estas tecnologías estén accesibles y adaptadas para ser aprovechadas por todas las personas, independientemente del tamaño de la empresa o su sector de operación.

La democratización de la tecnología juega un papel crucial en la búsqueda de la eficiencia y la competitividad empresarial, permitiendo a las organizaciones, incluso en economías emergentes, optimizar sus operaciones y tomar decisiones informadas basadas en datos precisos y análisis avanzados.

1.2 DELIMITACIÓN

Particularmente en las áreas de Compra de las empresas colombianas, el proceso de evaluación y selección de proveedores es una tarea crítica que implica la comparación de múltiples cotizaciones para asegurar la adquisición de productos y servicios al mejor costo y calidad. Sin embargo, este proceso presenta varios desafíos, como la subjetividad en la toma de decisiones y la falta de estandarización, lo que puede llevar a inconsistencias y errores en la selección de cada proveedor, afectando así los costos y la eficiencia operativa.

El volumen de trabajo relacionado con la gestión de compras puede ser abrumador para los analistas del área, especialmente en organizaciones con un alto volumen de adquisiciones y proveedores. Esta carga de trabajo puede implicar un retraso significativo en las operaciones, ya que naturalmente no se logra ser eficiente, conllevando a elevar la inversión en talento humano por parte de la empresa debido a la necesidad de contratar personal adicional.

Asimismo, otro desafío importante es la falta de herramientas adecuadas para monitorear y evaluar el desempeño de los proveedores de manera continua, lo que dificulta la implementación de mejoras en el proceso de selección. Paralelamente, la falta de transparencia y control en el proceso de selección de proveedores puede abrir la puerta a prácticas deshonestas, como el fraude o el

favoritismo, comprometiendo la integridad del proceso de compra y de la compañía en cuestión.

1.3 FORMULACIÓN

En este contexto, surge la necesidad de una solución digital que optimice los procesos de selección y compra a proveedores, optimizando el sistema y mejorando la eficiencia, factor crítico en las áreas de compra de las empresas colombianas. ¿Cómo integrar un modelo de inteligencia artificial con un software de compras para optimizar las actividades de sus usuarios en términos de eficiencia operativa?

2. OBJETIVOS

2.1 OBJETIVO GENERAL

Integrar el modelo de inteligencia artificial GPT-4o mini con el software HandelBay para mejorar la eficiencia operativa de sus usuarios en la gestión de compras.

2.2 OBJETIVOS ESPECÍFICOS

- Identificar y analizar la necesidad de optimización del sistema HandelBay, en términos computacionales y de la eficiencia que ofrece a sus usuarios, así como el modelo de GPT más adecuado para la integración.
- Diseñar una solución que permita la integración fluida entre el modelo de GPT y el software de compras HandelBay.
- Desarrollar la integración de la API de GPT con la base de datos del software de compras HandelBay, configurando el modelo, construyendo una interfaz intuitiva, y validando los roles y permisos respectivamente.
- Ejecutar un protocolo de pruebas del sistema integrado para verificar su eficiencia y utilidad para los usuarios.
- Crear documentación del aplicativo que facilite a todos los usuarios el manejo de esta funcionalidad con las mejores prácticas recomendadas.

3. JUSTIFICACIÓN

3.1 NOVEDAD

Al utilizar una inteligencia artificial avanzada, las empresas pueden automatizar gran parte del proceso de análisis y comparación de cotizaciones, asegurando una selección de proveedores más objetiva y basada en datos precisos. La integración de GPT 4o con el software de compras colombiano HandelBay propone una solución innovadora para estos problemas, reduciendo el tiempo invertido en tareas hechas manualmente y disminuyendo la subjetividad en las decisiones de los usuarios.

3.2 INTERÉS

La implementación de esta solución ofrece la estandarización de una buena parte del proceso de selección de proveedores, mejorando la consistencia y rapidez en las decisiones de compra, dado que la capacidad de GPT-4o mini para manejar grandes volúmenes de datos y realizar análisis complejos en tiempo real proporciona una ventaja significativa, permitiendo a los analistas del área enfocarse en tareas de mayor valor estratégico.

Asimismo, la reducción de costos operativos es otro beneficio clave de esta integración. Al disminuir la carga de trabajo manual y minimizar los errores, las empresas pueden optimizar sus recursos humanos y financieros. Mejorar la precisión y rapidez del proceso de compras contribuye a evitar costos adicionales derivados de decisiones arriesgadas, como la adquisición de productos de menor calidad o a precios más altos.

3.3 UTILIDAD

Actualmente, la capacidad de tomar decisiones rápidas y basadas en datos es crucial para que los usuarios del software HandelBay mantengan una ventaja competitiva, por lo que esta moderna solución digital busca ofrecer las herramientas necesarias para lograrlo. En definitiva, la propuesta de desarrollo de esta integración busca beneficiar a las empresas colombianas en términos de eficiencia, así como también adaptarse mejor a un entorno empresarial cada vez más competitivo y dinámico.

4. MARCO TEÓRICO

4.1 INTELIGENCIA ARTIFICIAL (IA)

La inteligencia es un concepto multifacético e intrigante que ha desafiado durante mucho tiempo a psicólogos, filósofos y científicos de la computación.

En palabras de Bubeck et. al. (2023):

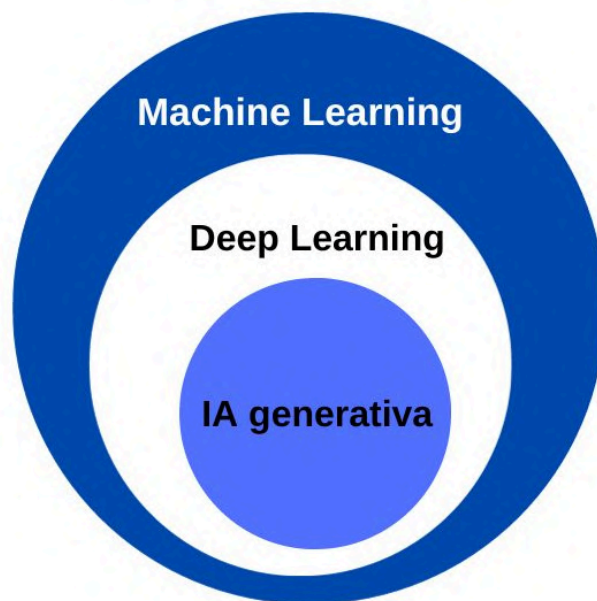
No existe una definición de inteligencia generalmente aceptada, pero hay un aspecto que es ampliamente aceptado. Se acepta que la inteligencia no se limita a un dominio o tarea específica, sino que abarca una amplia gama de habilidades y capacidades cognitivas. (pág. 4).

Específicamente, la inteligencia artificial (IA) es una rama de la informática que se ocupa de la creación de sistemas capaces de realizar tareas que normalmente requieren inteligencia humana. Estas tareas incluyen el reconocimiento de voz, el procesamiento del lenguaje natural, la toma de decisiones y la resolución de problemas.

En la actualidad, la IA comprende tres ramas destacadas conocidas como Machine Learning, Deep Learning y la Inteligencia Artificial Generativa, las cuales poseen características propias y están influenciando la forma en como vivimos en muchos sentidos.

Figura 1. Ramas de la Inteligencia Artificial

INTELIGENCIA ARTIFICIAL



smartmind

Fuente: Smartmind. (s. f.). *Descubriendo las ramas de la inteligencia artificial*. Smartmind. <https://www.smartmind.net/blog/ramas-ia/>

Particularmente el aprendizaje automático o Machine Learning se enfoca en la creación de algoritmos y modelos capaces de aprender y mejorar en tareas específicas usando datos y experiencias. Sin la necesidad de programar directamente una tarea en un sistema, “Machine Learning permite que las máquinas identifiquen patrones y tomen decisiones basadas en datos. La clave de esta tecnología es su habilidad para perfeccionarse y adaptarse conforme recibe más datos y experiencias.” (Smartmind, s. f.)

Por su parte, el Deep Learning es un subcampo específico dentro del Machine Learning cuya función es imitar la complejidad del cerebro humano para procesar datos complejos, imágenes y texto. Esta técnica se basa en redes neuronales artificiales para aprender y encontrar patrones de forma automática y profunda.

4.1.1 Inteligencia Artificial Generativa. “La IA generativa es un subconjunto del aprendizaje profundo o Deep Learning, lo que significa que utiliza redes neuronales artificiales y puede procesar datos etiquetados y no etiquetados usando métodos supervisados y no supervisados.” Smartmind (s. f.)

La integración de tecnologías avanzadas como la inteligencia artificial (IA) se está implementando en diversas áreas para una amplia gama de usos. Desde la medicina y la educación hasta las finanzas y la manufactura, la IA está revolucionando la manera en que se abordan los procesos, mejorando la eficiencia, la precisión y la capacidad de tomar decisiones informadas. Esta tendencia global ha estado reflejando la capacidad de la IA para analizar grandes volúmenes de datos, identificar patrones complejos y ofrecer soluciones innovadoras a problemas tradicionales. “El rápido avance de la inteligencia artificial (IA) y procesamiento del lenguaje natural (PNL) ha llevado al desarrollo de modelos de lenguaje cada vez más sofisticados y versátiles.” (Ray, 2023, pág 1).

Figura 2 .Aplicaciones de la Inteligencia Artificial Generativa



Fuente: Smartmind. (s. f.). *Descubriendo las ramas de la inteligencia artificial*. Smartmind. <https://www.smartmind.net/blog/ramas-ia/>

En el ámbito empresarial, especialmente en las áreas de compras, la adopción de la IA está generando grandes expectativas entre sus integrantes. La gestión de compras empresariales implica numerosas tareas para asegurar la correcta adquisición de bienes y servicios en las organizaciones, entre ellas la oportuna evaluación y selección de proveedores, y la adecuada comparación de cotizaciones y análisis de datos históricos, tareas que pueden ser optimizadas significativamente con el uso de IA.

Cada vez más se está comprobando que la implementación de tecnologías innovadoras con la IA permite automatizar procesos repetitivos, reducir errores humanos y mejorar la objetividad en la toma de decisiones, aportando así beneficios tangibles en relación a la eficiencia operativa y el ahorro de costos de las empresas. El auge de esta tecnología está teniendo un impacto significativo en

varios sectores empresariales como servicio al cliente, comercio electrónico, salud, finanzas y marketing. (Arman & Lamiyar, 2023).

La importancia de este proyecto reside en su potencial para mostrar cómo el uso de tecnologías avanzadas puede revolucionar aspectos clave dentro de la operación empresarial. Al ofrecer una solución novedosa y eficiente, logra integrar de manera efectiva los principios teóricos con su aplicación práctica, específicamente en el ámbito de la ingeniería de sistemas. Además, este enfoque permite no solo mejorar procesos empresariales críticos, sino también servir como un modelo para futuros desarrollos tecnológicos que busquen optimizar la eficiencia y competitividad en las organizaciones.

4.1.2 Modelos de Lenguaje Grande (LLM). Tal como señala Sallam (2023) en su proyecto de investigación, “La inteligencia artificial (IA) puede definirse como el enfoque multidisciplinario de la informática y la lingüística que aspira a crear máquinas capaces de realizar tareas que normalmente requieren inteligencia humana”. (pág. 1).

En otras palabras, la IA puede abarcar una gran variedad de capacidades, desde aprender, adaptarse, racionalizar y comprender conceptos abstractos, así como la reactividad ante atributos humanos complejos como la atención, las emociones, la creatividad, entre otros.

Además, la IA que comprende varias disciplinas como el aprendizaje automático, el aprendizaje profundo y los sistemas de recomendación, ofrece con cada una de sus áreas una fuerte capacidad a las máquinas para aprender y adaptarse a partir de datos.

Bryant (2023), ha afirmado en su investigación:

Recientemente, el desarrollo de modelos de lenguajes grandes (LLM) de inteligencia artificial generativa (IA) ha abierto nuevas vías para la validación de teorías e ideas científicas que entran en conflicto con teorías aceptadas y convicciones profundamente arraigadas. GPT-4, un modelo de lenguaje muy avanzado desarrollado por OpenAI, es un ejemplo de ello. (pág 1).

El aprendizaje automático, una de las áreas más prominentes en las investigaciones de IA, se basa en el desarrollo de algoritmos que permiten a las

computadoras aprender de los datos y hacer predicciones o tomar decisiones sin ser explícitamente programadas para cada tarea específica. El aprendizaje profundo, una subdisciplina del aprendizaje automático, utiliza redes neuronales artificiales con múltiples capas para modelar y entender patrones complejos en grandes conjuntos de datos. Estos avances han llevado al desarrollo de modelos como GPT-4o mini, que pueden procesar y generar lenguaje humano con un alto grado de precisión.

4.1.3 Procesamiento del Lenguaje Natural (PLN). “Es una área multidisciplinaria (de ciencias de la computación, inteligencia artificial y lingüística aplicada) que se encarga de estudiar las interacciones que existen entre personas y computadoras mediante algún lenguaje humano como español, inglés, chino, etc.” (Hubspot, 2023). Su objetivo es ayudar a las máquinas a comprender, interpretar y generar un lenguaje útil para sus usuarios.

El PLN suele utilizarse para extraer información y conocimientos a partir de grandes cantidades de texto. Sus aplicaciones van desde la traducción automática, el análisis de sentimientos, los chatbots y los asistentes virtuales; además, cabe resaltar que en la búsqueda por mejorar cada vez más la experiencia del usuario, los proyectos de PLN se esfuerzan en ofrecer interfaces naturales e intuitivas que soporten sus funcionalidades.

Asimismo, los grandes modelos lingüísticos (LLM) partieron de la base de los PLN como modelos avanzados de IA cuyo entrenamiento requiere enormes cantidades de datos de texto para comprender y generar lenguaje humano. Estos modelos pueden interpretar, predecir y crear textos muy similares al lenguaje humano empleando técnicas de aprendizaje profundo; además cuentan con la capacidad de ejecutar diversas tareas desde responder preguntas, redactar textos, traducir el lenguaje y mucho más.

Según la investigación de Nori et. al. (2023), expresaron:

“También ha habido un gran interés en la capacidad de los LLM para hacer inferencias útiles para una amplia gama de tareas especializadas sin necesidad de realizar ajustes específicos. El desempeño de los LLM de propósito general que utilizan pocas o incluso ninguna indicación resalta su potencial para ayudar con tareas en todos los tipos de problemas, áreas de especialidad y disciplinas.” (pág. 2).

La implementación de PLN en el contexto empresarial mejora potencialmente la comunicación entre los sistemas y los usuarios, y también facilita la automatización de tareas repetitivas y la extracción de información valiosa a partir de grandes volúmenes de datos. Esto es particularmente beneficioso en áreas como las compras, donde la capacidad de analizar rápidamente múltiples cotizaciones y seleccionar proveedores óptimos puede significar una ventaja competitiva significativa.

Particularmente GPT-4, el modelo de IA desarrollado por OpenAI, es un ejemplo avanzado de PLN, ya que utiliza una arquitectura de transformadores para procesar grandes cantidades de texto y generar respuestas coherentes y contextualmente relevantes. Siguiendo a Arman & Lamiyar (2023), “los beneficios potenciales de ChatGPT AI para las empresas la convierten en una tecnología atractiva para explorar y adoptar.” (pág. 65).

4.1.4 GPT (Generative Pre-trained Transformer). OpenAI, fundada en 2015 por Elon Musk, Sam Altman y otros, ha sido pionera en el desarrollo de inteligencia artificial general (AGI) con el objetivo de beneficiar a la humanidad. Como una organización dedicada a la investigación y desarrollo de la inteligencia artificial, OpenAI ha trabajado en la mejora continua de sus modelos a lo largo de los años, logrando con GPT-4, su último lanzamiento a la fecha de este proyecto, un notable avance en la calidad de las respuestas generadas.

El modelo de lenguaje más popular desarrollado por OpenAI es conocido como Chat GPT. Durante su historia, OpenAI ha lanzado una serie de modelos innovadores que han transformado el campo de la inteligencia artificial, como GPT-2 y GPT-3. Estos modelos son capaces de generar texto coherente y comprensible a partir de entradas en lenguaje natural, lo que ha abierto nuevas posibilidades en áreas como la automatización de tareas, la generación de contenido y el procesamiento del lenguaje. Con el desarrollo de ChatGPT-4, OpenAI ha logrado un avance significativo al mejorar las capacidades de conversación, haciendo que el modelo sea más eficiente en el entendimiento del contexto y en la generación de respuestas más coherentes. (Ray, 2023).

El modelo ChatGPT-4 se basa en la arquitectura GPT-4, la cual introdujo mejoras clave respecto a sus predecesores. Este modelo ha sido diseñado para ofrecer una mejor comprensión contextual, lo que le permite generar respuestas más relevantes y precisas en conversaciones más complejas. Además, ChatGPT-4 se

destaca por su capacidad de mantener la coherencia en conversaciones largas, algo que GPT-3 presentaba con ciertas limitaciones. (OpenIA, s.f.).

Con su enfoque en la investigación aplicada y el impacto social positivo, la organización ha demostrado cómo la IA puede ser una herramienta poderosa para resolver problemas complejos y optimizar procesos en diversos campos, lo que refuerza su compromiso con la creación de tecnología accesible y útil. Su constante evolución ha permitido que GPT-4 no solo sea más eficiente para tareas de procesamiento de texto, sino también más adaptable a las interacciones con los usuarios en tiempo real.

La combinación de la generalidad de las capacidades de GPT-4 con numerosas habilidades que abarcan una amplia gama de dominios y su desempeño en un amplio espectro de tareas a nivel humano o más allá, nos hace sentir cómodos al decir que GPT-4 es un paso importante hacia la inteligencia artificial general. (Bubeck et. al., 2023, pág 4).

OpenIA* (2023) señalan las expectativas generadas antes de lanzar este modelo de IA: “esta primera versión de GPT-4 demuestra capacidades notables en una variedad de dominios y tareas, incluida la abstracción, la comprensión, la visión, la codificación, las matemáticas, la medicina, el derecho, la comprensión de los motivos y emociones humanos y más.” (pág 4).

Según el artículo publicado por OpenAI (s.f.) sobre el lanzamiento de GPT-4o mini en mayo de 2024, se trata de su último paso para ampliar los límites del aprendizaje profundo y supervisado en relación con la usabilidad práctica que suelen proponer en sus productos digitales:

Dedicamos mucho esfuerzo durante los últimos dos años a trabajar en mejoras de eficiencia en cada capa de la pila. Como primer fruto de esta investigación, podemos hacer que un modelo de nivel GPT-4 esté disponible de manera mucho más amplia. (pág 1).

En efecto, y siguiendo el proyecto desarrollado por Bryant (2023):

Los avances en los modelos de lenguajes grandes, ejemplificados por GPT-4, han posicionado a la IA como un colaborador dinámico, catalizando un cambio transformador dentro de la investigación académica. Este cambio de paradigma genera un nuevo marco conceptual para capturar las funciones diversificadas de la IA. Este marco bidimensional gira en torno al nivel de interacción y la naturaleza de la tarea. (pág. 2).

La evolución de los modelos de GPT desarrollados por OpenAI, pueden analizarse en el siguiente cuadro:

Cuadro 1. Evolución de los modelos de GPT

Modelo	Fecha de lanzamiento	Características principales	Valores aproximados	Diferencias clave
GPT-1	Junio 2018	- Primer modelo basado en el enfoque de transformadores. - Utiliza 117M de parámetros.	No disponible públicamente	- Introdujo el enfoque de transformadores en modelos de lenguaje. - Limitada capacidad de comprensión y generación de texto.
GPT-2	Febrero 2019	- Escalado a 1.5B de parámetros. - Capaz de generar texto con mayor coherencia	No comercializado	- Mayor capacidad de generación de texto coherente. - Introdujo preocupaciones sobre mal uso de IA, lo que limitó su acceso.
GPT-3	Junio 2020	- Utiliza 175B de parámetros. - Gran capacidad para generar texto natural y realizar tareas de lenguaje complejo.	API desde \$100 al mes, según uso	- Mejora significativa en comprensión y generación de texto. - Amplia capacidad para realizar tareas de lenguaje natural.
GPT-4	Marzo 2023	- Basado en un modelo multimodal con mejoras en precisión y comprensión contextual. - Mejor rendimiento conversacional.	Planes de API desde \$20 al mes en ChatGPT Plus (marzo 2024)	- Aumenta la coherencia en conversaciones largas. - Mayor capacidad para entender y generar respuestas precisas en diferentes contextos.

Fuente: OpenAI. (s.f.). *Documentación y tutoriales en la página oficial de OpenAI* <https://www.openai.com/>

Ahora bien, GPT-4o mini es la evolución más reciente de los modelos GPT, desarrollado por OpenAI para mejorar el rendimiento en velocidad y capacidades en comparación con su predecesor GPT-4. Este modelo no solo mantiene el nivel

de inteligencia del GPT-4, sino que ofrece un procesamiento más rápido y mejoras significativas en tareas relacionadas con texto, voz y visión. Una de las principales innovaciones de su versión más reciente, GPT-4o mini, es su capacidad para entender y discutir imágenes de manera más eficaz, lo que permite aplicaciones como la traducción de menús en tiempo real o la explicación de elementos visuales, ampliando así el rango de interacción con los usuarios. (OpenAI, s.f.).

En el siguiente cuadro se pueden observar algunos detalles relevantes de las recomendaciones y uso de GPT-4o mini, en relación a su desarrollo e innovación:

Cuadro 2. Aspectos relevantes del modelo GPT-4o mini

Aspecto	Descripción
Modelo	GPT-4o mini
Funcionalidades principales	Procesamiento de lenguaje natural avanzado, respuesta a consultas complejas, generación de texto con lógica de negocio.
Integración con API	Se utiliza la API de OpenAI para realizar solicitudes, utilizando gemas como httparty para facilitar la integración.
Capacidad de respuesta	Soporta respuestas categorizadas para preguntas de procesos, cotizaciones, requisiciones, proveedores.
Tiempos de respuesta	Depende de la carga del sistema, pero optimizado para alto rendimiento con tiempos de respuesta adecuados.
Control de permisos y roles	Implementación de restricciones de acceso según la compañía y roles asignados.

Fuente: OpenAI. (s.f.). *Documentación y tutoriales en la página oficial de OpenAI*
<https://www.openai.com/>

A diferencia de GPT-4, GPT-4o mini integra mejoras en la interacción conversacional en tiempo real, permitiendo la integración de modos de voz y video, lo cual promete llevar la IA a un nivel más natural y fluido. Además, su capacidad para procesar datos a mayor velocidad lo hace adecuado para aplicaciones en tiempo crítico, como la explicación de eventos deportivos en vivo o el análisis de datos complejos en tiempo real. Esta versatilidad lo convierte en una herramienta poderosa para una amplia variedad de sectores, desde la educación hasta los negocios y la tecnología.

Para desarrolladores y empresas, GPT-4o mini también ofrece opciones de ajuste fino (fine-tuning), permitiendo personalizar el modelo para tareas específicas con un costo asociado por el uso de tokens. Esta característica ha demostrado ser exitosa en aplicaciones empresariales como la generación de código y la automatización de procesos, donde se ha destacado por su precisión y capacidad de adaptación

Ahora bien, la capacidad de la versión GPT-4o mini para comprender y producir lenguaje natural lo hace ideal para aplicaciones en las que la interacción con el usuario es clave, como en la integración con sistemas de compras empresariales. Por ejemplo, al responder a consultas de usuarios sobre la gestión de compras y proveedores, GPT-4o mini puede proporcionar información precisa y rápida basada en los datos disponibles en la base de datos del software HandelBay.

4.2 SISTEMAS DE GESTIÓN DE COMPRAS

Los sistemas de gestión de compras son plataformas tecnológicas diseñadas para facilitar y optimizar el proceso de adquisición de bienes y servicios en una organización. Estos sistemas abarcan una variedad de funcionalidades, incluyendo la solicitud de cotizaciones, la evaluación de proveedores, la gestión de contratos y la supervisión de órdenes de compra. Al centralizar y automatizar estas tareas, los sistemas de gestión de compras ayudan a las empresas a mejorar la eficiencia, reducir costos y asegurar la transparencia en sus operaciones.

HandelBay es un ejemplo de software de gestión de compras que permite a las empresas manejar sus procesos de adquisición de manera más efectiva. (HandelBay, s. f.). Justamente este desarrollo hecho en Colombia, está orientado a facilitar las compras empresariales y relación con proveedores en empresas de diversos sectores. Entre sus principales características, HandelBay permite la automatización de procesos de compra, generación de reportes en tiempo real y un sistema de control de calidad en las adquisiciones. Su promesa es el enfoque en eficiencia y precisión para cada transacción, ayudando a las empresas a ejecutar las mejores prácticas en Compras, optimizando sus recursos y facilitando un flujo de trabajo ordenado.

Además de integrar herramientas para realizar compras eficientes, también permite gestionar documentos de proveedores de forma centralizada, hacer comparativos de cotizaciones en línea, seguimiento de pedidos y evaluar el desempeño de proveedores. La integración de HandelBay con modelos de IA como GPT-4o mini puede llevar estas capacidades al siguiente nivel, permitiendo análisis más rápidos y precisos y mejorando la toma de decisiones.

La incorporación de IA en sistemas de gestión de compras ofrece numerosos beneficios. Por ejemplo, la capacidad de realizar análisis predictivos puede ayudar a las empresas a anticipar tendencias de mercado y ajustar sus estrategias de compra en consecuencia. Además, la automatización de tareas como la evaluación de proveedores y la comparación de cotizaciones reduce la carga de trabajo manual, liberando tiempo para que los profesionales de compras se enfoquen en actividades estratégicas de mayor valor.

4.3 TECNOLOGÍA EN COMPRAS EMPRESARIALES

Gracias a numerosos estudios e interés en los modelos de inteligencia artificial, es notable el avance significativo durante los últimos años tanto en el campo de sistemas y tecnología, como en otras áreas profesionales.

En el sector financiero, la inteligencia artificial de ChatGPT se está implementando para optimizar tanto el servicio al cliente como la detección de fraudes. Los chatbots basados en ChatGPT facilitan a los usuarios realizar tareas bancarias comunes, como verificar los saldos de sus cuentas, transferir fondos o realizar pagos de facturas, todo de manera rápida y eficiente. Esta automatización no solo mejora la experiencia del usuario, sino que también permite a las instituciones financieras ofrecer un servicio más ágil y accesible. (Arman y Lamiyar, 2023, pág. 66)

Además, GPT-4o mini es capaz de procesar grandes volúmenes de datos, lo que le permite identificar patrones sospechosos que podrían indicar actividades fraudulentas. De esta manera, se reduce significativamente el riesgo de pérdidas financieras para las entidades. Otro beneficio clave es su capacidad para ofrecer asesoramiento personalizado en materia de decisiones estratégicas, lo que ayuda a sus usuarios a tomar decisiones financieras más informadas y objetivas.

También en palabras de Arman & Lamiyar (2023):

ChatGPT AI se ha vuelto cada vez más popular en diversas operaciones comerciales, ofreciendo el potencial para mejorar la eficiencia, la precisión y la satisfacción del cliente. Sin embargo, el uso de la IA en las operaciones comerciales plantea varias consideraciones éticas y legales que deben evaluarse para garantizar que la tecnología se utilice de manera responsable y ética. (pág 78).

De esta manera, es posible percibir cómo la IA está siendo incorporada en una amplia gama de sectores industriales con el objetivo de automatizar tareas y optimizar la eficiencia operativa.

En el área de compras empresariales, la IA tiene la capacidad de procesar y analizar grandes cantidades de datos relacionados con proveedores y cotizaciones, lo que le permite identificar patrones y tendencias relevantes. A partir de este análisis, la IA puede ofrecer recomendaciones basadas en datos precisos, facilitando una toma de decisiones más informada. Este enfoque no solo permite un ahorro significativo de tiempo y recursos, sino que también incrementa la precisión y la imparcialidad en la selección de proveedores y en la gestión global del proceso de compras, asegurando una mejora continua en la operación de las empresas.

4.4 REFERENTE NORMATIVO Y LEGAL

En el contexto del desarrollo e implementación de sistemas de inteligencia artificial como GPT-4o mini, es fundamental considerar el marco legal vigente que regula tanto el uso de tecnologías avanzadas como la protección de datos y derechos de los usuarios. Particularmente en Colombia existen varias leyes y regulaciones relevantes que impactan directamente esta investigación, asegurando que su desarrollo y operación cumplan con los estándares legales y éticos establecidos.

Una de las principales normativas a destacar es la Ley 1581 de 2012, conocida como la Ley de Protección de Datos Personales. (Alcaldía de Bogotá, s.f.). Esta ley establece los principios, derechos y procedimientos para el manejo adecuado de datos personales, asegurando que cualquier sistema que maneje información de usuarios lo haga de manera segura y respetuosa de la privacidad. En el

contexto de la integración de GPT-4o mini con HandelBay, es crucial que los datos de proveedores y transacciones sean manejados conforme a esta ley para evitar vulneraciones a la privacidad y confidencialidad de la información.

Además, la Resolución 256 de 2013 de la Superintendencia de Industria y Comercio reglamenta los estándares para la seguridad de la información y la implementación de medidas técnicas y administrativas que deben seguir las empresas para proteger los datos que manejan. (Departamento Administrativo de la Función Pública, s.f.).

Esta resolución es particularmente relevante para el proyecto, ya que la integración de sistemas de IA en procesos empresariales puede involucrar grandes volúmenes de datos sensibles. Asegurarse de que estas medidas de seguridad sean implementadas adecuadamente es esencial para proteger tanto la integridad de los datos como la confianza de los usuarios en el sistema.

En el ámbito de la contratación pública, la Ley 80 de 1993, o Estatuto General de la Contratación de la Administración Pública, es otro marco legal relevante. Esta ley regula los procesos de contratación pública, estableciendo principios de transparencia, economía y responsabilidad que deben ser observados en la selección de proveedores y la adjudicación de contratos. La integración de GPT-4 en el software de compras HandelBay puede facilitar el cumplimiento de estos principios al automatizar y transparentar el proceso de selección de proveedores, garantizando que las decisiones se basen en datos objetivos y verificables. (Departamento Administrativo de la Función Pública, n.d.).

Finalmente, es importante considerar las normativas internacionales y las mejores prácticas en el uso de IA, como las recomendaciones de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) sobre IA. Estas recomendaciones enfatizan la importancia de diseñar y utilizar sistemas de IA de manera responsable, asegurando la equidad, la transparencia y la rendición de cuentas. (Organización para la Cooperación y el Desarrollo Económicos, 2019)

Adoptar estas prácticas en el desarrollo del proyecto no solo garantiza el cumplimiento de estándares globales, sino que también contribuye a la legitimidad y aceptación del sistema por parte de los usuarios y la comunidad en general.

4.5 ANTECEDENTES

4.5.1 Big Data decision support system. En su tesis doctoral, Ma (2023) propone la creación de un sistema de soporte de decisiones basado en big data como proyecto para su tesis doctoral en Filosofía en la Universidad de Colorado. Su contribución va dirigida principalmente hacia el análisis de la inteligencia accionable como el siguiente nivel de análisis de Big Data, donde los datos se utilizan para la toma de decisiones, entendiendo que la cantidad de datos producidos por sensores, medios sociales y digitales e Internet de las cosas aumenta rápidamente año tras año.

4.5.2 Sparks of Artificial General Intelligence: Early experiments with GPT-4. Bubeck et al. (2023) afirman que en su investigación de una versión temprana de GPT-4, cuando todavía estaba en desarrollo activo por OpenAI, el modelo exhibía una inteligencia más general que los modelos de IA anteriores, logrando resolver tareas nuevas y difíciles que abarcan matemáticas, codificación, visión, medicina, derecho, psicología y más, sin necesidad de ninguna indicación especial, mostrando así un desempeño sorprendentemente cercano al nivel humano. Adicionalmente, proponen varias reflexiones sobre las influencias sociales del reciente salto tecnológico y las futuras direcciones de investigación.

4.5.3 Exploring the Implication of ChatGPT AI for Business: Efficiency and Challenges. Arman & Lamiyar (2023), en su investigación sobre el impacto de ChatGPT en varios sectores comerciales para examinar sus beneficios y desafíos, especialmente en sectores de servicio al cliente, comercio electrónico, atención médica, finanzas, marketing y desarrollo, concluyeron que ChatGPT está revolucionando el mundo empresarial y su impacto es significativo, brindando soluciones a diversos desafíos que enfrentan las empresas, otorgando mayor eficiencia, una mejor satisfacción del cliente y una reducción de costos. Asimismo, se revelaron algunas consideraciones éticas y legales en torno al uso de esta herramienta, que incluyen cuestiones relacionadas con la privacidad, la protección de datos y el sesgo. Entre sus recomendaciones, proponen examinar el desarrollo de programas de capacitación al personal de las empresas sobre las habilidades necesarias para administrar y usar ChatGPT.

4.5.4 Assessing GPT-4's Role as a Co-Collaborator in Scientific Research: A Case Study Analyzing Einstein's Special Theory of Relativity. Según Bryant (2023), el objetivo de este estudio ha sido explorar el papel de GPT-4 como

colaborador en la investigación científica, analizando su capacidad para realizar tareas como compilación bibliográfica y resúmenes de texto de forma autónoma. Adicionalmente, destaca modos de colaboración donde GPT-4 y los investigadores co-crean productos compartidos, impulsando la innovación y eficiencia en el ámbito académico.

4.5.5 The Utility of ChatGPT as an Example of Large Language Models in Healthcare Education, Research and Practice: Systematic Review on the Future Perspectives and Potential Limitations. Según la investigación de Sallam (2023), los mayores beneficios del modelo de inteligencia artificial ChatGPT se hallaron dentro del campo de la escritura académica y/o científica, al menos en un 51,7%, encontrando eficiencia y versatilidad en la escritura con texto de alta calidad, lenguaje mejorado, legibilidad y traducción, lo que facilita la equidad en la investigación y revisión bibliográfica acelerada. De igual manera, se hicieron hallazgos interesantes en relación a la capacidad de analizar datos masivos. Esta revisión sistemática evalúa la utilidad del GPT como modelo de lenguaje destacando su potencial, pero también las preocupaciones, teniendo en cuenta valores fundamentales ante la implementación de la herramienta como la responsabilidad, integridad, transparencia y honestidad de la comunidad científica.

4.5.6 Capabilities of GPT-4 on Medical Challenge Problems. De acuerdo con Nori et al. (2023), el objetivo de esta investigación fue evaluar el rendimiento de GPT-4 en preguntas de examen de competencia médica, comparándolo con modelos anteriores y técnicas avanzadas. Los resultados arrojaron que GPT-4 supera tanto a los humanos como a otros modelos en precisión y eficiencia, resaltando su potencial en aplicaciones médicas y educativas.

4.5.7 Using AI-generated suggestions from ChatGPT to optimize clinical decision support. El artículo de Liu et al. (2023), que explora el uso de sugerencias generadas por inteligencia artificial para mejorar la lógica de alertas en sistemas de apoyo a decisiones clínicas, evaluando paralelamente la no inferioridad en comparación con las sugerencias generadas por humanos, arroja hallazgos interesantes en relación al gran potencial que demostró el modelo de lenguaje GPT, así como también sobre la importancia del aprendizaje por refuerzo a partir de la retroalimentación humana que se le ofrece. Este estudio se muestra como referente del presente proyecto, en busca de mostrar cómo se pueden utilizar modelos de IA para optimizar sistemas ya existentes, especialmente

aquellos que involucran una lógica compleja y sirviendo de apoyo para decisiones humanas.

4.5.8 Performance of ChatGPT as an AI-assisted decision support tool in medicine: a proof-of-concept study for interpreting symptoms and management of common cardiac conditions. Harskamp & De Clercq (2024), decidieron probar la capacidad de ChatGPT para abordar preguntas médicas en su investigación; la misma pretende proyectar el futuro de ChatGPT, un modelo de lenguaje avanzado desarrollado por Open AI, como herramienta de apoyo a la toma de decisiones asistida por IA en el área de medicina. Luego de una serie de pruebas exhaustivas a la herramienta, donde la IA respondió correctamente el 74% de las preguntas planteadas, el estudio concluyó que ChatGPT tiene potencial como herramienta de apoyo particularmente para preguntas médicas sencillas y de baja complejidad, señalando que de igual manejar se necesita hacer más investigación para evaluar completamente su potencial y limitaciones.

4.5.9 Computer-Supported Collaborative Decision-Making. Filip et al. (2017), manifiestan en su libro cómo los sistemas informáticos pueden facilitar la toma de decisiones colaborativas en entornos donde varias personas interactúan. Lo que sin duda alguna funciona como referente para la presente investigación, es cómo los autores definen el concepto de “decisión”, siendo este un resultado clave de las actividades humanas conscientes, que apuntan a seleccionar un curso de acción con el fin de alcanzar una meta o conjunto de metas determinadas. Además, resulta interesante el análisis del impacto de las tecnologías de la información en el diseño y uso de sistemas de apoyo a las decisiones de los equipos de trabajo.

4.5.10 ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Ray (2023), señala que ChatGPT ha demostrado ser muy prometedor para mejorar la eficiencia en varias aplicaciones y campos de investigación científica, así como facilitar la colaboración e impulsar la innovación. Específicamente, resalta los avances que aporta este modelo a la IA generativa, como la mejor comprensión contextual, mejor generación de lenguaje y su adaptabilidad a tareas. De igual manera, aunque destaca las contribuciones significativas al avance de la investigación científica y su potencial de seguir transformando dicho campo en el futuro, también hace énfasis en los desafíos y las preocupaciones éticas asociadas con su uso, proponiendo abordarlos de manera responsable.

5. METODOLOGÍA

5.1 TIPO DE TRABAJO

El proyecto se clasifica como una investigación aplicada en desarrollo tecnológico, cuyo objetivo fue desarrollar e implementar una integración entre el modelo de inteligencia artificial GPT-4o mini y el software de HandelBay, para optimizar la consulta de información por parte de los usuarios, agilizando su gestión de compras y selección de proveedores.

Para el desarrollo se tuvo en cuenta la metodología Waterfall o Cascada (European Knowledge Center For Information Technology, 2022) que permite la implementación del proyecto por fases, y también algunas disciplinas como la ingeniería de requerimientos, el desarrollo de software y la inteligencia artificial generativa.

Este proyecto pertenece al grupo de investigación y desarrollo en informática y telecomunicaciones de la Universidad de Manizales.

5.2 PROCEDIMIENTO

El proyecto se realizó en cinco fases, de la siguiente manera:

5.2.1 Fase 1. Revisión de la necesidad. Esta etapa permitió la recolección de toda la información para el levantamiento de los indicadores necesarios que permitieron dar soluciones a los objetivos indicados anteriormente:

- **Actividad 1. Análisis del requerimiento.** Estudio y comprensión de indicadores de eficiencia del sistema operativo actual sobre acciones específicas de los usuarios para su gestión de Compras, con el fin de contrastar versus las mismas acciones realizadas con el apoyo de la IA.
- **Actividad 2. Análisis de GPT como modelo de Inteligencia Artificial.** Investigación exhaustiva para confirmar la selección del modelo de GPT más adecuado para la integración, analizando y comparando sus características y costos.

5.2.2 Fase 2. Diseño y Arquitectura. Creación de la lógica de negocio de la integración entre ambos desarrollos tecnológicos, con el fin de permitir a los usuarios que puedan realizar consultas sobre la base de datos. Abarcó las actividades:

- **Actividad 1. Definición de Casos de Uso.** Se buscó identificar los puntos y procesos mediante los cuales los usuarios interactúan con el desarrollo, construyendo una descripción detallada de los mismos.
- **Actividad 2. Comprensión del Modelo Relacional.** Para profundizar mejor en la interacción del usuario con la nueva funcionalidad, se diseñó un diagrama de flujo que ilustra la relación del usuario con el aplicativo, otro diagrama de flujo del sistema que detalla el funcionamiento interno del mismo, y otro diagrama de procesos que representa acciones clave que el usuario puede realizar dentro del sistema, como por ejemplo, interactuar con un modal.
- **Actividad 3. Definición de consultas y acceso a la base de datos.** Se consolidaron las consultas SQL específicas y las tablas de la base de datos del software HandelBay que serán accesibles para el GPT-4o mini, con el objetivo de asegurar la precisión y relevancia de las respuestas proporcionadas por el modelo de IA.
- **Actividad 4. Diseño de la interfaz de usuario.** Se concretó el diseño de una interfaz intuitiva para interactuar con la API del GPT, a través de un mockup y facilitando la adaptación a la interfaz del software existente y garantizando una interacción eficiente con la integración.

5.2.3 Fase 3. Desarrollo. Configuración del modelo GPT, que abarca la implementación de la API de OpenAI y otros parámetros necesarios para que pueda procesar consultas en lenguaje natural y convertirlas en acciones ejecutables, desarrollo de la interfaz y revisión de roles y permisos. Se llevaron a cabo las siguientes actividades:

- **Actividad 1. Creación de un servicio para GPT-4o mini.** Desarrollo de un servicio para manejar las solicitudes de la API de GPT-4o mini usando el framework Ruby On Rails.

- **Actividad 2. Programación de controladores y rutas.** Se preparó el código que maneja las solicitudes de consultas del usuario, usando el framework Ruby On Rails y su respectivo lenguaje de programación Ruby.
- **Actividad 3. Desarrollo de la interfaz.** Con uso de tecnologías como AngularJS y Bootstrap 3.4 para el desarrollo de la interfaz, asegurando su integración con la API de GPT-4o mini, proporcionando a los usuarios una herramienta funcional y accesible para interactuar con el sistema.
- **Actividad 4. Configuración de permisos y roles.** Implementación de validaciones para configurar los permisos en el backend en base a los roles necesarios, asegurando que cada usuario tenga acceso únicamente a las funciones y datos pertinentes para garantizar la seguridad y adecuada gestión de acceso.

5.2.4. Fase 4. Pruebas. Ejecutar un protocolo de pruebas para asegurar la eficiencia del desarrollo. Comprendió la actividad:

- **Actividad 1. Ejecución de pruebas internas del sistema integrado.** Se buscó asegurar que el sistema cumpla con los requisitos y expectativas antes de su lanzamiento, ejecutando un protocolo de pruebas funcionales de rendimiento y seguridad, así como auditorías de código para asegurar que los permisos y roles estén configurados correctamente y no existan vulnerabilidades.

5.2.5 Fase 5. Documentación. Se reunieron todos los componentes que integran la aplicación para la elaboración de un manual técnico y un manual de usuario.

- **Actividad 1. Creación del manual de usuario:** Esta actividad permitió generar un manual en el cual se describen las funcionalidades que tiene la integración con GPT-4o mini con las mejores prácticas recomendadas para el usuario.
- **Actividad 2. Elaboración del manual técnico:** Con esta actividad se generó la documentación de la parte técnica respecto al funcionamiento de la integración con la plataforma.

6. RESULTADOS

6.1 DESCRIPCIÓN DE RESULTADOS

Se ha desarrollado e implementado con éxito la integración del modelo GPT-4o mini y el software de compras HandelBay. Para lograr este resultado final, se inició analizando el rendimiento del aplicativo actual, entendiendo que se necesitaba una mejora del sistema debido al volumen de datos que manejan las áreas de Compras de sus clientes, y la rapidez con la que deben resolver sus procesos de manejo y adjudicación de proveedores.

Con mayor especificidad y con el fin de establecer métricas de eficiencia y de optimización de tiempo, se procedió a medir tres variables en relación con el uso de la herramienta HandelBay en tres diferentes casos para la gestión de Compras, los cuales pueden consultarse en el Cuadro 3.

Comprendiendo a detalle cada una de las acciones que realizan los usuarios para cumplir sus objetivos en la plataforma, se determinó la cantidad de consultas SQL que el sistema realiza internamente, así como la cantidad de clics que debe hacer el usuario mientras navega, y finalmente el tiempo promedio invertido en cada caso respectivamente.

Es importante destacar que dichos indicadores de eficiencia se obtuvieron mediante pruebas de usuario realizadas en un entorno de pruebas controlado. Para ello, se solicitó a un grupo de usuarios representativos que ejecutaran tareas específicas en la plataforma, registrando el número de clics, tiempo promedio de ejecución y cantidad de consultas SQL realizadas. Durante las pruebas, el equipo de QA (aseguramiento de calidad) y el área de Customer Success (éxito del cliente) observaron las interacciones para identificar posibles obstáculos y recopilar información precisa sobre el desempeño del sistema.

Cuadro 3. Indicadores de eficiencia sin la integración con GPT-4o mini

Objetivo del caso de uso	Acciones para cumplir objetivo	Cantidad de clics	Tiempo promedio	Cantidad de consultas SQL
Consultar información general de un proceso de compra	<ul style="list-style-type: none"> ● Iniciar sesión ● Pestaña “Procesos” ● Pestaña “En Revisión” ● Uso de buscador ● Selección de proceso 	5	180 seg.	4
Verificar estado de respuesta de 3 proveedores de un proceso de compra	<ul style="list-style-type: none"> ● Iniciar sesión ● Pestaña “Procesos” ● Pestaña “En Revisión” ● Selección de proceso ● Pestaña “Invitados” ● Botón “Ver más” 	6	240 seg.	5
Comparar y analizar 3 cotizaciones de un proceso de compra	<ul style="list-style-type: none"> ● Iniciar sesión ● Pestaña “Procesos” ● Pestaña “En Revisión” ● Selección de proceso ● Pestaña “Cotizaciones” ● Botón “Ver Cotización” (Proveedor 1) ● Botón “Volver” ● Botón “Ver Cotización” (Proveedor 2) ● Botón “Volver” ● Botón “Ver Cotización” (Proveedor 3) 	10	420 seg.	9

Fuente: Autores

Seguidamente, para seleccionar el modelo de GPT más adecuado a la plataforma y sus usuarios, se realizó un análisis comparativo de sus distintas versiones, considerando tanto capacidad como costo. Se optó por el modelo GPT-4o mini, descrito en la documentación oficial de OpenAI como una versión más avanzada en comparación con modelos anteriores, y con una eficiencia de costo superior. (Open AI, 2024).

La decisión fue basada en varios factores clave:

- El GPT-4o mini ofrece una alta tasa de procesamiento, lo que resulta idóneo para el proyecto.
- El costo por millón de tokens es de \$0.150 USD para los tokens de entrada y \$0.600 USD para los de salida, lo que lo convierte en una opción económicamente viable (Figura 3 y Figura 4).
- Se eligió el Tier 1, que soporta hasta 200,000 tokens por minuto (TPM) y un máximo de 10,000 consultas diarias, garantizando que las necesidades del desarrollo sean cubiertas de manera eficiente (Figura 5).

Figura 3. Precio de modelo GPT-4o mini

Model	Pricing	Pricing with Batch API*
gpt-4o-mini	\$0.150 / 1M input tokens	\$0.075 / 1M input tokens
	\$0.075 / 1M cached** input tokens	
	\$0.600 / 1M output tokens	\$0.300 / 1M output tokens
gpt-4o-mini-2024-07-18	\$0.150 / 1M input tokens	\$0.075 / 1M input tokens
	\$0.075 / 1M cached** input tokens	
	\$0.600 / 1M output tokens	\$0.300 / 1M output tokens

Fuente: OpenIA. (2024). GPT-4o Mini Pricing. <https://openai.com/api/pricing/>

Figura 4. Precio de modelo GPT-4o mini (continuación)

GPT-4o

GPT-4o is our most advanced multimodal model that's faster and cheaper than GPT-4 Turbo with stronger vision capabilities. The model has 128K context and an October 2023 knowledge cutoff.

[Learn about GPT-4o ↗](#)

Model	Pricing	Pricing with Batch API*
gpt-4o	\$5.00 / 1M input tokens	\$2.50 / 1M input tokens
	\$15.00 / 1M output tokens	\$7.50 / 1M output tokens
gpt-4o-2024-08-06	\$2.50 / 1M input tokens	\$1.25 / 1M input tokens
	\$1.25 / 1M cached** input tokens	
gpt-4o-2024-05-13	\$10.00 / 1M output tokens	\$5.00 / 1M output tokens
	\$5.00 / 1M input tokens	\$2.50 / 1M input tokens
	\$15.00 / 1M output tokens	\$7.50 / 1M output tokens

Fuente: OpenIA. (2024). GPT-4o Mini Pricing. <https://openai.com/api/pricing/>

Figura 5. Tabla de límites de cada modelo en Tier 1

Tier 1 rate limits

This is a high level summary and there are per-model exceptions to these limits (e.g. some legacy models or models with larger context windows have different rate limits). To view the exact rate limits per model for your account, visit the [limits](#) section of your account settings.

MODEL	RPM	RPD	TPM	BATCH QUEUE LIMIT
gpt-4o	500	-	30,000	90,000
gpt-4o-mini	500	10,000	200,000	2,000,000
gpt-4o-realtime-preview	100	100	20,000	-
gpt-4-turbo	500	-	30,000	90,000
gpt-4	500	10,000	10,000	100,000
gpt-3.5-turbo	3,500	10,000	200,000	2,000,000
omni-moderation-*	500	10,000	10,000	-
text-embedding-3-large	3,000	-	1,000,000	3,000,000
text-embedding-3-small	3,000	-	1,000,000	3,000,000
text-embedding-ada-002	3,000	-	1,000,000	3,000,000
omni-moderation-*	500	10,000	10,000	-

Fuente: OpenIA. (2024). GPT-4o Mini Pricing. <https://openai.com/api/pricing/>

Esta decisión asegura que el sistema pueda manejar un volumen considerable de datos con rapidez, manteniendo control sobre los costos operativos.

Una vez analizada esta información y continuando con la segunda fase del proyecto, se generaron los diferentes casos de uso, permitiendo identificar de manera clara la interacción de los usuarios con el desarrollo, en la búsqueda por facilitarles respuestas rápidas de sus procesos de compra así como análisis completos de las cotizaciones recibidas por sus proveedores, lo que puede evidenciarse en el Anexo A.

Asimismo, para cumplir la actividad 2 se elaboraron tres diagramas para la comprensión del flujo de información, en principio uno sobre el modelo entidad-relación, otro que permite visualizar la estructura y arquitectura de software, y finalmente otro que abarca las acciones clave que el usuario puede realizar en el sistema; los mismos pueden ser consultados en el Anexo B. Este

modelo en formato diagrama permite una comprensión más clara de las relaciones entre la data y el usuario, facilitando la interacción (De Miguel Castaño, 1999).

Para la tercera actividad de la Fase 2, se definieron cuatro consultas SQL específicas y estáticas que sirvieron de base para las interacciones entre el usuario y el modelo GPT-4o mini; adicionalmente el sistema puede responder a preguntas abiertas más dinámicas donde implique análisis de la data entregada.

Cuadro 4. Tipos de consultas al asistente virtual

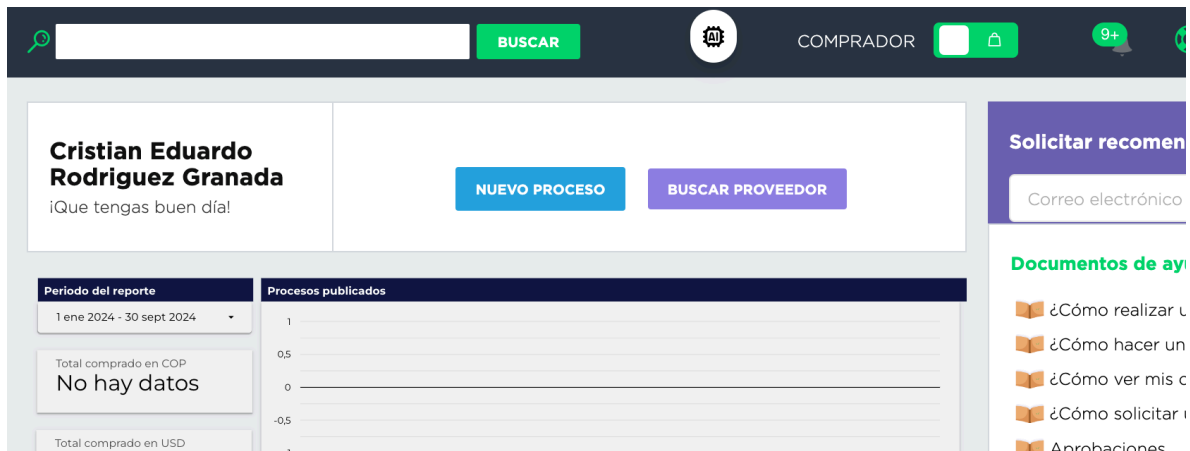
Tipos de consulta	Respuesta
<p>Tipo 1: Preguntas estáticas</p>	Devolver el estado de un proceso de compra o requisición.
	Proporcionar detalles sobre un proceso o requisición.
	Mostrar información actualizada sobre las cotizaciones de un proceso o requisición.
	Listar los proveedores invitados a un proceso de compra y su estado respecto al proceso.
<p>Tipo 2: Preguntas dinámicas</p>	La respuesta es abierta y dinámica, dependiendo de la consulta del usuario, por ejemplo, comparar un grupo de cotizaciones por sus valores y sugerir la mejor oferta.

Fuente: Autores

Estas consultas fueron diseñadas para cubrir los aspectos más relevantes de la funcionalidad del sistema, garantizando que la plataforma proporcione información relevante y precisa a los usuarios, mejorando su experiencia y la eficiencia de sus decisiones.

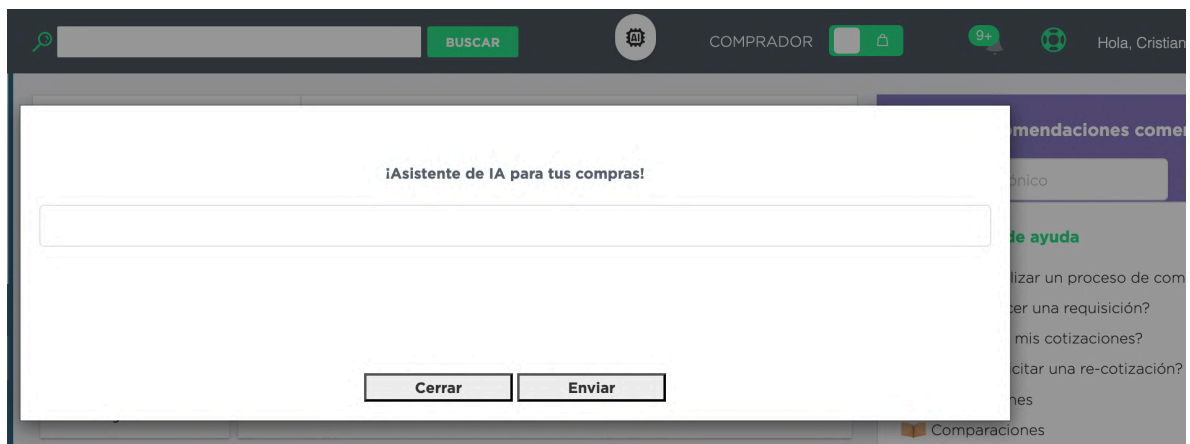
Durante la actividad 3 de la segunda fase, se llevaron a cabo diseños de mockups que representan fielmente la interfaz del sistema, asegurando una experiencia de usuario intuitiva.

Figura 6. Mockup de botón en header



Fuente: Autores

Figura 7. Mockup de modal



Fuente: Autores

Se trabajó en la distribución de los elementos visuales, mostrando cómo se presentarían las respuestas generadas por el modelo GPT-4o mini y la forma en que los usuarios interactuarían con la plataforma. Los mockups fueron alineados con la estética y el diseño de la plataforma existente, manteniendo coherencia con los lineamientos visuales ya establecidos y garantizando una fácil adaptación por parte de los usuarios.

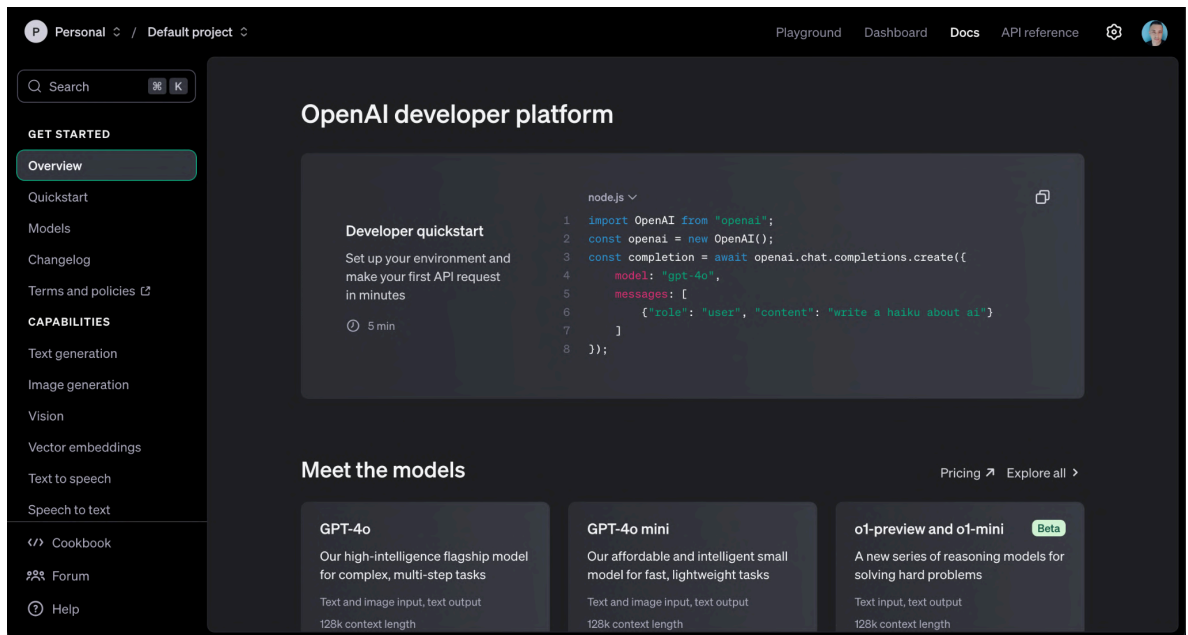
Figura 8. Mockup de modal con respuesta



Fuente: Autores

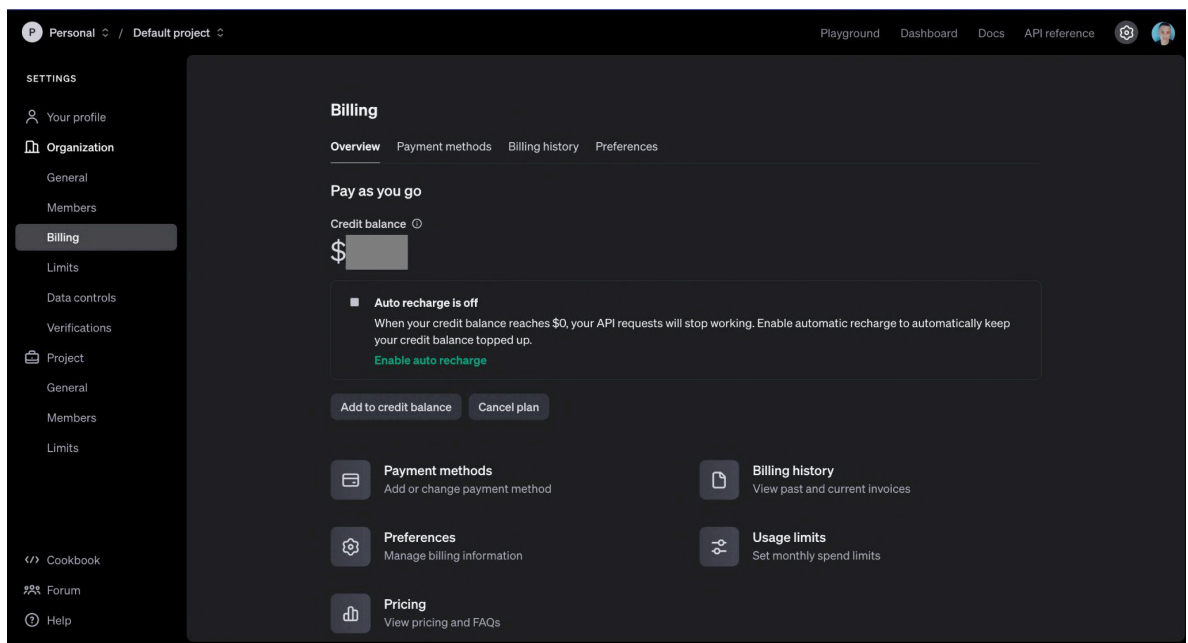
Para la ejecución de la Fase 3, en primer lugar se creó una cuenta en la plataforma de OpenAI a través del sitio <https://platform.openai.com/>, donde se accedió a la sección de Dashboard y API Keys para generar las credenciales necesarias que permitirían la integración con la API de GPT-4o mini. Para garantizar el uso sin restricciones del Tier 1, se añadió un método de pago y se realizó una recarga de \$20 USD a la cuenta, habilitando las características completas del plan.

Figura 9. Dashboard de OpenAI



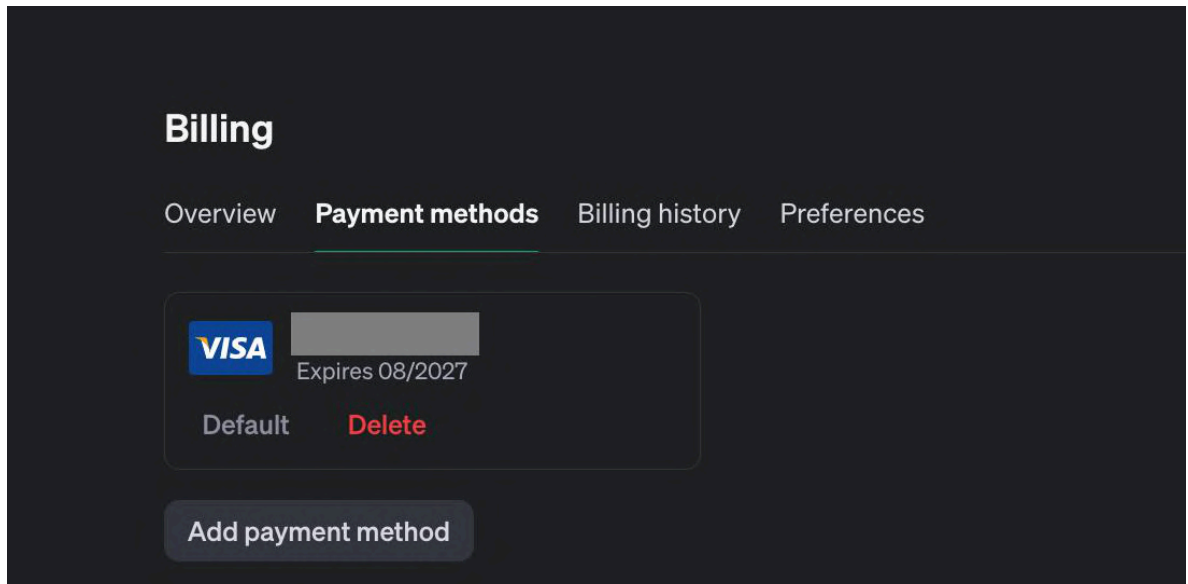
Fuente: OpenAI. (s.f.). OpenAI API documentation. OpenAI.
<https://platform.openai.com/>

Figura 10. Sección de facturación



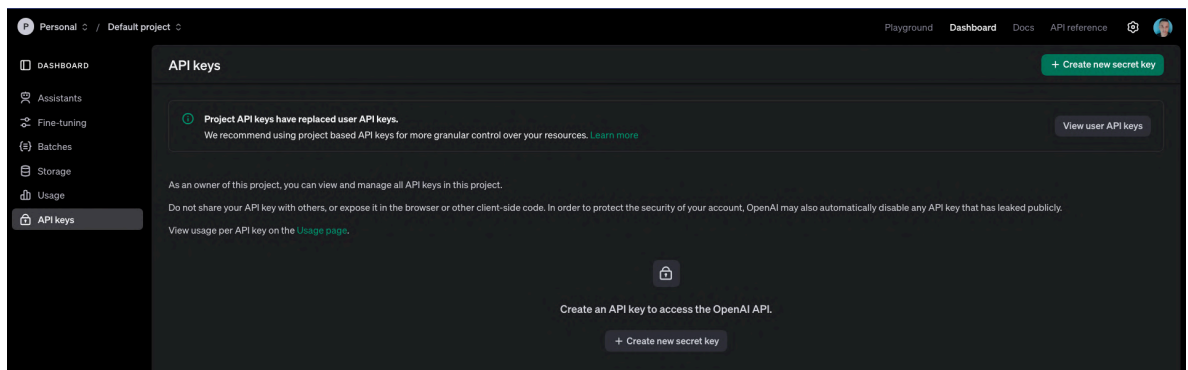
Fuente: OpenAI. (s.f.). OpenAI API documentation. OpenAI.
<https://platform.openai.com/>

Figura 11. Sección de facturación - Agregar método de pago



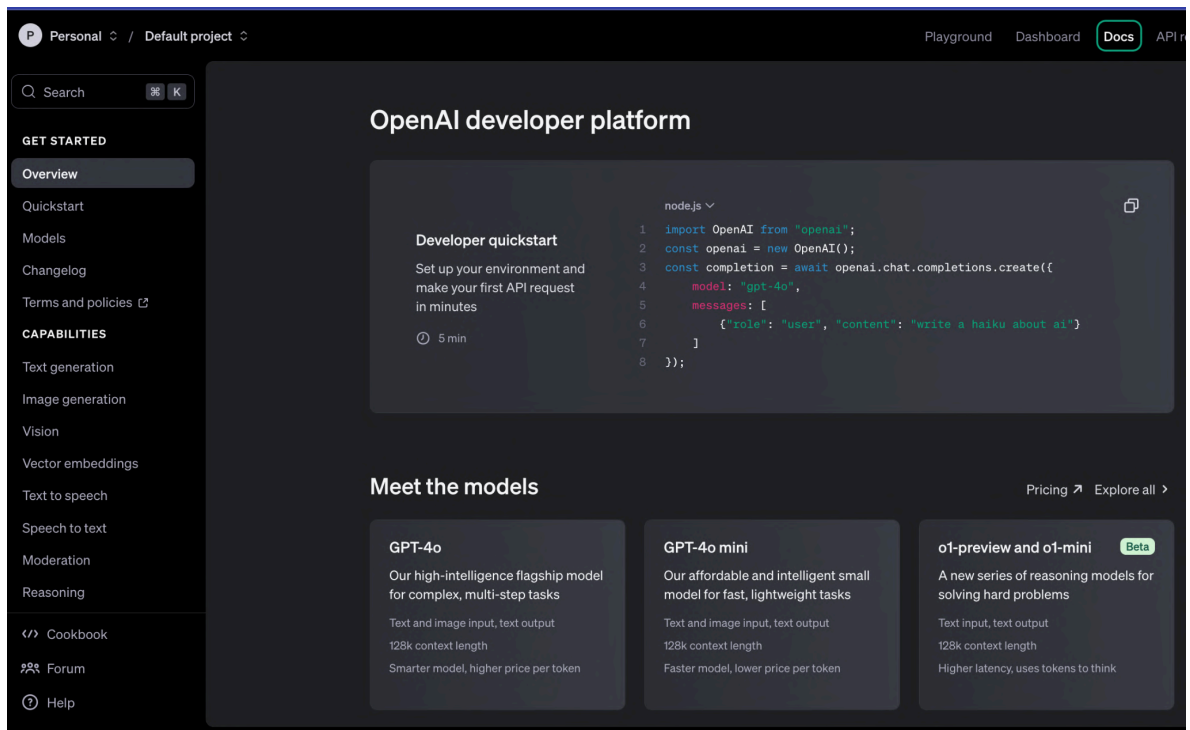
Fuente: OpenAI. (s.f.). OpenAI API documentation. OpenAI.
<https://platform.openai.com/>

Figura 12. Sección de crear credenciales (API Keys)



Fuente: OpenAI. (s.f.). OpenAI API documentation. OpenAI.
<https://platform.openai.com/>

Figura 13. Documentación de OpenAI



Fuente: OpenAI. (s.f.). OpenAI API documentation. OpenAI. <https://platform.openai.com/>

Concluyendo la actividad 1 de la fase 3, se instaló la gema `httparty` para facilitar la integración con la API. Posteriormente, se desarrolló un servicio en el archivo `app/services/chatgpt_service.rb`, cuya función principal es gestionar las solicitudes a la API de GPT-4o mini (Figura 14). Este servicio organiza las preguntas o solicitudes realizadas por los usuarios y las envía a la API para su procesamiento. (Figura 16).

Figura 14. Servicio de ChatGPT - Prompt por defecto parte 1

```
# Iniciar la conversación con el prompt de sistema
def self.initiate_conversation
  # Esto es el prompt del sistema que le dice a ChatGPT cómo responder
  system_prompt = <<<PROMPT
  Eres un asistente virtual para una plataforma B2B. Todas tus respuestas deben seguir estrictamente el siguiente formato JSON:

  {
    "action": "nombre_de_la_accion",
    "data": {
      "query": "valor"
    }
  }

  No incluyas ninguna explicación o texto adicional fuera del objeto JSON. Solo devuelve el JSON según la acción correspondiente.

  Las acciones posibles son:
  - "process_status": Devuelve el estado de un proceso de compra o requisición.
  - "process_info": Devuelve información sobre el proceso o requisición.
  - "process_quotation": Devuelve información sobre las cotizaciones del proceso o requisición.
  - "process_quotation_analysis": Analiza las cotizaciones de un proceso, selecciona la más favorable, ordena las cotizaciones por criterios como el precio o tiempo de entrega, y permite preguntas como "ordena de la más costosa a la más barata", "filtra por menor tiempo de entrega", etc.
  - "process_guests": En esta sección encontrarás todos los proveedores invitados y la información actualizada de su proceso de cotización también llamado tracker de invitados.
  - "error": Esta acción debe usarse cuando la pregunta no corresponde a ninguna de las acciones definidas, o si no se entiende la pregunta. En el campo "query", devuelve un mensaje explicativo dentro de los límites de JSON.

  Si la acción es "process_status", incluye en "data" una clave "query" con el término de búsqueda, para las acciones process_status, process_info, process_quotation y process_guests tiene que ser un query de tipo numerico, sino que sea vacío, ejm:
  "query": "",

  ## Instrucciones para interpretar las consultas de cotización (process_quotation_analysis):
  - Siempre que la pregunta esté relacionada con cotizaciones (precios, tiempo de entrega, análisis), utiliza la acción "process_quotation_analysis".
  - Si la acción es "process_quotation_analysis" debes extraer la intención del usuario. Si la pregunta está relacionada con cotizaciones y análisis, extrae la acción deseada y devuélvela en el campo "user_intent". Ejemplo de intención: "analizar y elegir la cotización más favorable y explicar por qué".
  - Ejemplos de intenciones incluyen: "ordenar cotizaciones por precio descendente", "mostrar la cotización con el menor tiempo de entrega", etc.

  Ejemplo:
  Usuario: "Quiero saber el estado de mi proceso número 12345."
  Respuesta esperada:
  {
    "action": "process_status",
    "data": {
      "query": "12345"
    }
  }

  Ejemplo de error:
  Usuario: "Quiero saber algo que no entiendo bien."
  Respuesta esperada:
  {
    "action": "error",
    "data": {
      "query": "La pregunta no corresponde a ninguna acción válida."
    }
  }
  >>>
end
```

Fuente: Autores

Figura 15. Servicio de ChatGPT - Prompt por defecto parte 2

```
Ejemplo:
Usuario: "Quiero saber cuál es la mejor cotización para el proceso número 12345 y por qué debería elegirla."
Respuesta esperada:
{
  "action": "process_quotation_analysis",
  "data": {
    "query": "12345",
    "user_intent": "analizar y elegir la cotización más favorable y por qué"
  }
}

Asegúrate de devolver siempre el formato exacto de JSON sin añadir texto adicional.
PROMPT

# Iniciamos la conversación con el prompt del sistema
@conversation_messages = [{ role: 'system', content: system_prompt }]
end
```

Fuente: Autores

Dentro del servicio, se creó el método *self.initiate_conversation*, que se encarga de estructurar las consultas del usuario para que el modelo responda en uno de los formatos previamente establecidos, los cuales corresponden a las diferentes respuestas que el sistema puede generar. Las credenciales de la API fueron almacenadas de manera segura en el archivo *config/secrets.yml* (Figura 18) para ser utilizadas en el momento de realizar peticiones a OpenAI.

Figura 16. Servicio de ChatGPT - Envía solicitudes

```
46 # Método que interactúa con la API de OpenAI
47 def self.ask(prompt)
48   # Aseguramos que la conversación haya sido iniciada con el sistema prompt
49   initiate_conversation() if @conversation_messages.nil?
50
51   # Añadir el mensaje del usuario a la conversación
52   @conversation_messages << { role: 'user', content: prompt }
53
54   # Se hace la petición a la API con el prompt recibido
55   response = post('/chat/completions',
56     headers: {
57       'Content-Type' => 'application/json',
58       'Authorization' => "Bearer #{ENV['CHATGPT_API_KEY']}"
59     },
60     body: {
61       model: 'gpt-4o-mini', # Puedes cambiar el modelo según tus necesidades
62       messages: @conversation_messages,
63       max_tokens: 150
64     }.to_json)
65
66   # Guardamos la respuesta de ChatGPT en la conversación
67   @conversation_messages << { role: 'assistant', content: response['choices'][0]['message']['content'] }
68
69   # Devuelve el texto generado por ChatGPT
70   response.parsed_response['choices'][0]['message']['content'] rescue "Lo siento, hubo un problema con la solicitud."
71 end
```

Fuente: Autores

Asimismo, se creó el método encargado de dar las instrucciones a las preguntas de tipo 2 (dinámicas) y posteriormente enviar la petición con los datos (instrucciones, datos de cotizaciones y pregunta hecha por el usuario) para ser procesado y generar una respuesta para el usuario (Figura 17). Este proceso está basado en la disciplina conocida como prompt engineering, la cual permite configurar las respuestas que da un sistema inteligente. (Hubspot, 2024)

Figura 17. Servicio de ChatGPT - Instrucciones para preguntas tipo 2

```
# Método que interactúa con la API para analizar las cotizaciones
def self.analyze_ask(quotations_data, user_intent)
  prompt = <<<PROMPT
  Aquí tienes las cotizaciones para un proceso de compra en formato JSON:

  #{@quotations_data.to_json}

  La intención del usuario es: "#{user_intent}".

  ### Instrucciones para el análisis:
  - Si solo hay una cotización en los datos, devuelve la respuesta con etiquetas <strong> para resaltar puntos importantes y cualquier otra etiqueta para organizar y que se vea mejor la data.
  - Si solo hay una cotización en los datos, no se puede realizar un análisis comparativo. En ese caso, devuelve el siguiente mensaje: "¡opps! Solo hay una cotización disponible.</pp>".
  - Si hay más de una cotización, realiza las siguientes acciones:
  - Analiza esta data teniendo presente lo que quiere hacer el usuario con ella en la intención del usuario.
  - Además, devuelve la respuesta en formato HTML con etiquetas <strong> para resaltar puntos importantes y <ul><li> para listar las ventajas o desventajas de cada cotización y cualquier otra etiqueta para organizar y que se vea mejor la data.
  - Además formatea los precios, ponles el formato correcto teniendo cuidado con las decimales y teniendo presente su moneda.
  - Ten presente, nunca incluyas enlaces en la respuesta y asegura que el formato sea correcto para HTML.
  - Las respuestas dadas siempre en español.
  - No expongas en la respuesta el funcionamiento de la estructura y arquitectura del data.

  PROMPT

  response = post('/chat/completions',
    headers: {
      'Content-Type' => 'application/json',
      'Authorization' => "Bearer #{ENV['CHATGPT_API_KEY']}"
    },
    body: {
      model: 'gpt-4o-mini', # Puedes cambiar el modelo si lo deseas
      messages: [{ role: 'user', content: prompt }],
      max_tokens: 2000
    }.to_json)
  )

  # Devuelve la respuesta generada por ChatGPT
  response.parsed_response['choices'][0]['message']['content'] rescue "Lo siento, hubo un problema con el análisis."
end
```

Fuente: Autores

Figura 18. Archivo secrets.yml

```
1  development:
2    secret_key_base: cdf4f9a1a3sdfsdfs65c41fd9
3  test:
4    secret_key_base: 12823dfsdfsdf3bcf44
5  production:
6    secret_key_base: fa740b5dfsdfsfd
7    CHATGPT_API_KEY: 'sk-proj-PIpdwsdIHG5ksdfsQfvLcuqelG74M8fYJt1G
8
9
```

Fuente: Autores

Asimismo, para la actividad 2 que corresponde a programar controladores y rutas, se crearon los controladores y las rutas necesarias para establecer la conexión entre el servicio encargado de manejar las solicitudes a la API de GPT-4o mini y la vista de usuario. El controlador, ubicado en su correspondiente directorio, se encarga de recibir las consultas que el usuario realiza en la interfaz (Figura 20) y las envía al servicio mencionado en la actividad 1. Este servicio procesa la solicitud, la envía a la API junto con el formato predefinido, y recibe la respuesta que posteriormente es evaluada en el controlador (Figura 21).

Figura 19. Ruta del controlador

```
namespace :landing do
  namespace :api do
    namespace :v1 do
      resources :chatgpt do
        collection do
          post 'chatgpt_ask'
        end
      end
    end
  end
end
```

Fuente: Autores

Figura 20. Controlador de ChatGPT - Recibe pregunta

```
1 class Landing::Api::V1::ChatgptController < ApiController
2
3   # Se encarga de mandar el prompt a ChatGPT y devolver la respuesta
4   def chatgpt_ask
5     response_content = ChatgptService.ask(params[:prompt])
6
7     response_content = process_answer(response_content)
8
9     render json: { response: response_content }, status: :ok
10  end
11
```

Fuente: Autores

Figura 21. Controlador de ChatGPT - Procesa respuesta de ChatGPT

```
19 private
20
21 # Procesa la respuesta del servicio de ChatGPT
22 def process_answer(response_content)
23   parsed_response = parse_json(response_content)
24   return response_content unless parsed_response # Si no se puede parsear el JSON, devolver tal cual
25
26   action = parsed_response['action']
27   data = parsed_response['data']
28   company_id = current_user.company_id
29
30   case action
31   when 'process_status'
32     process_status(data['query'], company_id)
33   when 'process_info'
34     process_info(data['query'], company_id)
35   when 'process_quotation'
36     process_quotation(data['query'], company_id)
37   when 'process_guests'
38     process_guests(data['query'], company_id)
39   when 'process_quotation_analysis'
40     process_quotation_analysis(data['query'], data['user_intent'], company_id)
41   else
42     "<strong>#{data['query']}</strong>"
43   end
44 end
```

Fuente: Autores

Dependiendo de la respuesta que el modelo devuelva, el sistema ejecuta una de las consultas SQL definidas previamente (Figura 22 y Figura 23). Estas consultas permiten devolver información clave sobre el estado de un proceso de compra, detalles de un proceso o requisición, información sobre cotizaciones o detalles actualizados de los proveedores.

Figura 22. Controlador de ChatGPT - Llamado de consultas parte 1

```
50 # Método para manejar el estado de un proceso
51 def process_status(process_id, company_id)
52   # Validaciones
53   return valid_process_id(process_id) if valid_process_id(process_id).present?
54
55   proceso = Req.search_process_company(process_id, company_id)[]
56
57   if proceso
58     "El estado de tu proceso es: " + proceso[:status]
59   else
60     "No se encontró el proceso con ID: " + proceso[:id]
61   end
62 end
63
64 # Método para manejar la información de un proceso
65 def process_info(process_id, company_id)
66   # Validaciones
67   return valid_process_id(process_id) if valid_process_id(process_id).present?
68
69   proceso = Req.process_info(process_id, company_id)
70
71   if proceso.present?
72     "Nombre del proceso: " + proceso[:name] + ", Estado: " + proceso[:status] + ", Fecha de publicación: " + proceso[:date_publication] + ", El proceso cierra: " + proceso[:end_date]
73   else
74     "No se encontró el proceso con ID: " + proceso[:id]
75   end
76 end
77
78 # Método para manejar las cotizaciones de un proceso
79 def process_quotation(process_id, company_id)
80   # Validaciones
81   return valid_process_id(process_id) if valid_process_id(process_id).present?
82
83   quotations = Quotation.search_quotations_of_process(process_id, company_id)
84
85   if quotations.present?
86     response = "El proceso " + proceso[:id] + " tiene las siguientes cotizaciones:"
87     quotations.each do |quot|
88       response += "Compañía: " + quot[:company] + "\n"
89       response += "Estado: " + quot[:status] + "\n"
90       response += "Recotización: " + quot[:recotization] + " (Sí / No)\n"
91       response += "-----\n"
92     end
93     response += "\n"
94   else
95     response = "No se encontró el proceso con ID " + proceso[:id] + " o no tiene cotizaciones"
96   end
97 end
98 end
```

Fuente: Autores

Figura 23. Controlador de ChatGPT - Llamado de consultas parte 2

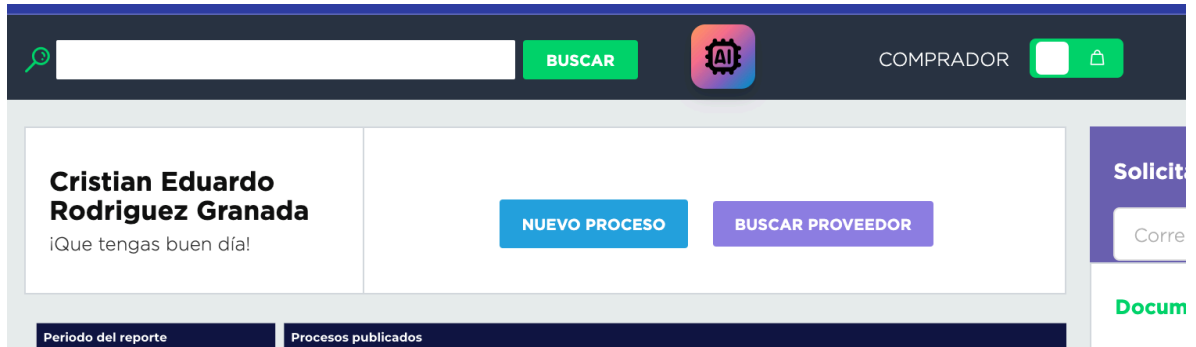
```
102 # Método para manejar la información de los invitados a un proceso
103 def process_guests(process_id, company_id)
104 # Validaciones
105 return valid_process_id(process_id) if valid_process_id(process_id).present?
106
107 guests_info = PurchaseProvider.guests_info_tracker(process_id, company_id)
108
109 if guests_info.present?
110 response = "<strong>Invitados al proceso</strong> ##{process_id}:<br>"
111
112 guests_info.each do |guest|
113 response += "• <strong>Compañía:</strong> #{guest[:user_company]}, "
114 response += "<strong>Nombre:</strong> #{guest[:user_name]}, "
115 response += "<strong>Teléfono:</strong> #{guest[:phone]}, "
116 response += "<strong>Email:</strong> #{guest[:email]}<br>"
117 response += "• <strong>Lectura:</strong> #{guest[:read]}<br>"
118 response += "• <strong>Intención:</strong> #{guest[:intention]}<br>"
119 response += "• <strong>Cotización:</strong> #{guest[:send_quotation]}<br>"
120 response += "<br>"
121 end
122
123 response
124 else
125 response = "<strong>No se encontró información de invitados para el proceso con ID</strong> #{process_id}."
126 end
127 end
128
129 # Se encarga de analizar la data de las cotizaciones para procesar y mostrar respuestas
130 def process_quotation_analysis(process_id, user_intent, company_id)
131 # Validación del process_id
132 return valid_process_id(process_id) if valid_process_id(process_id).present?
133
134 # Obtenemos las cotizaciones del proceso
135 quotations_data = Quotation.analyze_quotations(process_id, company_id)
136
137 if quotations_data.present?
138 # Enviamos los datos y la intención del usuario a ChatGPT para el análisis
139 analysis = ChatgptService.analyze_ask(quotations_data, user_intent)
140
141 # Devolver el análisis al usuario
142 response = "<strong>Análisis: </strong><br>#{analysis}"
143 else
144 response = "No se encontraron cotizaciones para el proceso con ID <strong>#{process_id}</strong>."
145 end
146 end
147
148
149
150 # Método para validar que el process_id sea un número entero válido
151 def valid_process_id(process_id)
152 unless process_id.to_s =~ /\A\d+\z/
153 response = "<strong>El ID del proceso debe ser un número entero válido.</strong>"
154 end
155 end
156
157 # Método para parsear el JSON
158 def parse_json(response_content)
159 JSON.parse(response_content)
160 rescue JSON::ParserError
161 nil
162 end
163
```

Fuente: Autores

En caso de que la respuesta de la API no coincida con los datos existentes en la base, el sistema valida la respuesta, evitando errores y garantizando que el usuario siempre obtenga una respuesta adecuada. Finalmente, los resultados obtenidos son enviados a la vista, donde se presentan en un modal para su visualización por parte del usuario.

Para continuar con la tercera actividad de la Fase 3 del proyecto, se desarrolló una interfaz gráfica que permite a los usuarios interactuar con el módulo de integración de manera clara y concisa. La interfaz fue diseñada tomando como base los mockups previamente creados, los cuales sirvieron como guía para asegurar una experiencia de usuario intuitiva y funcional.

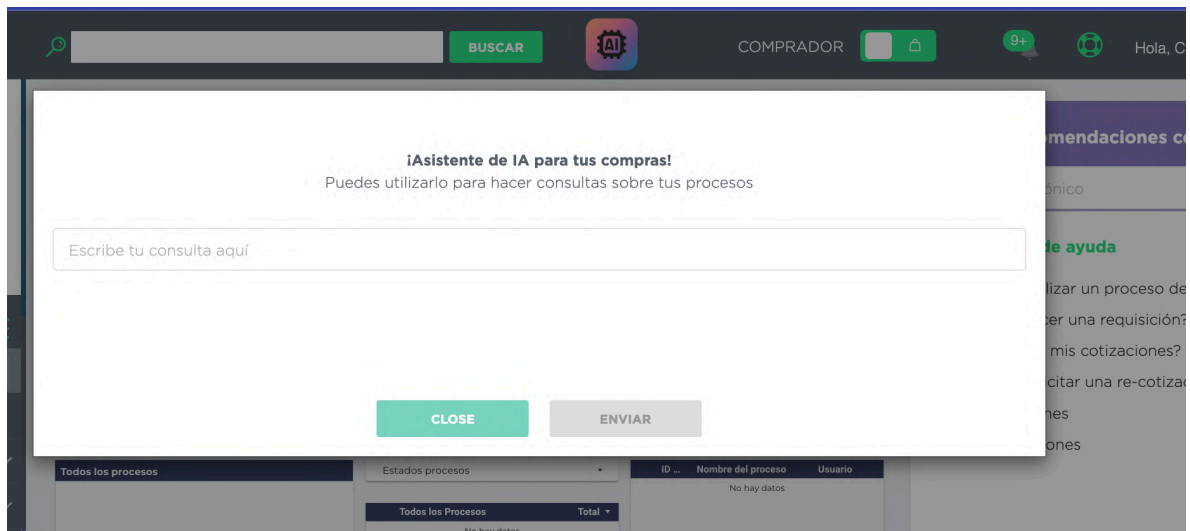
Figura 24. Interfaz gráfica del botón



Fuente: Autores

Durante el desarrollo, se añadieron validaciones importantes para mejorar la interacción con el sistema. Por ejemplo, se incluyó una validación que impide la ejecución de consultas si la casilla de preguntas está vacía, asegurando que el usuario proporcione la información necesaria antes de enviar una solicitud.

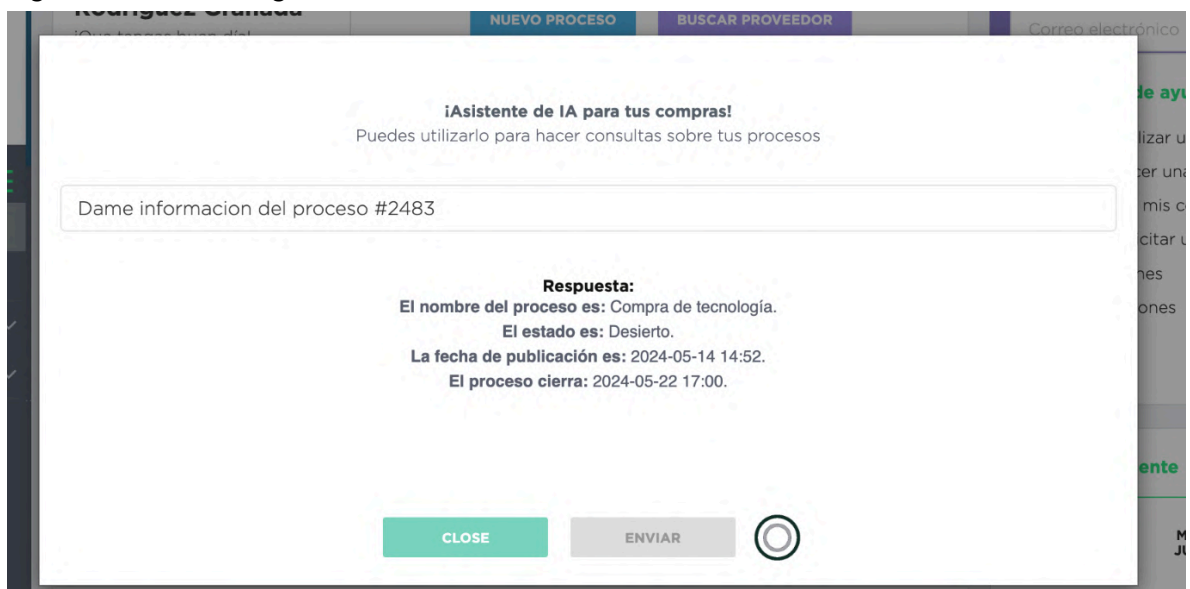
Figura 25. Interfaz gráfica modal



Fuente: Autores

Además, se implementó un loader que ofrece retroalimentación visual al usuario mientras la consulta está siendo procesada por el sistema (Figura 26). Esto mejora la experiencia, ya que el usuario entiende que su solicitud está siendo atendida y debe esperar la respuesta.

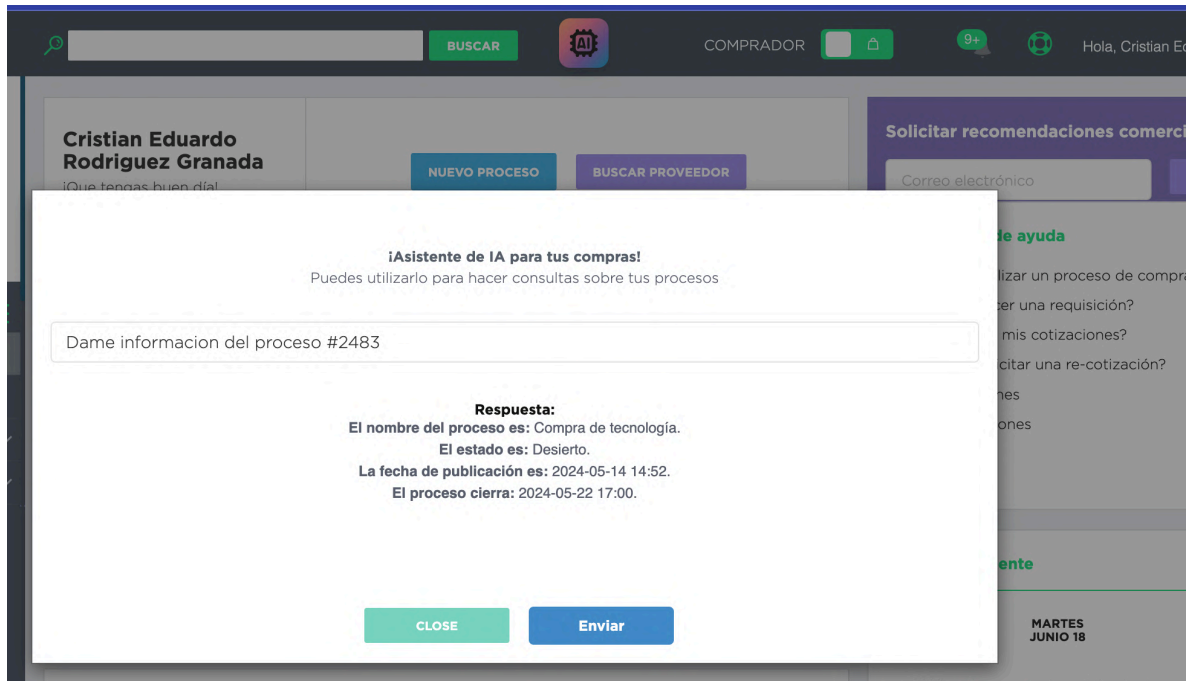
Figura 26. Interfaz gráfica modal con loader



Fuente: Autores

Por último, se agregó una función que limpia automáticamente el historial de preguntas cuando el usuario cierra el modal, garantizando una interfaz limpia y lista para nuevas consultas en cada interacción.

Figura 27. Interfaz gráfica con respuesta - parte 1



Fuente: Autores

Figura 28. Interfaz gráfica con respuesta - parte 2

Puedes utilizarlo para hacer consultas sobre tus procesos

dame las cotizaciones del proceso 2686 y dime cual seria la mejor opcion y porque

Respuesta:
Análisis:

De acuerdo con la información proporcionada, se han recibido dos cotizaciones diferentes, una del proveedor X y la otra del proveedor Y. A continuación, se mostrarán las características principales de cada una: **Cotización del Proveedor Y:**

- Identificación de cotización: 1303
- Fecha de validez de oferta: **2024-09-27**
- Plazo de entrega: **60 días**
- Costo de flete: **\$800,000.00 pesos**
- Plazo de pago: **30 días**
- Total de la cotización: **\$22,548,864.00 pesos**

Cotización del Proveedor X:

- Identificación de cotización: 1302
- Fecha de validez de oferta: **2024-09-06**
- Plazo de entrega: **12 días**
- Costo de flete: **\$0.00 pesos**
- Plazo de pago: **45 días**
- Total de la cotización: **\$27,251,070.00 pesos**

Después de analizar ambas cotizaciones, la cotización del **Proveedor Y parece más favorable debido a las siguientes razones:**

- La cotización del Proveedor Y es **menor** en comparación con la del Proveedor X, lo que le permitiría ahorrar costos.
- A pesar de que el Proveedor Y tiene un costo de flete más alto, el monto total aún es menor en comparación con el Proveedor X.
- El Proveedor Y ofrece un plazo de pago más corto. Esto podría ser beneficioso si se desea liquidar el pago rápidamente.

Dicho esto, existen otros factores a considerar que podrían inclinar la balanza. El plazo

CLOSE **Enviar**

Fuente: Autores

Para culminar la fase 3 de la investigación, se procedió a dar cumplimiento de la actividad 4, cuyo propósito fue implementar los mecanismos de control de acceso basados en permisos y roles de los usuarios, con el fin de asegurar que cada empresa sólo pueda visualizar sus propios datos y no los de otras compañías.

Para lograr esto, se realizaron ajustes en las consultas SQL utilizadas en el sistema, donde se pasó como parámetro el ID de la empresa correspondiente (Figura 29). De esta manera, las consultas se limitan a devolver únicamente la información pertinente a la empresa del usuario actual, garantizando un acceso controlado y seguro. Esta implementación contribuye a cumplir con las políticas de seguridad y protección de datos dentro del sistema, evitando fugas de información

y manteniendo la privacidad de los usuarios.

Figura 29. Consultas SQL con restricción de data - parte 1

```
# Trae el proceso siempre y cuando pertenezca a la empresa a la que se busca
def self.search_process_company(process_id, company_id)
  req = self.search_process(process_id, company_id)
  req.includes(:user).where("reqs.id = ? AND companies.id = ?", process_id, company_id).references(:user)
end

# Trae la información de un proceso
def self.process_info(process_id, company_id)
  req = self.search_process_company(process_id, company_id).first
  return nil unless req

  { id: req.id, name: req.name, status: req.status_name, date_publication: req.purchase.start_date, end_date: req.purchase.end_date }
end

# Trae las cotizaciones de una compra
def self.search_quotations_of_process(req_id, company_id)
  self.search_quotations_of_purchase(req_id, company_id).where("reqs.id = ? AND companies.id = ? AND quotations.status IN (?)", req_id, company_id, [Quotation::COTIZADO, Quotation::SELECCIONADA]).references(:purchase).references(:user)
end

return nil unless quot.any?

quot.map { |quot| { company: quot.company_name, status: quot.status_name, recotizacion: quot.version.present? } }
end

# Se trae todos los proveedores (invitados) y la información actualizada de su proceso de cotización también llamado tracker de invitados
def self.guests_info_tracker(req_id, company_id)
  guests = PurchaseProviders.includes(:purchase).where("reqs.id = ? AND companies.id = ? AND purchases.id = ?")
  SELECT id FROM purchases WHERE req_id = reqs.id ORDER BY created_at DESC LIMIT 1
  }, req_id, company_id).references(:purchase).references(:user)
end

return nil unless guests.any?

guests.map do |guest|
  quotation = Quotation.find_by(user_id: guest.user_id, purchase_id: guest.purchase_id)

  {
    user_company: guest.company_name, user_name: guest.user6.full_name,
    phone: guest.user6.phone.presence || "Sin definir",
    email: guest.user6.email,
    read: quotation.read ? "El proveedor ya accedió al proceso y revisó los detalles." : "El proveedor no ha accedido al proceso y no ha visto los detalles.",
    intention: guest.intention_name,
    send_questions: (quotation.status == Quotation::COTIZADO || quotation.status == Quotation::SELECCIONADA) ? "El proveedor ya envió su cotización" : "El proveedor no ha cotizado"
  }
end
end
```

Fuente: Autores

Figura 30. Consultas SQL con restricción de data - parte 2

```
324 # Trae las cotizaciones para analizar y procesar
325 def self.analyze_quotations(req_id, company_id)
326   quots = Quotation.includes(:purchase).where("reqs.id = ? AND companies.id = ?") # Es más eficiente si solo se necesita acceder a las relaciones cargadas directamente en el código, solo filtrar por ellas.
327   .includes(:product_quotes) # Solo product_quotes, ya que se necesita para acceder a esos datos para el mapeo de productos
328   .where("reqs.id = ? AND companies.id = ? AND quotations.status IN (?)", req_id, company_id, [Quotation::COTIZADO, Quotation::SELECCIONADA])
329   .select("quotations.id, quotations.company_id, quotations.payment_deadline, quotations.offer_validity, quotations.delivery_term, quotations.freight, quotations.is_negotiable")
330   .references(:product_quotes, :purchase).references(:user)
331 end
332 return nil unless quots.present?
333
334 quots.map do |quot|
335   products = ProductQuote.where(quotation_id: quot.id).pluck(:id, :amount_quoted, :unit_value, :subtotal, :value_iva, :total, :brand, :description, :deliver_date, :warranty_offered, :days_to_deliver, :currency)
336 end
337
338 quotation_id: quot.id, company_id: quot.company_name, payment_deadline: quot.payment_deadline, offer_validity: quot.offer_validity, delivery_term: quot.delivery_term, freight: quot.freight, is_negotiable: quot.is_negotiable,
339 products: products.map do |product|
340   {
341     product_id: product[0],
342     amount_quoted: product[1],
343     unit_value: product[2],
344     subtotal: product[3],
345     value_iva: product[4],
346     total: product[5],
347     brand: product[6].presence || "Sin información",
348     description: product[7],
349     deliver_date: product[8].presence || "Sin información",
350     warranty_offered: product[9].presence || "Sin información",
351     days_to_deliver: product[10],
352     currency: product[11].presence || "Sin información"
353   }
354 end
355 }
356 end
357 end
```

Fuente: Autores

Continuando con la fase 4, se buscó validar la correcta implementación del sistema y su integración con el modelo de Inteligencia Artificial GPT-4o mini llevando a cabo un protocolo de pruebas para medir indicadores y contrastar con los resultados iniciales, así como también una serie de pruebas que evaluaban si la IA respondía de manera correcta a datos válidos y si identificaba de manera precisa los errores al procesar datos inválidos, insuficientes o fuera del formato establecido. De esta forma, se buscó garantizar la robustez del sistema y su eficiencia en torno a interacciones claras y útiles para el usuario.

Las pruebas fueron realizadas bajo la supervisión de Juan Camilo Valencia Henao, en representación del equipo de desarrollo del proyecto, con el apoyo de Maria Ximena Cárdenas, como usuario experto en representación de la empresa HandelBay, y también por Katherine Urrea, en representación de uno de los clientes actuales de la compañía.

Se tomaron en cuenta una serie de precondiciones para ejecutar dicho protocolo de pruebas, entre ellas, configurar el ambiente de pruebas en el servidor con la dirección <https://testing.handelbay.com.co/> para que funcionara bajo parámetros estándar de producción, con acceso pleno a la integración con GPT-4o mini, una conexión estable a la base de datos para tener disponibilidad de datos históricos, y las credenciales necesarias para la comunicación con el sistema de backend de HandelBay. También se garantizó que los datos simulados para cada caso de prueba no presentaran conflictos ni datos repetidos, de manera que el sistema pudiera procesarlos y devolver respuestas con precisión.

Cuadro 5. Protocolo de pruebas funcionales y de rendimiento

Categoría	Código	Descripción	Objetivo
Pruebas funcionales y de rendimiento previas a la integración con GPT-4o mini	F01	Evaluar los indicadores de eficiencia en consulta general de tipo 1.	Establecer una línea base de rendimiento en términos de precisión, tiempo de ejecución, número de clics y cantidad de consultas internas SQL.
	F02	Evaluar los indicadores de eficiencia en consulta específica de tipo 1.	
	F03	Evaluar los indicadores de eficiencia en consulta de tipo 2.	
Pruebas funcionales y de rendimiento posterior a la integración	V01	Medir los indicadores de eficiencia en consulta general de tipo 1.	Validar y contrastar datos de precisión de la respuesta, tiempo de ejecución, reducción de pasos y consultas internas SQL.

con GPT-4o mini	V02	Medir los indicadores de eficiencia en consulta específica de tipo 1	
	V03	Medir los indicadores de eficiencia en consulta de tipo 2.	
Pruebas de validación de entrada	E01	Probar la respuesta del sistema ante datos incompletos.	Garantizar mensajes de error claros.
	E02	Verificar la respuesta del sistema ante datos ambiguos y/o cortos.	Validar que el desarrollo logre interpretar y dar solución.
	E03	Probar la respuesta del sistema ante datos ajenos (de otra compañía).	Garantizar la robustez y seguridad del desarrollo.
	E04	Verificar la respuesta del sistema ante datos con errores gramaticales.	Validar que el desarrollo logre interpretar y dar solución.
	E05	Revisar la respuesta del sistema ante datos inválidos.	Asegurar que la funcionalidad guíe al usuario correctamente.
	E06	Probar la respuesta del sistema ante datos fuera de contexto.	Garantizar mensajes de error claros.
Pruebas de revisión técnica	T01	Analizar el comportamiento ante varias solicitudes simultáneas.	Confirmar la eficiencia del sistema en condiciones reales.
	T02	Comprender el comportamiento ante peticiones extensas.	Validar la escalabilidad del sistema.
	T03	Validar el comportamiento ante la pérdida de conexión a internet.	Confirmar el correcto funcionamiento de la funcionalidad ante reanudaciones.

Pruebas de análisis interno para respuestas complejas	A01	Evaluar la precisión en el procesamiento de comparativos de cotizaciones.	Comprobar que las respuestas sean consistentes y útiles.
	A02	Analizar la precisión en el procesamiento y promedio de datos.	Mejorar la experiencia de usuario en análisis complejos.
	A03	Analizar la precisión en la entrega de sugerencias y recomendaciones al usuario.	Verificar que las sugerencias de la funcionalidad sean apropiadas.

Fuente: Autores

Vale resaltar que para que cada prueba se considerara exitosa, la solución debía interpretar correctamente la entrada, mostrando los mensajes de validación apropiados que orienten al usuario, por tanto, el éxito se midió en la exactitud de las respuestas y en la claridad de los mensajes de error. Cualquier variación o falla en los mensajes de error, en la detección de inconsistencias o en la interpretación de datos, se consideró como un fallo, con las consideraciones pertinentes que su caso amerite para la optimización en el sistema de interpretación de entrada de datos.

En relación a las pruebas centradas en evaluar la robustez técnica y la capacidad analítica del sistema, los criterios de éxito consideraron la capacidad del sistema para procesar varias consultas simultáneamente, ofrecer respuestas dentro del tiempo esperado, y realizar análisis complejos que fueran precisos y realistas. En caso de fallos, se evaluó la claridad de los mensajes de error y la posibilidad de mejorar la experiencia del usuario mediante ajustes técnicos y analíticos.

Una vez ejecutadas, las pruebas internas concluyeron que el desarrollo es funcionalmente sólido y responde adecuadamente a las consultas principales relacionadas con los procesos de compra y cotizaciones. Todas las consultas clave fueron aprobadas sin incidentes, y el rendimiento bajo carga fue satisfactorio y los indicadores de eficiencia fueron contrastados apropiadamente. El informe interno de dichos resultados puede ser consultado en el Anexo C.

Para finalizar, se realizó un manual de usuario para el manejo adecuado de la herramienta en términos de uso y buenas prácticas Anexo D, con su respectivo

video Anexo E. Adicionalmente, se construyó un manual técnico para la consulta de desarrolladores y profesionales de ingeniería (Anexo F) también con su respectivo video explicativo (Anexo G).

6.2 DISCUSIÓN DE RESULTADOS

Uno de los resultados más destacados en la integración de GPT-4o mini con la plataforma HandelBay es la mejora significativa en la interacción con el usuario, así como en la precisión y velocidad del procesamiento de consultas complejas. El asistente virtual, basado en la API de GPT-4o mini, facilita la interpretación de solicitudes en lenguaje natural, lo que permite a los usuarios obtener respuestas rápidas y pertinentes sobre procesos de compra, proveedores y cotizaciones. La implementación de esta solución tiene un gran potencial para optimizar el sistema de gestión de compras dentro de HandelBay, proporcionando una mayor eficiencia en la consulta y manejo de datos.

Cuadro 6. Indicadores de eficiencia con la integración de GPT-4o mini

Objetivo del caso de uso	Acciones para cumplir objetivo	Cantidad de clics	Tiempo promedio	Cantidad de consultas SQL
Consultar información general de un proceso de compra	Iniciar sesión Botón "Asistente virtual" Botón "Enviar Consulta"	3	30 seg.	2
Verificar estado de respuesta de los proveedores de un proceso de compra	Iniciar sesión Botón "Asistente virtual" Botón "Enviar Consulta"	3	30 seg.	3
Comparar y analizar las cotizaciones de un proceso de compra	Iniciar sesión Botón "Asistente virtual" Botón "Enviar Consulta"	3	45 seg.	2

Fuente: Autores

La capacidad de GPT-4o mini para adaptarse a las preguntas contextuales de los usuarios reduce los tiempos de búsqueda y permite un acceso más dinámico a la información, lo que mejora la toma de decisiones dentro de la organización. La comparación y análisis de los indicadores de eficiencia en torno a las alternativas

que posee el usuario para cumplir sus objetivos en gestión de Compras con HandelBay se pueden consultar en el Anexo H.

Este avance puede contrastarse con otras implementaciones de IA en el sector del comercio electrónico, donde se observa un creciente uso de asistentes virtuales para automatizar el servicio al cliente y la gestión de inventarios. Sin embargo, la integración de GPT-4o mini se distingue por su capacidad de aprendizaje supervisado y adaptación en tiempo real, lo que le otorga una ventaja competitiva al ser capaz de manejar un mayor volumen de consultas con menos intervención humana.

Algunos podrían argumentar que el uso de inteligencia artificial en la gestión de datos podría plantear riesgos en cuanto a la precisión de las respuestas y la seguridad de los datos almacenados. No obstante, este desafío ha sido mitigado mediante la implementación de algoritmos avanzados de control de calidad y protocolos de seguridad en HandelBay, lo que asegura la protección de la información y la fiabilidad de las respuestas.

Por otro lado, es esencial evaluar si esta integración satisface las necesidades específicas de HandelBay en cuanto a la gestión de compras y la interacción con el cliente. Aunque los resultados iniciales son prometedores, se deben realizar estudios adicionales para confirmar que la solución basada en GPT-4o mini responde de manera eficiente a los cambios y demandas del mercado. De hecho, para continuidad del proyecto, se recomendaría continuar con pruebas de usabilidad con usuarios reales y realizar ajustes finales en base a la retroalimentación obtenida.

Se espera que esta integración aporte grandes beneficios a la empresa, tanto en términos de competitividad como de productividad, y que la adopción de este tipo de IA represente un paso significativo en la automatización e innovación de las áreas de Compra colombianas.

7. CONCLUSIONES

El proyecto de integración del modelo GPT-4o mini con el software de compras HandelBay ha demostrado ser una solución eficiente, cumpliendo con los objetivos propuestos en términos de eficiencia y optimización de tiempos. La selección del modelo de IA adaptado a esas consultas recurrentes de los usuarios del software permitió optimizar el uso de tokens y solicitudes, lo que fue clave para impactar positivamente en la experiencia del usuario al facilitar respuestas precisas y reducir tiempos de espera. No obstante, el esquema utilizado, aunque eficiente, presenta limitaciones en cuanto a la flexibilidad de las consultas, lo que requerirá ajustes en el futuro para ampliar el alcance de preguntas más variadas.

La integración técnica del modelo GPT-4o mini con HandelBay fue robusta y exitosa, facilitando una interacción fluida entre la API de la IA y la base de datos del software, validando adecuadamente las solicitudes del usuario con la precisión requerida para apoyar decisiones de compra en tiempo real. Tras la validación interna, esta integración se encuentra ya en producción, disponible para los usuarios de HandelBay, demostrando que el desarrollo cumple con estándares empresariales de funcionalidad.

La interfaz de usuario, basada en los mockups iniciales, resultó clara y funcional, proporcionando una experiencia intuitiva y optimizada para el usuario final. Las mejoras visuales como el loader de carga y las validaciones de consultas enriquecieron la experiencia del usuario, asegurando que las interacciones fueran eficientes, manteniendo el flujo de trabajo diario sin interrupciones, reduciendo los clics y el tiempo requerido en las consultas.

En términos de seguridad, la implementación de roles y permisos garantiza que la solución cumpla con la promesa de optimizar procesos al tiempo que asegura la privacidad e integridad de la información sensible. El sistema está alineado con las necesidades del modelo de negocio y ha demostrado ser una solución escalable capaz de evolucionar con el sistema, adecuándose a mejoras o expansiones futuras que podrían ampliar su aplicabilidad en otros módulos.

8. RECOMENDACIONES

- Se recomienda aumentar la variedad de consultas disponibles para cubrir una gama más amplia de preguntas posibles, asegurando mayor flexibilidad en el uso del sistema por parte de los usuarios, paralelamente con el uso de la funcionalidad de GPT de ajuste fino (fine-tuning) para el entrenamiento customizado de datos. Por supuesto, esto deberá ser validado con el grupo de usuarios finales.
- Una mayor optimización del uso de tokens permitirá mantener los costos bajo control a medida que se expanda la funcionalidad del sistema. Por tanto, se recomienda monitorear y ajustar el uso de tokens para evitar un consumo excesivo en consultas más complejas.
- Será necesario implementar pruebas de usabilidad adicionales; realizar pruebas de usabilidad con un mayor número de usuarios finales para identificar posibles mejoras en la experiencia de interacción y optimización de la interfaz.
- Evaluar la posibilidad de permitir que los usuarios realicen consultas más dinámicas y personalizadas sin necesidad de ampliar el esquema predefinido, puede ser de gran utilidad y beneficios, especialmente utilizando un enfoque que combine procesamiento en la API y validaciones locales, además de integrarlas a aplicativos de comunicación muy usados como Whatsapp.
- Mantener un seguimiento y monitoreo constante del rendimiento del sistema, especialmente en condiciones de carga elevada, para asegurar que el sistema se mantenga estable y responda eficientemente a la creciente demanda.

REFERENCIAS BIBLIOGRÁFICAS

- Alcaldía de Bogotá. (s. f.). *Normas y documentación en la página oficial de la Alcaldía de Bogotá*.
<https://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=49981>
- Arman, M., & Lamiyar, U. R. (2023). Exploring the implication of ChatGPT AI for business: Efficiency and challenges. *Journal of Innovation Information Technology and Application*, 5(1), 52-64.
<https://app.dimensions.ai/details/publication/pub.1160827384>
- Bryant, S. (2023). Assessing GPT-4's role as a co-collaborator in scientific research: A case study analyzing Einstein's special theory of relativity. *Discov Artif Intell*, 3(26). <https://doi.org/10.1007/s44163-023-00075-3>
- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lungberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. *arXiv*. <https://arxiv.org/abs/2303.12712>
- De Miguel Castaño, A., Velthuis, M., Martínez, E. (1999) Diseño de bases de datos relacionales. Editorial Ra-ma. ISBN 9788478973859
- Departamento Administrativo de la Función Pública. (s. f.). *Decreto Ley 256 de 2013*.
<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=52014>
- Departamento Administrativo de la Función Pública. (s.f.). *Ley 1150 de 2007*.
<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=304>
- Departamento Administrativo de la Función Pública. (s.f.). *Ley 1581 de 2012*.
<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=49981>
- European Knowledge Center For Information Technology (2022). Waterfall: metodología para el desarrollo secuencial de tareas. TIC Portal.
<https://www.ticportal.es/glosariotic/waterfall-metodologia-desarrollo-secuencial>

- Filip, F. G., Zamfirescu, C. B., & Ciurea, C. (2017). *Computer-supported collaborative decision-making* (1st ed., pp. 216). Springer. <https://doi.org/10.1007/978-3-319-47221-8>
- Harskamp, R. E., & De Clercq, L. (2024). Performance of ChatGPT as an AI-assisted decision support tool in medicine: A proof-of-concept study for interpreting symptoms and management of common cardiac conditions (AMSTELHEART-2). *Acta Cardiologica*, 79(3), 358–366. <https://doi.org/10.1080/00015385.2024.2303528>
- HandelBay. (s. f.). *HandelBay: Software de gestión y automatización*. HandelBay. <https://handelbay.com/>
- HubSpot. (2023). *Procesamiento de lenguaje natural: qué es, ejemplos y herramientas*. HubSpot. <https://blog.hubspot.es/marketing/procesamiento-de-lenguaje-natural>
- HubSpot. (2024). *Prompt Engineering: qué es y cómo dominarlo con 5 cursos gratis*. HubSpot. <https://blog.hubspot.es/website/prompt-engineering#que-hace>
- Liu, S., Wright, A. P., Patterson, B. L., Wanderer, J. P., Turer, R. W., Nelson, S. D., McCoy, A. B., Sittig, D. F., & Wright, A. (2023). Using AI-generated suggestions from ChatGPT to optimize clinical decision support. *Journal of the American Medical Informatics Association*, 30(7), 1237-1245. <https://doi.org/10.1093/jamia/ocad072>
- Ma, T. J. (2023). *Big data decision support system*. (Trabajo de grado, Doctorado en Filosofía, Universidad Estatal de Colorado, Departamento de Ingeniería de Sistemas). <https://www.osti.gov/biblio/1875023>
- Nori, H., King, N., Mayer McKinney, S., Carignan, D., & Horvitz, E. (2023). Capabilities of GPT-4 on medical challenge problems. *arXiv*. <https://arxiv.org/abs/2303.13375>
- OpenAI. (s.f.). *Documentación y tutoriales en la página oficial de OpenAI*. <https://www.openai.com/>
- OpenAI. (s.f.). *OpenAI API documentation*. OpenAI. <https://platform.openai.com/>

OpenAI. (2023). *GPT-4 Technical Report*. <https://arxiv.org/abs/2303.08774>

OpenAI. (2024). *GPT-4o Mini Pricing*. <https://openai.com/api/pricing/>

Organización para la Cooperación y el Desarrollo Económicos. (2019). *Principios de la OCDE sobre la Inteligencia Artificial*. <https://www.oecd.org/going-digital/ai/principles>

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121-154. <https://www.sciencedirect.com/science/article/pii/S266734522300024X>

Ruby. (s.f.). *Documentación y tutoriales en la página oficial del lenguaje*. <https://www.ruby-lang.org/>

Ruby on Rails. (s.f.). *Documentación y tutoriales en la página oficial del framework*. <https://rubyonrails.org/>

Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare*, 11(6), 887. <https://doi.org/10.3390/healthcare11060887>

Smartmind. (s. f.). *Descubriendo las ramas de la inteligencia artificial*. Smartmind. <https://www.smartmind.net/blog/ramas-ia/>

ANEXO A CASOS DE USO

Caso de Uso 1: Consultar Estado de un Proceso

ID	CU01
Nombre	Consultar Estado de un Proceso
Descripción	El usuario consulta el estado actual de un proceso de compra o requisición.
Actor Principal	Maikol Rico Perez
Actores Secundarios	Sistema, Base de datos
Precondiciones	El usuario ha iniciado sesión y tiene permisos para consultar procesos.
Flujo Principal	<ol style="list-style-type: none"> 1. El usuario accede al módulo de ChatGPT. 2. El usuario pregunta "¿Cuál es el estado del proceso número X?". 3. El sistema interpreta la pregunta y ejecuta la acción process_status. 4. El sistema consulta la base de datos y obtiene el estado del proceso. 5. El sistema devuelve el estado del proceso.
Flujo Alternativo	Si el proceso no se encuentra, el sistema muestra un mensaje de error indicando que el proceso no existe o que no tiene acceso.
Postcondiciones	El usuario recibe la información sobre el estado del proceso.

Prioridad	Alta
Excepciones	El usuario no tiene permisos para ver el estado del proceso.

Caso de Uso 2: Análisis de Cotizaciones

ID	CU02
Nombre	Análisis de Cotizaciones
Descripción	El usuario solicita el análisis de las cotizaciones de un proceso de compra, obteniendo una recomendación sobre cuál elegir.
Actor Principal	Maikol Rico Perez
Actores Secundarios	Sistema, Base de datos
Precondiciones	El usuario ha iniciado sesión y tiene permisos para acceder a las cotizaciones.
Flujo Principal	<ol style="list-style-type: none"> 1. El usuario accede al módulo de ChatGPT. 2. El usuario pregunta "Analiza las cotizaciones del proceso número X y dime cuál debo elegir". 3. El sistema interpreta la pregunta y ejecuta la acción process_quotation_analysis. 4. El sistema recupera las cotizaciones de la base de datos. 5. El sistema analiza las cotizaciones y recomienda la mejor opción.
Flujo Alternativo	Si no hay suficientes cotizaciones, el sistema informa al usuario que no es posible realizar el análisis.

Postcondiciones	El usuario recibe el análisis de las cotizaciones y una recomendación.
Prioridad	Alta
Excepciones	No se encuentran cotizaciones para el proceso.

Caso de Uso 3: Listar Invitados a un Proceso

ID	CU03
Nombre	Listar Invitados a un Proceso
Descripción	El usuario solicita la lista de proveedores invitados a un proceso de compra a través del asistente virtual.
Actor Principal	Maikol Rico Perez
Actores Secundarios	Sistema, Base de datos
Precondiciones	El usuario ha iniciado sesión y tiene permisos para ver la lista de invitados.
Flujo Principal	<ol style="list-style-type: none"> 1. El usuario accede al módulo de ChatGPT. 2. El usuario pregunta "Lista los invitados al proceso número X". 3. El sistema interpreta la pregunta y ejecuta la acción process_guests. 4. El sistema consulta la base de datos y obtiene la lista de invitados. 5. El sistema devuelve la lista de invitados.

Flujo Alternativo	Si no se encuentran invitados, el sistema informa al usuario que no hay proveedores invitados en el proceso.
Postcondiciones	El usuario recibe la lista de proveedores invitados.
Prioridad	Media
Excepciones	El usuario no tiene permisos para ver la lista de invitados.

Caso de Uso 4: Ordenar Cotizaciones

ID	CU04
Nombre	Ordenar Cotizaciones
Descripción	El usuario solicita que las cotizaciones de un proceso se ordenen por un criterio específico, como precio o tiempo de entrega.
Actor Principal	Maikol Rico Perez
Actores Secundarios	Sistema, Base de datos
Precondiciones	El usuario ha iniciado sesión y tiene permisos para ver las cotizaciones.
Flujo Principal	<ol style="list-style-type: none"> 1. El usuario accede al módulo de ChatGPT. 2. El usuario pregunta "Ordena las cotizaciones del proceso número X por precio de mayor a menor". 3. El sistema interpreta la pregunta y ejecuta la acción process_quotation_analysis con la intención de orden. 4. El sistema consulta la base de datos y ordena las cotizaciones según el criterio solicitado. 5. El sistema devuelve las cotizaciones ordenadas.

Flujo Alternativo	Si no hay cotizaciones, el sistema informa al usuario que no hay cotizaciones para ordenar.
Postcondiciones	El usuario recibe las cotizaciones ordenadas según el criterio solicitado.
Prioridad	Alta
Excepciones	No se encuentran cotizaciones para el proceso.

Caso de Uso: Consultar Información de un Proceso

ID	CU05
Nombre	Consultar Información de un Proceso
Descripción	El usuario solicita información detallada sobre un proceso de compra o requisición.
Actor Principal	Maikol Rico Perez
Actores Secundarios	Sistema, Base de datos
Precondiciones	El usuario ha iniciado sesión y tiene permisos para ver la información del proceso.
Flujo Principal	<ol style="list-style-type: none"> 1. El usuario accede al módulo de ChatGPT. 2. El usuario pregunta "Dame información sobre el proceso número X". 3. El sistema interpreta la pregunta y ejecuta la acción process_info. 4. El sistema consulta la base de datos y obtiene la información del proceso.

	5. El sistema devuelve la información detallada del proceso al usuario.
Flujo Alternativo	Si el proceso no existe, el sistema muestra un mensaje indicando que el proceso no se encontró.
Postcondiciones	El usuario recibe la información detallada del proceso, incluyendo datos como el nombre, estado, fecha de publicación, y fecha de cierre.
Prioridad	Media
Excepciones	El usuario no tiene permisos para ver la información del proceso, o el proceso no existe.

ANEXO B DIAGRAMAS

Diagrama de Modelo Relacional:

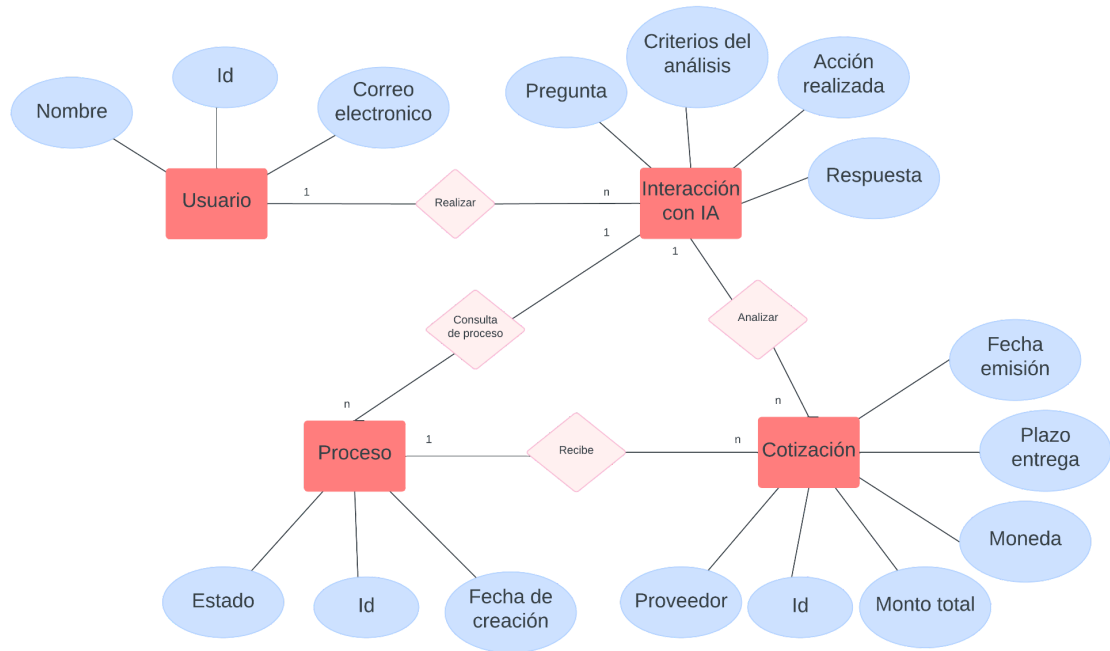


Diagrama de flujo del sistema:

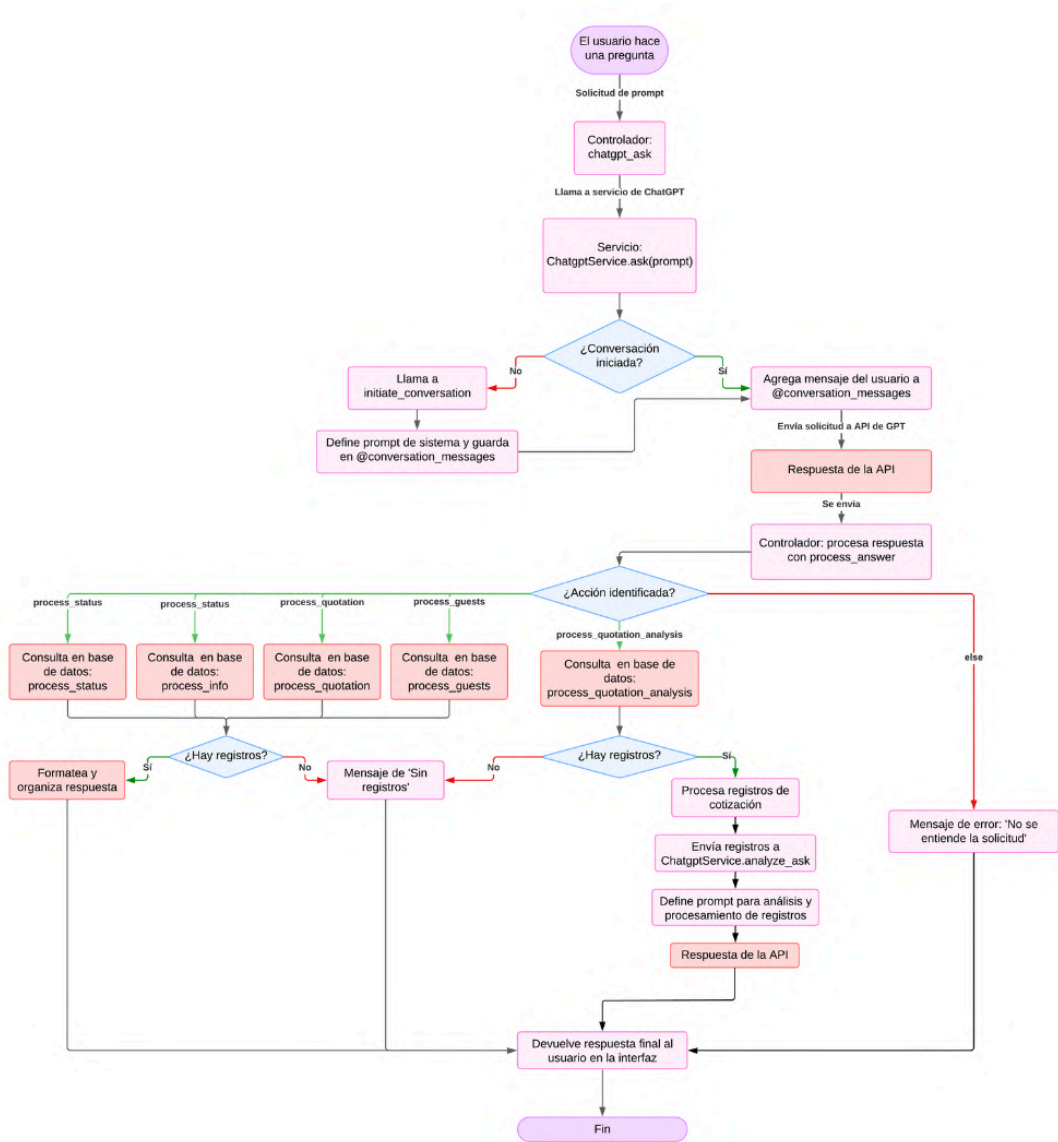
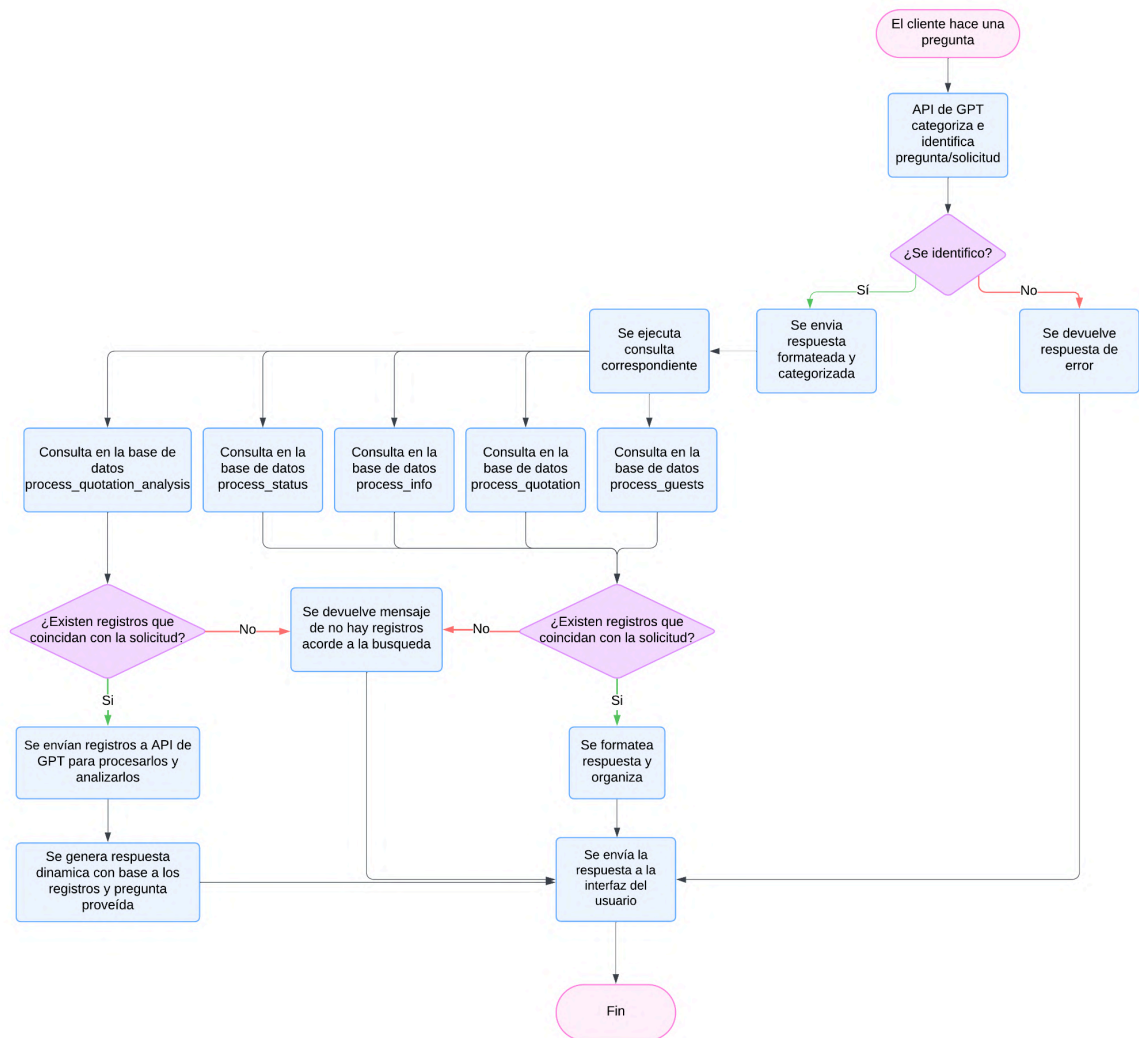


Diagrama de flujo del proceso de consulta y respuesta:



ANEXO C PRUEBAS



INFORME DE RESULTADO DE PRUEBAS

Objetivo general: Evaluar la eficiencia y precisión del software de gestión de compras de la plataforma HandelBay mediante un análisis comparativo de indicadores clave antes y después de la integración de GPT-4o mini, con el fin de determinar mejoras en el rendimiento, reducción de consultas SQL y optimización en la experiencia del usuario.

Responsable: Juan Camilo Valencia Henao

1. Pruebas funcionales y de rendimiento previas a la integración con GPT-4o mini

Objetivo: Evaluar los indicadores de eficiencia en la consulta de información de compras antes de la implementación de la IA, para establecer una línea base de rendimiento en términos de precisión, tiempo de ejecución, número de clics y cantidad de consultas internas SQL.

Fecha de la prueba: 26 de julio de 2024

Caso de uso	Ejecución	Indicador	Descripción	Resultado obtenido	Resultado esperado
F01 - Consultar información general de un proceso de compra.	El usuario consulta datos de compra en el gestor de procesos sin intervención de la IA.	Precisión	Respuestas correctas frente a las consultas.	La información completa del proceso se muestra sin errores.	Respuestas completas y precisas según la consulta.
		Tiempo de ejecución	Duración de consulta desde inicio a fin.	180 seg.	Reducción en un 30%.
		Cantidad	Número de clics necesarios	5	Reducción del

		de clics	para consulta.		número de clics.
		Consultas SQL	Cantidad de consultas a la base de datos.	4	Optimización al reducir la cantidad de consultas internas.
F02- Verificar el estado de respuesta de 3 proveedores de un proceso de compra.	El usuario ingresa al gestor de invitados de un proceso para consultar el estado de respuesta de 3 proveedores.	Precisión	Respuestas correctas frente a las consultas.	La información de los proveedores se muestra completa sin errores.	Respuestas completas y precisas según la consulta.
		Tiempo de ejecución	Duración de consulta desde inicio a fin.	240 seg.	Reducción en un 40%.
		Cantidad de clics	Número de clics necesarios para consulta.	6	Reducción del número de clics.
		Consultas SQL	Cantidad de consultas a la base de datos.	5	Optimización al reducir la cantidad de consultas internas.
F03- Comparar y analizar 3 cotizaciones de un proceso de compra	El usuario ingresa al gestor de cotizaciones del proceso para consultar la información de 3 cotizaciones, una por una.	Precisión	Respuestas correctas frente a las consultas.	La información de las cotizaciones se muestra completa sin errores.	Respuestas completas y precisas según la consulta.
		Tiempo de ejecución	Duración de consulta desde inicio a fin.	420 seg.	Reducción en un 50%.
		Cantidad de clics	Número de clics necesarios para consulta.	10	Reducción del número de clics.
		Consultas SQL	Cantidad de consultas a la base de datos.	9	Optimización al reducir la cantidad de consultas internas.

Observaciones: Los resultados iniciales muestran que la precisión de las respuestas es correcta, aunque la experiencia puede ser mejorable debido al tiempo de ejecución prolongado y a la cantidad significativa de clics para completar cada actividad, así como también por la cantidad de consultas internas SQL que el sistema debe hacer para generar la respuesta al usuario, en términos de rendimiento.

2. Pruebas funcionales y de rendimiento posterior a la integración con GPT-4o mini

Objetivo: Medir la eficiencia tras la implementación de la IA en la consulta de información de compras, con el fin de comparar mejoras en precisión de la respuesta, tiempo de ejecución, reducción de pasos y consultas internas SQL.

Fecha de la prueba: 30 de septiembre de 2024

Caso de uso	Ejecución	Indicador	Descripción	Resultado obtenido	Resultado esperado
V01- Consultar información general de un proceso de compra.	El usuario consulta datos de compra usando el asistente virtual.	Precisión	Respuestas correctas frente a las consultas.	La funcionalidad presenta la información solicitada del proceso sin errores.	Respuestas completas y precisas según la consulta.
		Tiempo de ejecución	Duración de consulta desde inicio a fin.	30 seg.	Reducción en un 30%.
		Cantidad de clics	Número de clics necesarios para consulta.	3	Reducción del número de clics.
		Consultas SQL	Cantidad de consultas a la base de datos.	2	Optimización al reducir la cantidad de consultas internas.
V02- Verificar el estado de	El usuario usa el asistente virtual para	Precisión	Respuestas correctas frente a las	La funcionalidad presenta la información	Respuestas completas y

respuesta de 3 proveedores de un proceso de compra.	consultar el estado de 3 proveedores invitados a un proceso de compra.		consultas.	solicitada de los proveedores sin errores.	precisas según la consulta.
		Tiempo de ejecución	Duración de consulta desde inicio a fin.	30 seg.	Reducción en un 40%.
		Cantidad de clics	Número de clics necesarios para consulta.	3	Reducción del número de clics.
		Consultas SQL	Cantidad de consultas a la base de datos.	3	Optimización al reducir la cantidad de consultas internas.
V03- Comparar y analizar 3 cotizaciones de un proceso de compra .	El usuario emplea el asistente virtual para consultar, comparar y analizar las 3 cotizaciones recibidas en un proceso de compra.	Precisión	Respuestas correctas frente a las consultas.	La funcionalidad presenta la información solicitada de las cotizaciones sin errores.	Respuestas completas y precisas según la consulta.
		Tiempo de ejecución	Duración de consulta desde inicio a fin.	45 seg.	Reducción en un 50%.
		Cantidad de clics	Número de clics necesarios para consulta.	3	Reducción del número de clics.
		Consultas SQL	Cantidad de consultas a la base de datos.	2	Optimización al reducir la cantidad de consultas internas.
Observaciones: Con la IA integrada, las respuestas continúan siendo precisas pero ahora de manera centralizada; el tiempo de ejecución se redujo considerablemente, y las acciones a ejecutar ahora					

requieren menos clics. Asimismo, se observa una notable optimización del sistema al reducir la cantidad de consultas internas SQL.

3. Pruebas de validación de entrada de la integración con GPT-4o mini

Objetivo: Verificar que la funcionalidad maneje correctamente datos inválidos, insuficientes o fuera de formato, devolviendo mensajes de error apropiados cuando se introducen entradas que no cumplen con los requisitos.

Fecha de la prueba: 5 de octubre de 2024

Caso de prueba	Entrada	Acción esperada	Resultado esperado	Resultado obtenido	Observaciones
E01 - Entrada insuficiente.	"2697"	Detectar una instrucción incompleta.	"La pregunta no corresponde a ninguna acción válida"	"La pregunta no corresponde a ninguna acción válida"	Funciona correctamente
E02 - Entrada corta / ambigua	"Información 2697"	Interpretar el ID de un proceso.	Presentar la información solicitada del proceso sin errores.	La funcionalidad presenta la información solicitada del proceso sin errores.	Funciona correctamente
E03- Instrucción con data de otra compañía	"¿Cuál es el estado del proceso 500?"	Detectar que la información solicitada no está disponible por seguridad.	"No se encontró el proceso con ID 500".	"No se encontró el proceso con ID 500".	Considerar ajuste en el mensaje de salida para informar de restricciones.
E04 - Instrucción con errores gramaticales	"dam información 2697"	Interpretar a pesar de errores gramaticales.	Presentar la información solicitada del proceso sin errores.	La funcionalidad presenta la información solicitada del proceso sin errores.	Funciona correctamente; explorar mejoras de NLP para este escenario.

E05 - Instrucción larga sin sentido	“Abcdefg hijklmn opqrstu”	Detectar una instrucción sin sentido.	"La pregunta no corresponde a ninguna acción válida"	"La pregunta no corresponde a ninguna acción válida"	Funciona correctamente; explorar mejoras de NLP para este escenario.
E06 - Instrucción fuera del contexto de la plataforma	“¿Cuál es el valor del dólar hoy?”	Detectar una instrucción fuera de contexto.	"La pregunta no corresponde a ninguna acción válida"	"La pregunta no corresponde a ninguna acción válida"	Considerar ajuste en el mensaje de salida para informar de restricciones.
<p>Observaciones: Las pruebas de validación de entrada revelan que la integración con GPT-4o mini responde adecuadamente ante datos insuficientes o incoherentes, mostrando respuestas claras que orientan al usuario sobre la invalidez de su instrucción; sin embargo, algunos casos específicos requieren de una mejora del mensaje de error para especificar con mayor claridad el ajuste necesario y/o como debe introducir preguntas más precisas. Estas observaciones sugieren que, al afinar estos detalles, la funcionalidad podrá ofrecer una experiencia de usuario cada vez más intuitiva y robusta en cuanto a entradas inválidas.</p>					

4. Pruebas de revisión técnica de la integración con GPT-4o mini

Objetivo: Evaluar el rendimiento y la estabilidad del sistema ante diferentes escenarios de carga y ejecución de consultas.

Fecha de la prueba: 6 de noviembre de 2024

Caso de prueba	Entrada	Acción esperada	Resultado esperado	Resultado obtenido	Observaciones
T01 - Carga simultánea de consultas	5 consultas en paralelo	Procesar múltiples consultas sin degradación del rendimiento.	Consultas procesadas sin errores ni retrasos significativos.	Consultas procesadas sin errores.	Funciona correctamente
T02 - Consulta extensa de datos	Solicitud de datos de +100 invitados a un proceso de compra	Entregar resultados completos y ordenados rápidamente.	Respuesta completa en menos de 2 segundos.	Respuesta completa en menos de 4 segundos.	Funciona correctamente; considerar acelerar el tiempo de espera.
T03- Interrupción de conexión	Consulta interrumpida intencionalmente.	Manejar la interrupción y permitir la reanudación	Reanudación sin errores.	Reanudación sin errores.	Funciona correctamente; considerar mensaje de reintento o de detección de falta de conexión..
<p>Observaciones: Las pruebas de revisión técnica han evidenciado la capacidad del sistema para manejar consultas simultáneas y solicitudes complejas sin comprometer el rendimiento. Sin embargo, se detectaron leves retrasos en escenarios de alta carga, sugiriendo la necesidad de optimizar las rutinas de manejo de solicitudes concurrentes. Adicionalmente, la recuperación ante interrupciones se mostró sólida, garantizando una experiencia continua para el usuario.</p>					

5. Pruebas de análisis interno para respuestas complejas a la integración con GPT-4o mini

Objetivo: Verificar la capacidad del sistema para realizar análisis avanzados basados en entradas complejas y generar respuestas procesadas.

Fecha de la prueba: 6 de noviembre de 2024

Caso de prueba	Entrada	Acción esperada	Resultado esperado	Resultado obtenido	Observaciones
A01 - Comparativo de cotizaciones	“Compara la cotizaciones de los proveedores para el proceso 2541”	Analizar los datos de las cotizaciones y generar un resumen comparativo.	Respuesta que resalte diferencias clave: precios, plazos y validez de la oferta.	Comparativo de las cotizaciones con texto adicional que resalta las diferencias.	Funciona correctamente
A02 - Análisis de rendimiento	“Analiza el tiempo de entrega promedio de los proveedores en el proceso 2686”	Calcular el tiempo promedio y analizar las recomendaciones para este caso en particular.	Respuesta con los tiempos de cada proveedor analizados, el promedio final y un análisis con recomendaciones	Cuadro con los tiempos de cada proveedor analizados, el promedio final y un análisis con recomendaciones	Funciona correctamente
A03- Análisis de costos	“Analiza los totales cotizados en el proceso 2630”	Usar datos históricos para proyectar un análisis y sugerencias.	Respuesta con los detalles de precios de cada cotización recibida y un análisis comparativo con sugerencias.	Respuesta con los detalles de precios de cada cotización recibida y un análisis comparativo con sugerencias.	Funciona correctamente ; considerar mensaje de reintento o de detección de falta de conexión..
Observaciones: El sistema demostró una notable capacidad para realizar análisis avanzados, como la comparación de cotizaciones y el análisis de costos y tiempos, brindando respuestas detalladas y útiles. No está demás para algunos casos específicos mejorar la profundidad del análisis para ofrecer mayor precisión en las recomendaciones, especialmente en escenarios con datos limitados.					

6. Conclusiones

Tras realizar las pruebas, se concluye que la integración de GPT-4o mini mantuvo y centralizó la precisión de las respuestas, redujo significativamente el tiempo de ejecución y disminuyó la cantidad de clics y consultas SQL. Los resultados confirman una experiencia de usuario optimizada

y mayor eficiencia en el proceso de consulta de información de compras, cumpliendo con los objetivos propuestos y verificando que los datos suministrados correspondan a los del usuario y sobre el contexto de su gestión de compras. Adicionalmente, las pruebas técnicas validaron la capacidad del sistema para manejar cargas concurrentes, mientras que el análisis interno resaltó su eficacia al procesar consultas complejas, como comparativos de cotizaciones, mejorando el rendimiento técnico y la satisfacción del usuario.

ANEXO D MANUAL DE USO

MANUAL DE USO DEL USUARIO



Introducción

Este manual está diseñado para ayudarte a usar el **asistente virtual** integrado en la plataforma B2B de **HandelBay**. A través de esta herramienta, puedes realizar consultas sobre tus procesos de compra, cotizaciones, proveedores invitados, y más, todo mediante preguntas sencillas.

El sistema utiliza inteligencia artificial (IA) para interpretar tus preguntas y ofrecerte información clara y precisa. A continuación, te explicamos paso a paso cómo puedes aprovechar esta funcionalidad.

Acceso al Asistente Virtual

1. Iniciar sesión en la plataforma:

- Dirígete a la página principal de HandelBay.
- Inicia sesión con tu usuario y contraseña.

2. Acceder al asistente de IA:

- Una vez que inicies sesión, encontrarás el **módulo de ChatGPT** a través de un botón destacado en la interfaz en la parte superior al lado del buscador.
- Haz clic en el botón de "**Asistente de IA**".

Consultas Disponibles

A través del asistente virtual, puedes realizar las siguientes consultas:

1. Consultar Estado de un Proceso:

Esta funcionalidad te permite consultar el estado de un proceso de compra o requisición.

- **Instrucciones:**
 - Escribe la pregunta: "**¿Cuál es el estado del proceso número [número de proceso]?**".
 - Presiona Enter o haz clic en "**Enviar**".
 - El sistema te devolverá el estado del proceso, por ejemplo: "**El estado de tu proceso 12345 es 'En Progreso'**".
- **Ejemplo:**
 - Pregunta: "**¿Cuál es el estado del proceso número 12345?**"
 - Respuesta: "**El estado de tu proceso 12345 es 'Finalizado'**".

2. Consultar Información de un Proceso

Puedes obtener detalles de un proceso de compra, incluyendo el nombre del proceso, estado, fecha de publicación y fecha de cierre.

- **Instrucciones:**
 - Escribe la pregunta: "**Dame información sobre el proceso número [número de proceso]**".
 - Presiona Enter o haz clic en "**Enviar**".
 - El sistema te proporcionará información detallada sobre el proceso.
- **Ejemplo:**
 - Pregunta: "**Dame información sobre el proceso número 67890.**"
 - Respuesta:
 - **Nombre del proceso:** Proceso de compra ABC.
 - **Estado:** En Progreso.
 - **Fecha de Publicación:** 01/08/2024.
 - **Fecha de Cierre:** 31/08/2024.

3. Análisis de Cotizaciones

Esta funcionalidad te permite obtener un análisis comparativo de las cotizaciones disponibles para un proceso de compra, recomendando la opción más favorable.

- **Instrucciones:**

- Escribe la pregunta: "**Analiza las cotizaciones del proceso número [número de proceso] y dime cuál debo elegir**".
- Presiona Enter o haz clic en "**Enviar**".
- El sistema te mostrará las cotizaciones disponibles, destacando la opción recomendada con una breve explicación.

- **Ejemplo:**

- Pregunta: "**Analiza las cotizaciones del proceso número 2840 y dime cuál debo elegir**".
- Respuesta: "**La cotización más favorable para el proceso 2840 es la de la Compañía XYZ, debido a su menor tiempo de entrega y mejor oferta económica.**"

También puedes pedir que las cotizaciones sean ordenadas de distintas maneras:

- **Pregunta: "Ordena las cotizaciones del proceso número 2840 de la más costosa a la más barata."**
- **Respuesta:** El sistema devolverá una lista de las cotizaciones ordenadas por el criterio solicitado.

4. Listar Proveedores Invitados a un Proceso

Puedes obtener una lista de los proveedores que han sido invitados a un proceso de compra y su estado en el proceso de cotización.

- **Instrucciones:**

- Escribe la pregunta: "**Lista los invitados al proceso número [número de proceso]**".
- Presiona Enter o haz clic en "**Enviar**".
- El sistema te devolverá la lista de invitados con sus respectivos nombres, empresas y el estado de su cotización.

- **Ejemplo:**

- Pregunta: "**Lista los invitados al proceso número 2840.**"
- Respuesta:
 - **Invitados al proceso 2840:**
 - Compañía A - **Estado:** Cotización enviada.
 - Compañía B - **Estado:** No ha enviado cotización

5. Consultar Cotizaciones de un Proceso

La funcionalidad "**process_quotation**" te permite consultar las cotizaciones asociadas a un proceso de compra o requisición. Puedes obtener detalles como el nombre de la compañía, el estado de la cotización, y si existe recotización.

- **Instrucciones:**
 - Escribe la pregunta: "**Dame las cotizaciones del proceso número [número de proceso]**".
 - Presiona Enter o haz clic en "**Enviar**".
 - El sistema te devolverá una lista de las cotizaciones asociadas al proceso solicitado.
- **Ejemplo:**
 - Pregunta: "**Dame las cotizaciones del proceso número 5678**".
 - Respuesta:
 - **El proceso 5678 tiene las siguientes cotizaciones:**
 - **Compañía A:** Estado: Cotizado. ¿Recotización?: No.
 - **Compañía B:** Estado: Cotizado. ¿Recotización?: Sí.
 - **Compañía C:** Estado: Seleccionada. ¿Recotización?: No.

Manejo de Errores

Si el sistema no puede interpretar tu pregunta, te devolverá un mensaje de error indicando que la pregunta no tiene correspondencia con ninguna de las opciones válidas.

- **Ejemplo:**
 - Pregunta: "**Quiero saber algo que no tiene sentido.**"
 - Respuesta: "**La pregunta no corresponde a ninguna acción válida. Por favor, intenta nuevamente con una pregunta más clara.**"

Preguntas Frecuentes

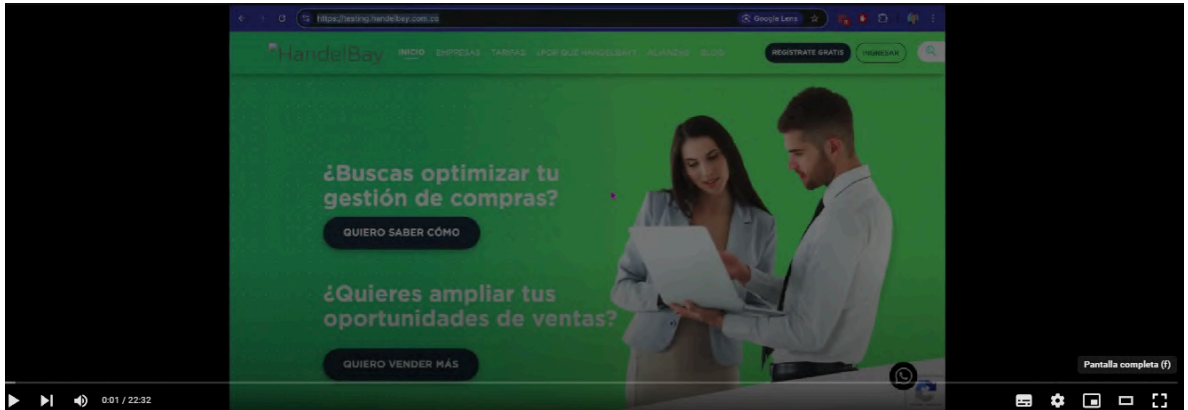
1. **¿Qué sucede si mi pregunta no es entendida por el asistente?**
 - Si el asistente no puede procesar tu solicitud, te devolverá un mensaje de error. Asegúrate de formular la pregunta en términos claros, como los ejemplos dados en este manual.
2. **¿Puedo realizar varias consultas consecutivas?**
 - No, sólo puedes hacer una pregunta a la vez.
3. **¿Cómo puedo realizar un análisis de cotizaciones?**
 - Puedes preguntar algo como: "**Analiza las cotizaciones del proceso número [número de proceso] y dime cuál es la más favorable.**" El asistente te

proporcionará una recomendación basándose en la oferta más conveniente, considerando el precio, tiempo de entrega y otros factores.

4. **¿Puedo ordenar las cotizaciones por diferentes criterios?**
 - Sí, puedes solicitar que las cotizaciones se ordenen según distintos criterios. Ejemplos de preguntas incluyen:
 - i. **"Ordena las cotizaciones del proceso número 5678 de mayor a menor precio."**
 - ii. **"Ordena las cotizaciones del proceso número 5678 por el tiempo de entrega más corto."**
 - El asistente devolverá la lista de cotizaciones en el orden que solicitaste.
5. **¿Qué debo hacer si el sistema no encuentra información relacionada con mi consulta?**
 - Si el sistema no puede encontrar un proceso o cotización, te devolverá un mensaje indicando que no se encontró información. Asegúrate de verificar el número del proceso o los datos que ingresaste.
6. **¿Qué tipo de datos puedo obtener sobre los proveedores invitados?**
 - Al consultar los invitados a un proceso, recibirás información como:
 - i. Nombre de la empresa invitada.
 - ii. Estado de participación (si ha enviado cotización o no).
 - iii. Datos de contacto, como nombre del representante y correo electrónico.
 - Pregunta ejemplo: **"Lista los invitados del proceso número [número de proceso]"**.
7. **¿Qué pasa si hay errores en la respuesta que recibo?**
 - Si el asistente no puede interpretar tu pregunta, te devolverá un mensaje de error. Puedes formular tu pregunta de manera más específica o usar las preguntas sugeridas como guía. Ejemplo de error: **"La pregunta no corresponde a ninguna acción válida."**
8. **¿Puedo realizar consultas sobre varios procesos al mismo tiempo?**
 - Actualmente, el asistente responde consultas sobre un proceso a la vez. Si deseas obtener información sobre múltiples procesos, deberás hacer una consulta por cada uno de ellos.
9. **¿El asistente virtual está disponible 24/7?**
 - Sí, el asistente virtual está disponible en todo momento para responder tus consultas. Sin embargo, si encuentras algún problema, puedes comunicarte con el equipo de soporte durante el horario laboral.
11. **¿Cómo puedo asegurarme de que mi pregunta sea clara para el asistente?**
 - Es importante formular preguntas directas y específicas. Por ejemplo, si deseas conocer el estado de un proceso, usa una pregunta como: **"¿Cuál es el estado del proceso número [número de proceso]?"**. Preguntas demasiado generales o ambiguas pueden generar un mensaje de error.

ANEXO E MANUAL DE USO (VIDEO EXPLICATIVO)

Enlace: https://youtu.be/8_AC1RIQILo



ANEXO F MANUAL TÉCNICO

MANUAL TÉCNICO PARA LA INTEGRACIÓN CON LA API DE CHATGPT

Introducción

Este manual describe los pasos y la estructura del código para integrar una aplicación con la API de OpenAI, utilizando Ruby on Rails como entorno de desarrollo. Se explicará la configuración y el funcionamiento del código del servicio **ChatgptService**, la manera de realizar peticiones a la API y procesar las respuestas. También se explica cómo se realiza la integración en la interfaz del usuario.

Requisitos previos

- Ruby on Rails versión 6.0 o superior.
- HTTParty Gem: utilizada para realizar solicitudes HTTP a la API de OpenAI.
- API Key de OpenAI: es fundamental para acceder a los servicios de la API y debe protegerse como una variable de entorno.

1. Configuración de la API

En este proyecto, se utiliza la librería **HTTParty** para hacer peticiones HTTP a la API de OpenAI. Se define la URL base (`base_url`) para interactuar con la API de OpenAI.

El archivo de configuración de la API es `app/services/chatgpt_service.rb`.

```
class ChatgptService
  include HTTParty
  base_uri 'https://api.openai.com/v1'
  ...
end
```

Los puntos importantes son:

- **Base URI:** Se especifica la dirección **https://api.openai.com/v1**, que es el punto de entrada a los servicios de OpenAI.
- **Token de autenticación:** Se requiere un token (API Key) que debe guardarse en una variable de entorno (**ENV['CHATGPT_API_KEY']**). Esto asegura que el acceso a la API sea seguro.

```
require 'httparty'

class ChatgptService
  include HTTParty
  base_uri 'https://api.openai.com/v1'

  # Iniciar la conversación con el prompt de sistema
  def self.initiate_conversation
    ...
  end

  # Método que interactúa con la API de OpenAI
  def self.ask(prompt)
    ...
  end

  # Método que interactúa con la API para analizar las cotizaciones
  def self.analyze_ask(quotations_data, user_intent)
    ...
  end
end
```

2. Estructura del Servicio ChatgptService

El archivo `chatgpt_service.rb` define dos métodos principales: `initiate_conversation` y `ask`. Veamos cómo funciona cada uno de ellos:

2.1 Método `initiate_conversation`

Este método inicializa la conversación definiendo el `system_prompt`. El `system_prompt` es un mensaje que indica a la API de ChatGPT qué tipo de respuestas debe proporcionar. Las instrucciones se centran en devolver las respuestas en un formato JSON con ciertas acciones que incluyen:

- **process_status**: Devuelve el estado de un proceso de compra o requisición.
- **process_info**: Proporciona información detallada de un proceso o requisición.
- **process_quotation**: Devuelve información sobre las cotizaciones.
- **process_quotation_analysis**: Analiza las cotizaciones disponibles.
- **process_guests**: Proporciona información sobre los proveedores invitados.
- **error**: Utilizado si la pregunta del usuario no coincide con ninguna acción definida.

```
# Inicializa la conversación con un prompt de sistema para guiar el comportamiento de ChatGPT
def self.initiate_conversation
  system_prompt = <<<PROMPT
  Eres un asistente virtual para una plataforma B2B. Todas tus respuestas deben seguir estrictamente el siguiente formato JSON:

  {
    "action": "nombre_de_la_accion",
    "data": {
      "query": "valor"
    }
  }

  Las acciones posibles son:
  - "process_status": Devuelve el estado de un proceso.
  - "process_info": Devuelve información detallada del proceso.
  - "process_quotation": Muestra cotizaciones.
  - "process_quotation_analysis": Analiza cotizaciones.
  - "process_guests": Información de proveedores invitados.
  - "error": Indica que la solicitud no es válida.

  Responde únicamente en formato JSON, sin ningún texto adicional.
  PROMPT

  # Guardamos el prompt inicial en el contexto de conversación
  @conversation_messages = [{ role: 'system', content: system_prompt }]
end
```

2.2 Método `ask`

El método `ask` se encarga de enviar las preguntas del usuario a la API y obtener una respuesta. El flujo del método `ask` es:

- **Inicialización de la conversación:** Verifica si la conversación ya fue iniciada; si no, llama al método `initiate_conversation()`.
- **Construcción de mensajes:** Agrega el mensaje del usuario a la conversación y lo guarda en la variable `@conversation_messages`.
- **Petición HTTP:** Hace una petición **POST** a la API usando **HTTParty**. Envía el prompt recibido, el modelo a usar (en este caso **gpt-4o-mini**), el contexto del chat y los encabezados de autenticación.
- **Manejo de Respuestas:** La respuesta de la API se guarda en la conversación, y se devuelve al usuario procesándola con `parsed_response`.

```
# Método que interactúa con la API de OpenAI
def self.ask(prompt)
  initiate_conversation() if @conversation_messages.nil?

  @conversation_messages << { role: 'user', content: prompt }

  response = post('/chat/completions',
    headers: {
      'Content-Type' => 'application/json',
      'Authorization' => "Bearer #{ENV['CHATGPT_API_KEY']}"
    },
    body: {
      model: 'gpt-4o-mini',
      messages: @conversation_messages,
      max_tokens: 150
    }.to_json)

  @conversation_messages << { role: 'assistant', content: response['choices'][0]['message']['content'] }
  response.parsed_response['choices'][0]['message']['content'] rescue "Lo siento, hubo un problema con la solicitud."
end
```

3. . Controlador ChatgptController

Este controlador se encarga de recibir las solicitudes de la vista y mandar las peticiones a la API. Veamos cómo funciona:

- **Método `chatgpt_ask`:** Recibe el **prompt** del usuario y llama al servicio `ChatgptService.ask(prompt)` para procesar la solicitud. Luego llama al método `process_answer` para trabajar con la respuesta antes de enviarla a la vista.

```
# Se encarga de mandar el prompt a ChatGPT y devolver la respuesta
def chatgpt_ask
  response_content = ChatgptService.ask(params[:prompt])

  response_content = process_answer(response_content)

  render json: { response: response_content }, status: :ok
end
```

- **Método process_answer:** Este método analiza la respuesta JSON de ChatGPT para identificar qué acción ejecutar. Dependiendo de la acción devuelta por la API (“process_status”, “process_info”, etc.), se llama a métodos específicos para procesar cada tipo de respuesta.

```
def process_answer(response_content)
  parsed_response = parse_json(response_content)
  return response_content unless parsed_response

  action = parsed_response['action']
  data = parsed_response['data']
  company_id = current_user.company_id

  case action
  when 'process_status'
    process_status(data['query'], company_id)
  when 'process_info'
    process_info(data['query'], company_id)
  when 'process_quotation'
    process_quotation(data['query'], company_id)
  when 'process_guests'
    process_guests(data['query'], company_id)
  when 'process_quotation_analysis'
    process_quotation_analysis(data['query'], data['user_intent'], company_id)
  else
    "<strong>#{data['query']}</strong>"
  end
end
```

4. Métodos en el case del Controlador

- **process_status:** Busca el estado de un proceso y devuelve el resultado.
- **process_info:** Devuelve información detallada de un proceso, incluyendo nombre, estado, fechas, etc.
- **process_quotation:** Lista las cotizaciones de un proceso y detalla cada una.
- **process_guests:** Muestra información de los proveedores invitados al proceso.
- **process_quotation_analysis:** Analiza las cotizaciones en función de la intención del usuario y usa el método **analyze_ask**.
- **else:** el caso error se activa cuando la pregunta no coincide con las acciones establecidas. Este bloque devuelve un mensaje informativo al usuario, guiándolo para mejorar su pregunta.

Explicación Detallada de “process_quotation_analysis”

Este método utiliza la función **analyze_ask** de **ChatgptService** para analizar las cotizaciones, compararlas y resaltar la mejor opción, o clasificarlas según la intención del usuario.

```
# Se encarga de analizar la data de las cotizaciones para procesar y mostrar respuestas
def process_quotation_analysis(process_id, user_intent, company_id)
  # Validación del process_id
  return valid_process_id(process_id) if valid_process_id(process_id).present?

  # Obtenemos las cotizaciones del proceso
  quotations_data = Quotation.analyze_quotations(process_id, company_id)

  if quotations_data.present?
    # Enviamos los datos y la intención del usuario a ChatGPT para el análisis
    analysis = ChatgptService.analyze_ask(quotations_data, user_intent)

    # Devolver el análisis al usuario
    "<strong>Análisis: </strong><br>#{analysis}"
  else
    "No se encontraron cotizaciones para el proceso con ID <strong>#{process_id}</strong>."
  end
end
```

5. Método analyze_ask en el Servicio ChatgptService

El método **analyze_ask** toma los datos de cotización y la intención del usuario y realiza un análisis utilizando la API de OpenAI.

```
# Método que interactúa con la API para analizar las cotizaciones
def self.analyze_ask(quotations_data, user_intent)
  prompt = <<<PROMPT
  Aquí tienes las cotizaciones para un proceso de compra en formato JSON:
  #{quotations_data.to_json}

  La intención del usuario es: "#{user_intent}".

  ### Instrucciones para el análisis:
  - Si solo hay una cotización en los datos, devuelve la respuesta con etiquetas <strong> para resaltar puntos importantes y cualquier otra etiqueta para organizar y que se vea mejor la data,
  - Si solo hay una cotización en los datos, no se puede realizar un análisis comparativo. En ese caso, devuelve el siguiente mensaje: "<strong>No se puede realizar un análisis</strong> Solo hay una cotización disponible.</strong>".
  - Si hay más de una cotización, realiza las siguientes acciones:
    - Analiza esta data teniendo presente lo que quiere hacer el usuario con ella en la intención del usuario.
    - Además, devuelve la respuesta en formato HTML con etiquetas <strong> para resaltar puntos importantes y <ul> <li> para listar las ventajas o desventajas de cada cotización y cualquier otra etiqueta para organizar y que se vea mejor la data.
    - Además formatea los precios, ponles el formato correcto teniendo cuidado con las decimales y teniendo presente su moneda.
    - Ten presente, nunca incluyas enlaces en la respuesta y asegura que el formato sea correcto para HTML.
    - Las respuestas dadas siempre en español.
    - No pongas en la respuesta el funcionamiento de la estructura y arquitectura del data.
  PROMPT

  response = post('/chat/completions',
    headers: {
      'Content-Type' => 'application/json',
      'Authorization' => "Bearer #{ENV['CHATGPT_API_KEY']}"
    },
    body: {
      model: 'gpt-4o-mini', # Puedes cambiar el modelo si lo deseas
      messages: [{ role: 'user', content: prompt }],
      max_tokens: 2000
    }.to_json
  )

  # Devuelve la respuesta generada por ChatGPT
  response.parsed_response['choices'][0]['message']['content'] rescue "Lo siento, hubo un problema con el análisis."
end
```

Explicación del Flujo de analyze_ask

1. **Preparación del Prompt:** Se crea un prompt que contiene los datos de cotización en JSON y la intención del usuario, como “ordenar de mayor a menor precio” o “seleccionar la cotización más favorable”.
2. **Envío de la solicitud:** La petición se realiza mediante una llamada **POST** a la API de OpenAI.
3. **Manejo de la Respuesta:** La respuesta es formateada en HTML para mostrarla en la interfaz.

6. Interfaz del Usuario (General.js.erb)

La integración en la vista se realiza con código JavaScript (**General.js.erb**). La interacción principal es a través del método **chatgpt_ask**, que se ejecuta al presionar el botón de “Enviar” o al presionar la tecla “Enter”. Este método realiza lo siguiente:

1. **Enviar Petición HTTP:** Hace una petición **POST** al endpoint del controlador (**chatgpt_ask_landing_api_v1_chatgpt_index_url**) enviando el **prompt** del usuario.
2. **Mostrar Respuesta:** Recibe la respuesta del servidor y la renderiza en la interfaz usando **ng-bind-html** para mostrar el HTML generado por la API.

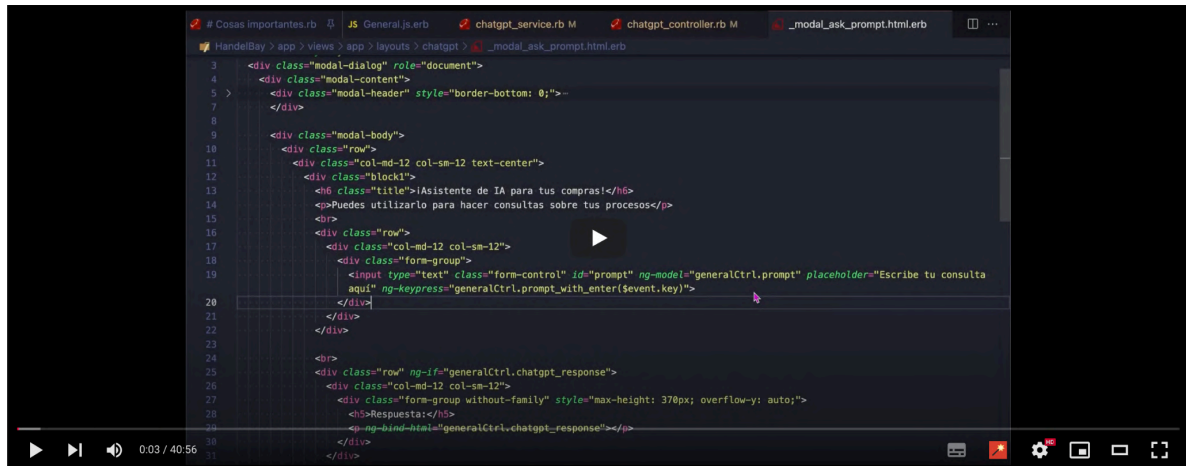
```
this.chatgpt_ask = function(prompt) {  
  this.show_loader_AI = true;  
  
  let params = {  
    user_email: self.user_email,  
    user_token: self.user_token,  
    prompt: prompt  
  }  
  
  $http.post("<%= chatgpt_ask_landing_api_v1_chatgpt_index_url(format: :json) %>", params)  
  .success(function(data, status) {  
    self.chatgpt_response = $sce.trustAsHtml(data.response);  
    self.show_loader_AI = false;  
  })  
  .error(function(data, status) {  
    self.chatgpt_response = "<p>Error: No se pudo procesar tu solicitud.</p>";  
    self.show_loader_AI = false;  
  })  
}
```


9. Mejoras y Buenas Prácticas

- **Manejo de Errores:** Es importante mejorar el manejo de errores en caso de fallos en la API, asegurando que los usuarios reciban mensajes adecuados si algo sale mal.
- **Validación del prompt:** Se recomienda validar el **prompt** en el lado del cliente antes de enviarlo para evitar solicitudes innecesarias.
- **Control de Tokens:** Asegúrate de que el token de la API esté almacenado de manera segura, evitando riesgos de exposición.
- **Optimizar Conversaciones:** Para evitar problemas de contexto, se puede almacenar el historial de mensajes de forma controlada para tener un mejor contexto de la conversación.

ANEXO G MANUAL TÉCNICO (VIDEO EXPLICATIVO)

Enlace: <https://youtu.be/KtDMbAzRuz4>



The image shows a video player interface with a code editor overlay. The code is written in ERB (Embedded Ruby) and defines a modal dialog for asking a prompt. The code includes the following elements:

```
3 <div class="modal-dialog" role="document">
4 <div class="modal-content">
5 <div class="modal-header" style="border-bottom: 0;">
6 </div>
7
8
9 <div class="modal-body">
10 <div class="row">
11 <div class="col-md-12 col-sm-12 text-center">
12 <div class="block1">
13 <h4 class="title">¡Asistente de IA para tus compras!</h4>
14 <p>Puedes utilizarlo para hacer consultas sobre tus procesos</p>
15 <br>
16 <div class="row">
17 <div class="col-md-12 col-sm-12">
18 <div class="form-group">
19 <input type="text" class="form-control" id="prompt" ng-model="generalCtrl.prompt" placeholder="Escribe tu consulta
aquí" ng-keypress="generalCtrl.prompt_with_enter($event.key)">
20 </div>
21 </div>
22 </div>
23
24 <br>
25 <div class="row ng-if="generalCtrl.chatgpt_response">
26 <div class="col-md-12 col-sm-12">
27 <div class="form-group without-family" style="max-height: 378px; overflow-y: auto;">
28 <div class="Resposta">
29 <ng-bind-html="generalCtrl.chatgpt_response"></div>
30 </div>
31 </div>
```

ANEXO H COMPARATIVO DE INDICADORES DE EFICIENCIA

Objetivo del caso de uso	Desarrollo empleado	Acciones para cumplir objetivo	Cantidad de clics	Tiempo promedio	Cantidad de consultas SQL
Consultar información general de un proceso de compra	Sin integración con IA	<ul style="list-style-type: none"> ● Iniciar sesión ● Pestaña "Procesos" ● Pestaña "En Revisión" ● Uso de buscador ● Selección de proceso 	5	200 seg.	<ul style="list-style-type: none"> - Obtener datos de procesos. - Búsqueda de proceso específico. - Traer datos de "requisition". - Traer datos de "purchase". <p style="text-align: center;">Total: 4 consultas</p>
	Integrando a GPT-4o mini	<ul style="list-style-type: none"> ● Iniciar sesión ● Botón "Asistente virtual" ● Botón "Enviar Consulta" 	3	30 seg.	<ul style="list-style-type: none"> - Consulta información de "requisition". - Consulta de "purchase" <p style="text-align: center;">Total: 2 consultas</p>
Análisis de indicadores de eficiencia			<p>Con el uso de la funcionalidad que integra la IA, la cantidad de clics se reduce en un 40% además de permitir que el usuario permanezca en la misma interfaz; el tiempo estimado para cumplir este objetivo se reduce en un 85%, y a nivel técnico, el sistema actúa de manera más óptima minimizando la cantidad de consultas internas SQL en un 50%.</p>		
Verificar estado de respuesta de 3 proveedores	Sin integración con IA	<ul style="list-style-type: none"> ● Iniciar sesión ● Pestaña "Procesos" ● Pestaña "En Revisión" ● Selección de 	8	280 seg.	<ul style="list-style-type: none"> - Obtener datos de procesos. - Búsqueda de proceso específico. - Traer datos de

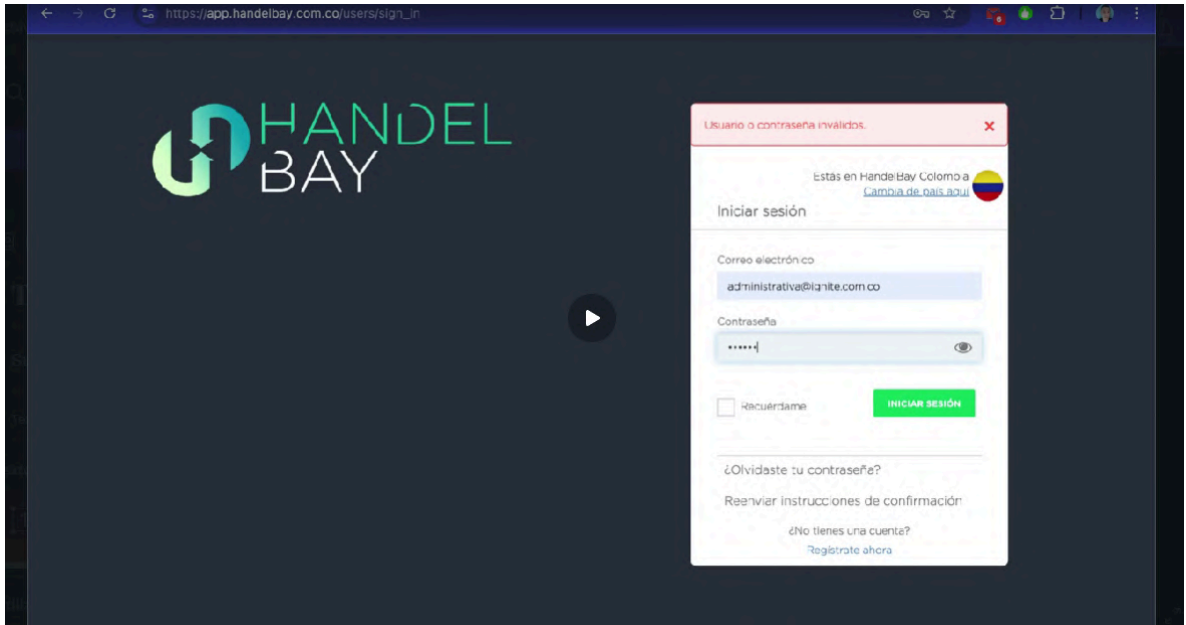
de un proceso de compra		<ul style="list-style-type: none"> proceso Pestaña “Invitados” Botón “Ver más” (Proveedor 1) Botón “Ver más” (Proveedor 2) Botón “Ver más” (Proveedor 3) 			“requisition”. - Traer datos de “purchase”. - Traer datos de “invitados”. Total: 5 consultas
	Integrando a GPT-4o mini	<ul style="list-style-type: none"> Iniciar sesión Botón “Asistente virtual” Botón “Enviar Consulta” 	3	30 seg.	- Consulta de información de “invitados”. - Consulta de “compañía”. - Consulta de “cotizaciones”. Total: 3 consultas
Análisis de indicadores de eficiencia			Con el uso de la funcionalidad que integra la IA, la cantidad de clics se reduce en un 62,5% además de permitir que el usuario permanezca en la misma interfaz; el tiempo estimado para cumplir el objetivo se reduce en un 89,29%, y a nivel técnico, el sistema actúa de manera más óptima minimizando la cantidad de consultas internas SQL en un 40%.		
Comparar y analizar 3 cotizaciones de un proceso de compra	Sin integración con IA	<ul style="list-style-type: none"> Iniciar sesión Pestaña “Procesos” Pestaña “En Revisión” Selección de proceso Pestaña “Cotizaciones” Botón “Ver Cotización” (Proveedor 1) Botón “Volver” Botón “Ver Cotización” 	10	420 seg.	- Obtener datos de procesos. - Búsqueda de proceso específico. - Traer datos de “requisition”. - Traer datos de “purchase”. - Traer datos de “cotizaciones”. - Entrar a cada cotización. Total:

		(Proveedor 2) ● Botón “Volver” ● Botón “Ver Cotización” (Proveedor 3)			9 consultas
	Integrando a GPT-4o mini	● Iniciar sesión ● Botón “Asistente virtual” ● Botón “Enviar Consulta”	3	45 seg.	- Consulta para traer toda la data de “cotizaciones”. - Consulta productos de cada cotización. Total: 2 consultas
Análisis de indicadores de eficiencia			Con el uso de la funcionalidad que integra la IA, la cantidad de clics se reduce en un 70% además de permitir que el usuario permanezca en la misma interfaz; el tiempo estimado para cumplir el objetivo se reduce en un 89,29%, y a nivel técnico, el sistema actúa de manera más óptima minimizando la cantidad de consultas internas SQL en un 77%.		

ANEXO I

DEMO DE LA INTEGRACIÓN EN PRODUCCIÓN

Enlace: <https://youtu.be/QxrGdgcpoxg>



ANEXO J

RESUMEN ANALÍTICO

Título del proyecto	Integración del modelo de inteligencia artificial GPT-4o mini con el software de compras HandelBay
Autor(es)	VALENCIA HENAO, Juan Camilo camilovalencia1229@gmail.com RICO PÉREZ, Maikol mricoperez9@gmail.com
Presidente	BETANCOURT CORREA, Carlos cbc@umanizales.edu.co
Tipo de documento	Trabajo de grado
Referencia documento	VALENCIA HENAO, Juan Camilo y RICO PÉREZ, Maikol. Integración del modelo de inteligencia artificial GPT-4o mini con el software de compras HandelBay, 2024, 85 p. Trabajo de grado (Ingeniero de Sistemas y Telecomunicaciones). Universidad de Manizales. Facultad de Ciencias e Ingeniería. Programa Ingeniería de Sistemas y Telecomunicaciones
Institución	Programa Ingeniería de Sistemas y Telecomunicaciones, Facultad de Ingeniería, Universidad de Manizales.
Palabras claves	Inteligencia Artificial, GPT-4o mini, Gestión de Compras, Automatización, HandelBay.
Descripción	En este proyecto se propone el diseño y desarrollo de una integración entre el modelo de inteligencia artificial GPT-4o mini y el software de compras colombiano HandelBay, con el objetivo de optimizar las tareas de

sus usuarios, permitiéndoles interactuar con una interfaz para realizar preguntas comunes sobre su gestión de compras y proveedores, y obteniendo respuestas precisas y rápidas a través de consultas a la base de datos del sistema. La investigación busca mejorar la eficiencia en las operaciones del área de compras, minimizando el tiempo y el esfuerzo requeridos para analizar múltiples cotizaciones y seleccionar proveedores. Además, se espera que la automatización de estos procesos reduzca los costos asociados a la contratación de analistas y practicantes para realizar estas tareas. Asimismo, este desarrollo pretende establecer un precedente en el uso de tecnologías avanzadas como la inteligencia artificial en el ámbito empresarial colombiano, promoviendo la adopción de soluciones tecnológicas que optimicen los procesos internos y mejoren la toma de decisiones. Aunque esta integración beneficia inicialmente a la empresa específica que utilice HandelBay, la expectativa a futuro es que sirva de referencia para otras organizaciones interesadas en la modernización y eficiencia de sus áreas de compras..

Fuentes

ARMAN, Md y LAMIYAR, Umama Rashid. Exploring the Implication of ChatGPT AI for Business: Efficiency and Challenges. En: Journal of Innovation Information Technology and Application. Vol. 5 No. 1 (2023).

BRYANT, Steven. Assessing GPT-4's Role as a Co-Collaborator in Scientific Research: A Case Study Analyzing Einstein's Special Theory of Relativity. [En línea] En: Discov Artif Intell 3, 26 (2023).

SALLAM, Malik. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. En: Healthcare 11, no. 6: 887 (2023).

Contenido

Introducción

1. Área Problemática
 2. Objetivos
 3. Justificación
 4. Marco Teórico (Marco Conceptual, Referente Normativo y Legal, y Antecedentes)
 5. Metodología (Tipo de trabajo y Procedimiento)
 6. Resultados
 7. Conclusiones
 8. Recomendaciones
- Referencias Bibliográficas
Anexos

Metodología

El proyecto se clasifica como una investigación aplicada en desarrollo tecnológico, basada en la metodología Waterfall o Cascada, cuyo objetivo fue desarrollar e implementar una integración entre el modelo de inteligencia artificial GPT-4o mini y el software de HandelBay para optimizar la consulta de información por parte de los usuarios, agilizando su gestión de compras y selección de proveedores. Para el procedimiento se inició con el análisis de la necesidad y de los modelos de GPT, siguiendo con la creación de la lógica de negocio para la integración entre ambos desarrollos tecnológicos revisando casos de uso, modelo relacional con el usuario, diseño de consultas e interfaz, para luego configurarlo usando el framework Ruby On Rails. Posteriormente se creó un servicio, se conectó con controladores y rutas, se construyó la interfaz para los usuarios y se establecieron los roles y permisos correspondientes; en las últimas fases se realizaron las pruebas pertinentes de rendimiento y eficiencia, así como la creación de la documentación técnica y el manual de uso.

Conclusiones

El proyecto de integración del modelo GPT-4o mini con el software de compras HandelBay ha demostrado ser una solución eficiente, cumpliendo con los objetivos propuestos en términos de funcionalidad, rendimiento y optimización de tiempos. La selección del modelo adecuado y la implementación de consultas predefinidas permitió optimizar el uso de tokens y solicitudes, lo que fue clave para mejorar la experiencia del usuario. No obstante, el esquema utilizado, aunque eficiente, presenta limitaciones en cuanto a la flexibilidad de las consultas, lo que requeriría ajustes si se desea ampliar el alcance de las preguntas que los usuarios pueden hacer.

La integración técnica del modelo GPT-4o mini con HandelBay fue robusta y exitosa, facilitando una interacción fluida entre el modelo de inteligencia artificial y la base de datos. El desarrollo de la API y las rutas garantiza una comunicación eficiente, validando correctamente las solicitudes del usuario y asegurando que las respuestas sean precisas. Vale resaltar que esta funcionalidad ya ha sido validada por la empresa y se encuentra en ambiente de producción disponible para los usuarios del sistema.

La interfaz de usuario, basada en los mockups iniciales, resultó clara y funcional, proporcionando una experiencia intuitiva para el usuario final. Las mejoras visuales como el loader de carga y las validaciones de consultas enriquecieron la experiencia, asegurando que las interacciones fueran eficientes sin interferir con el flujo de trabajo diario de los usuarios.

Posterior a la investigación, se llega a la conclusión de que el sistema desarrollado mejora la eficiencia de los usuarios en el manejo de los procesos de compra en HandelBay, buscando al mismo tiempo garantizar la privacidad y seguridad de los datos a través de una

adecuada implementación de permisos. El sistema está bien alineado con las necesidades del negocio y ha demostrado ser una solución escalable que puede adaptarse a futuras mejoras.

Anexos

Anexo A - Casos de uso

Anexo B - Diagramas

Anexo C - Pruebas

Anexo D - Manual de uso

Anexo E - Manual de uso (video explicativo)

Anexo F - Manual técnico

Anexo G - Manual técnico (video explicativo)

Anexo H - Comparativo de indicadores de eficiencia

Anexo I - Demo de la integración en producción

Anexo J - Resumen Analítico