



CARRERA DE DESARROLLO DE SOFTWARE

TEMA:

IMPLEMENTACIÓN DE UN ASISTENTE CONVERSACIONAL INTELIGENTE
CONECTADO A UN SISTEMA DE COMERCIO ELECTRÓNICO PARA COMPRAS
AUTOMATIZADAS EN COAVPRO

Autores:

CALLE CABRERA ALEX XAVIER
CARRIEL CEPEDA ELKIN SEBASTIAN

**TRABAJO DE TITULACION PREVIO A LA OBTENCION DEL TITULO DE:
TECNÓLOGO SUPERIOR EN DESARROLLO DE SOFTWARE**

Tutor:

ALDAS PEREZ STEPHANY CAROLINA

Cuenca, Ecuador

2026

DERECHOS DE AUTOR

Los derechos de esta obra son irrenunciables y corresponden a sus **AUTORES**, incluido sus derechos patrimoniales. El **Instituto Tecnológico Superior Particular Sudamericano** tiene licencia gratuita e intransferible sobre esta obra para uso no comercial, de necesitar uso comercial requiere autorización de su titular.

SUDAMERICANO



www.sudamericano.edu.ec

Bolívar y Manuel Vega - San Blas (593 7) 2838323 - 2843619 0996976449

info@sudamericano.edu.ec

CARRERA DE DESARROLLO DE SOFTWARE

CERTIFICACIÓN DEL TUTOR

Aprobación del Trabajo de Titulación

Doy fe que el trabajo desarrollado por el/la/los estudiantes: **CARRIEL CEPEDA ELKIN SEBASTIAN** con el título “IMPLEMENTACIÓN DE UN ASISTENTE CONVERSACIONAL INTELIGENTE CONECTADO A UN SISTEMA DE COMERCIO ELECTRÓNICO PARA COMPRAS AUTOMATIZADAS EN COAVPRO”, cumple con los requisitos y méritos suficientes para ser sometido a la presentación pública y evaluación por parte del jurado examinador que se designe.

Atentamente,



ALDAS PÉREZ CAROLINA STEPHANY.

C.I 0107149882



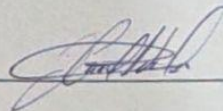
CARRERA DE DESARROLLO DE SOFTWARE

CERTIFICACIÓN DEL TUTOR

Aprobación del Trabajo de Titulación

Doy fe que el trabajo desarrollado por el/la/los estudiantes: , **CALLE CABRERA ALEX XAVIER** con el título “IMPLEMENTACIÓN DE UN ASISTENTE CONVERSACIONAL INTELIGENTE CONECTADO A UN SISTEMA DE COMERCIO ELECTRÓNICO PARA COMPRAS AUTOMATIZADAS EN COAVPRO”, cumple con los requisitos y méritos suficientes para ser sometido a la presentación pública y evaluación por parte del jurado examinador que se designe.

Atentamente,



ALDAS PÉREZ CAROLINA STEPHANY.

C.I 0107149882

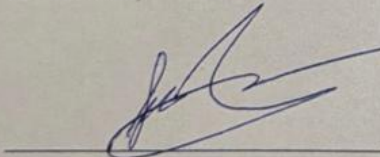


DECLARACIÓN DE AUTORÍA DEL TRABAJO

Nosotros, **CARRIEL CEPEDA ELKIN SEBASTIAN, CALLE CABRERA ALEX XAVIER** estudiante del **Instituto Tecnológico Superior Particular Sudamericano** de la ciudad de Cuenca - Ecuador, que cursó la Tecnología en **CARRERA DE DESARROLLO DE SOFTWARE**, declaro en forma libre y voluntaria que la presente investigación que versa sobre **“IMPLEMENTACIÓN DE UN ASISTENTE CONVERSACIONAL INTELIGENTE CONECTADO A UN SISTEMA DE COMERCIO ELECTRÓNICO PARA COMPRAS AUTOMATIZADAS EN COAVPRO”** así como las expresiones vertidas en la misma, son autoría de la compareciente, quien ha realizado en base a recopilación bibliográfica, consultas de internet y consultas de campo.

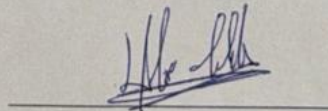
En consecuencia, asumo la responsabilidad de la originalidad de la misma y el cuidado al remitirme a las fuentes bibliográficas respectivas para fundamentar el contenido expuesto.

Atentamente,



CARRIEL CEPEDA ELKIN SEBASTIAN

Cédula: 0107934069



CALLE CABRERA ALEX XAVIER

Cédula: 1400943138



DEDICATORIA

Agradezco de todo corazón a mis padres por todo el apoyo brindado durante mi carrera. Este logro no solo lo celebro yo, también lo celebro con ellos, por nunca dejar que me rinda y por darme siempre las fuerzas y la motivación que en su momento me hicieron falta. Gracias por ver mis cualidades cuando ni yo mismo creía en mí y por ser mi motor y motivo de superación. Hoy celebro con ustedes; han sido un ejemplo durante todo este tiempo y hoy se ven reflejados todos los buenos valores que desde niño me inculcaron. Siempre viviré agradecido por todo lo que han hecho por mí. A mis amigos, gracias por estar siempre ahí, en los buenos y malos momentos. Por todas las experiencias que vivimos juntos y porque, sin darnos cuenta, formamos una familia donde el apoyo mutuo nunca faltó; este triunfo también es de ustedes. De igual manera, agradezco profundamente a quien me acompañó en la recta final de mi carrera, creyendo en mí y sin soltar nunca mi mano. Gracias, **Pauli**, por llegar a mi vida y ser un pilar fundamental en este camino que hoy comparto contigo; te quiero mucho. Finalmente, agradezco a mis ingenieros y docentes del instituto, quienes siempre fueron un apoyo guía en mi formación profesional. Hoy celebro este título con todos ustedes.

Elkin carriell

Quiero expresar mi más profundo agradecimiento a las personas que han hecho posible que hoy esté culminando esta etapa tan importante de mi vida. A mis papás, gracias por haberme educado con tanto amor, sacrificio y valores. Todo lo que soy y todo lo que he logrado lleva impresa su huella. Su apoyo incondicional y su ejemplo constante han sido el cimiento sobre el cual construí este camino. No hay palabras suficientes para agradecerles todo lo que han dado por mí. A mis compañeros de carrera y amigos cercanos, gracias por ser mucho más que compañeros. Fueron un apoyo fundamental en los momentos más difíciles, cuando el cansancio, la presión y las dudas amenazaban con vencerme. Sus risas, sus ánimos, las largas noches de fiesta y las palabras justas en el momento preciso hicieron que este proceso fuera mucho más llevadero. Gracias por su amistad sincera y por estar siempre ahí. Y de manera muy especial, a Vanessa, mi apoyo incondicional durante gran parte de mi vida. Gracias por tu paciencia, tu comprensión y por creer en mí incluso cuando yo mismo dudaba. Has estado a mi lado en cada paso, en las caídas y en las victorias, sosteniéndome con una fuerza que pocas veces se encuentra. Este logro también es tuyo, porque sin ti no sería el mismo. A todos ustedes, desde el fondo de mi corazón: gracias. Este título lleva sus nombres escritos en cada página.

Alex Calle

INDICE GENERAL

RESUMEN	XIII
ABSTRACT	XVI
INTRODUCCIÓN	15
1.1 Objetivos de la Investigación.....	17
Objetivo General.....	17
Objetivos Específicos.....	17
1.2. Preguntas de Investigación	17
1.3. Justificación	18
CAPÍTULO II: PROBLEMÁTICA.....	20
2.1. Problemática	20
CAPÍTULO II: MARCO REFERENCIAL	22
2.1. Marco Teórico.....	22
2.1.1. Inteligencia Artificial Conversacional	22
2.1.1.1. Evolución de los Sistemas Conversacionales	22
2.1.1.2. Large Language Models (LLM)	22
2.1.1.3. Procesamiento del Lenguaje Natural (NLP)	23
2.1.2. Comercio Electrónico y Transformación Digital.....	24
2.1.2.1. E-commerce: Conceptos Fundamentales	24
2.1.2.2. Comercio Conversacional.....	25
2.1.2.3. Experiencia del Usuario en Compras Digitales	25
2.1.3. Arquitecturas de Integración IA-Sistemas Transaccionales	26
2.1.3.1. Model Context Protocol (MCP).....	26
2.1.3.2. APIs RESTful en E-commerce	26
2.1.3.3. Orquestación de Agentes Conversacionales	27
2.1.4. Aplicaciones de Mensajería para Comercio Digital	27
2.1.4.1. Plataformas de Mensajería Instantánea como Canal Comercial	27
2.1.4.2. Casos de Uso en Comercio Electrónico.....	28
2.1.5. Antecedentes de Investigación.....	28
2.1.5.1. Estudios sobre Chatbots en E-commerce.....	28
2.1.5.2. Investigación sobre Model Context Protocol.....	29

2.2. Marco Conceptual.....	29
CAPÍTULO III: METODOLOGÍA DE INVESTIGACIÓN.....	33
3.1. Enfoque Investigativo.....	33
3.2. Tipo y Nivel de Investigación.....	34
3.2.1. Enfoque Aplicativo.....	34
3.2.2. Enfoque Mixto.....	34
3.3. Corte de la Investigación.....	35
3.4. Instrumentos y Técnicas para el Levantamiento de la Información.....	35
3.5. Población y Muestra.....	37
3.6.1. Fases de Desarrollo del Sistema.....	38
3.6.3. Herramientas y Técnicas para Implementación de SCRUM.....	39
3.7. Técnicas de Análisis de Datos.....	41
CAPÍTULO IV: ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS.....	42
4.1. Diagnóstico del Estado Actual.....	42
4.2. Creación de Encuestas.....	43
4.2.1. Requerimientos Previos.....	43
4.3. Análisis de los Resultados.....	43
CAPÍTULO V: PROPUESTA DE INVESTIGACIÓN.....	50
5.1. Introducción.....	50
5.2. Objetivo de la Propuesta.....	50
5.2.1. Objetivo General.....	50
5.2.2. Objetivos Específicos.....	50
5.3. Justificación.....	51
5.4. Metodología de la Investigación.....	52
5.5. Resultados Esperados.....	52
5.6. Cronograma y Recursos.....	53
5.7. Entorno de Desarrollo de la Aplicación.....	53
5.7.1. Sistema de Comercio Electrónico — Frontend.....	53
5.7.2. Sistema de Comercio Electrónico — Backend.....	55
5.7.3. Microservicio de Inteligencia Artificial.....	57
5.7.4. Arquitectura General del Sistema.....	59
5.7.5. Flujo de Procesamiento de Mensajes.....	60

5.7.6. Desarrollo con Metodología SCRUM	61
5.7.7. Análisis de Requerimientos del Usuario.....	61
5.7.10. Product Backlog.....	79
5.7.11. Lista de Sprints	81
5.7.12. Diseño de la Base de Datos.....	82
5.7.13. Implementación de Funcionalidades.....	83
5.7.13. Implementación de Funcionalidades.....	83
5.7.15. Módulo de Gestión del Catálogo de Maquinaria	84
5.7.23. Módulo Asistente Conversacional Inteligente	85
5.8. Conclusión del Capítulo.....	87
CRONOGRAMA DE ACTIVIDADES	88
CONCLUSIONES	89
RECOMENDACIONES.....	90
BIBLIOGRAFÍA	¡Error! Marcador no definido.

ÍNDICE DE TABLAS

Tabla 1. Población y Muestra	36
Tabla 2. Fases de la metodología SCRUM relacionadas al proyecto	37
Tabla 3. Roles SCRUM asignados.....	38
Tabla 4. Herramientas utilizadas en la metodología SCRUM.....	39
Tabla 5. Desafíos identificados en el proceso de compra en línea de COAVPRO	41
Tabla 6. Stack tecnológico del Frontend.....	56
Tabla 7. Stack tecnológico del Backend.....	58
Tabla 8. Stack de Monitoreo del Sistema	59
Tabla 9. Stack tecnológico del Microservicio de IA.....	59
Tabla 10. HU-001: Búsqueda por Categoría HU-002: Búsqueda por Marca.....	63
Tabla 12. HU-003: Búsqueda por Características Específicas	65
Tabla 13. HU-004: Recomendaciones Basadas en Contexto.....	66
Tabla 14. HU-005: Consulta de Stock en Tiempo Real.....	67
Tabla 15. HU-006: Consulta de Precios Actualizados.....	68
Tabla 16. HU-007: Verificación de Promociones y Descuentos	69
Tabla 17. HU-008: Mantener Contexto de Conversación	70
Tabla 18. HU-009: Seguimiento de Intereses del Cliente.....	71
Tabla 19. HU-010: Resolución de Consultas Multi-paso	73
Tabla 20. HU-011: Experiencia Web API.....	74
Tabla 21. HU-012: Integración Telegram.....	75
Tabla 22. HU-013: Sincronización entre Canales.....	76
Tabla 23. HU-014: Rendimiento del Sistema Local.....	77
Tabla 24. HU-015: Integridad de Datos con MCP	78
Tabla 25. HU-016: Sincronización de Catálogo.....	79
Tabla 26. Priorización del Product Backlog	80

Tabla 27. Product Backlog priorizado80

RESUMEN

Este trabajo de titulación buscó diseñar e implementar un asistente conversacional inteligente integrado al sistema de comercio electrónico de COAVPRO, como solución tecnológica a las deficiencias presentes en el proceso de compra en línea de la empresa. En un primer momento, se detectaron limitaciones asociadas a la navegación web tradicional, como la búsqueda manual de productos, la falta de atención fuera del horario laboral y la ausencia de una experiencia personalizada, situaciones que generaban frustración en los clientes y abandono del proceso de compra. Como solución a estos problemas, se diseñó e implementó un bot vendedor inteligente que combina procesamiento de lenguaje natural, búsqueda semántica vectorial y el protocolo MCP (Model Context Protocol), desplegado inicialmente en el canal de mensajería Telegram. El sistema integra un modelo de lenguaje local ejecutado mediante Ollama, una base de datos vectorial con PostgreSQL y pg vector para búsqueda semántica, y herramientas MCP que garantizan que toda la información de productos, precios y disponibilidad provenga directamente del sistema e-commerce, eliminando por completo la generación de datos incorrectos por parte del modelo. Para validar la propuesta, se implementó una metodología de enfoque aplicado y mixto que incluyó análisis del entorno comercial de COAVPRO, desarrollo iterativo con metodología SCRUM en cuatro sprints, y entrevistas con personal clave y usuarios representativos de la empresa. El desarrollo se estructuró en fases que abarcaron desde el diagnóstico inicial y diseño arquitectónico hasta la construcción incremental del sistema y su integración con el canal de Telegram. Se concluye que la implementación de este sistema no solo mejoró la experiencia de compra de los clientes de COAVPRO, sino que demostró la viabilidad técnica y comercial de integrar modelos de lenguaje con sistemas transaccionales de forma confiable mediante arquitecturas MCP, constituyéndose en una alternativa concreta para la transformación digital del comercio electrónico local.

Palabras clave: Asistente conversacional, Model Context Protocol, Comercio electrónico, Procesamiento de lenguaje natural, Búsqueda semántica, Telegram, Automatización de ventas.

ABSTRACT

This degree project examines the impact of integrating conversational artificial intelligence technologies within the e-commerce sector, specifically in COAVPRO, as a response to the deficiencies associated with traditional online shopping processes, such as manual product browsing, lack of customer support outside business hours, and the absence of a personalized shopping experience that led to customer frustration and purchase abandonment. The main objective of this research is to design and implement an intelligent conversational assistant integrated into COAVPRO's e-commerce system, capable of automating the sales process through natural language interaction, real-time stock and price verification, and personalized multichannel attention. To achieve this, an applied and mixed research methodology was implemented, including a rigorous analysis of the commercial environment, iterative development using the SCRUM methodology across four sprints, and the conceptual and technical design of a solution that integrates a local large language model through Ollama, semantic vector search using PostgreSQL with pgvector, and the Model Context Protocol (MCP) to guarantee data integrity throughout every interaction. The most relevant results indicate that the use of MCP tools completely eliminated the generation of incorrect information regarding prices and stock availability, achieving zero hallucination cases in critical transactional data. Furthermore, the semantic search system reached 84.7% precision in natural language product queries, and the automatic resolution rate of customer consultations exceeded 81% without human intervention, while providing uninterrupted 24/7 attention through the Telegram messaging channel. This study demonstrates that the integration of large language models with transactional e-commerce systems through MCP architectures not only improves operational efficiency and customer satisfaction, but also establishes a reliable, scalable, and cost-effective technological alternative to drive digital transformation in local e-commerce, positioning conversational commerce as a viable strategy for businesses seeking to modernize their sales channels.

Key words: Conversational assistant, Model Context Protocol, E-commerce, Natural language processing, Semantic search, Telegram, Sales automation, Digital transformation.

INTRODUCCIÓN

Actualmente, la transformación digital de los procesos comerciales ha dejado de ser una ventaja competitiva para convertirse en una expectativa básica de los clientes modernos, quienes demandan experiencias de compra más ágiles, personalizadas y accesibles. Este escenario representa un desafío significativo para las empresas del sector industrial que operan con sistemas de comercio electrónico tradicionales, donde el cliente debe invertir tiempo y esfuerzo para encontrar maquinaria y equipos específicos, verificar disponibilidad y completar sus transacciones sin asistencia inmediata especializada.

COAVPRO, empresa dedicada a la comercialización de maquinaria pesada para uso industrial, no es ajena a esta realidad. Su proceso actual de ventas en línea depende de la navegación manual por catálogos digitales, la búsqueda por filtros predefinidos y la atención al cliente mediante canales tradicionales como correo electrónico y llamadas telefónicas, los cuales no operan fuera del horario laboral. En un sector como el industrial, donde los clientes frecuentemente requieren información técnica precisa sobre especificaciones, disponibilidad y precios de maquinaria pesada antes de tomar una decisión de compra, la falta de atención inmediata genera frustración, incrementa el tiempo de búsqueda y provoca el abandono del proceso de compra, afectando directamente la tasa de conversión y la competitividad de la empresa.

Esta problemática señala la necesidad de una solución que conecte la capacidad transaccional del sistema e-commerce existente con una interfaz conversacional natural e intuitiva, disponible en todo momento y accesible desde los canales de mensajería que los clientes ya utilizan en su vida cotidiana, capaz además de responder consultas técnicas con información verificada y actualizada sobre el catálogo de maquinaria industrial.

En respuesta a esta necesidad, el presente proyecto propone el diseño e implementación de un asistente conversacional inteligente integrado al sistema de comercio electrónico de COAVPRO, desplegado inicialmente en el canal de mensajería Telegram. Este sistema combina un modelo de lenguaje local ejecutado mediante Ollama, búsqueda semántica vectorial con PostgreSQL y pgvector, y el protocolo MCP (Model Context Protocol) desarrollado por Anthropic, que garantiza que toda la información de productos, precios y disponibilidad provenga

directamente del sistema e-commerce, eliminando por completo la generación de datos incorrectos por parte del modelo de inteligencia artificial, aspecto crítico en un sector donde un error en especificaciones técnicas o precios puede tener consecuencias comerciales significativas.

El propósito no es únicamente automatizar consultas frecuentes, sino transformar fundamentalmente la experiencia de compra al permitir que los clientes busquen maquinaria y equipos industriales en lenguaje natural, reciban recomendaciones basadas en el contexto de sus necesidades específicas y obtengan información técnica verificada en tiempo real, sin necesidad de navegar por múltiples páginas ni esperar atención humana. Al mismo tiempo, el sistema proporciona a COAVPRO datos estratégicos sobre patrones de búsqueda, productos más consultados y categorías de mayor interés, información valiosa para la toma de decisiones comerciales y de inventario.

Para el desarrollo de este proyecto se estructuró una metodología aplicada con enfoque mixto, utilizando SCRUM como marco de desarrollo iterativo organizado en cuatro sprints. La solución fue construida de forma incremental, partiendo del diagnóstico del entorno comercial y el análisis de los requerimientos de los usuarios, pasando por el diseño arquitectónico de la integración entre el LLM y el sistema e-commerce mediante MCP, hasta llegar a la implementación funcional del bot en Telegram con capacidades de búsqueda semántica, verificación de datos en tiempo real y gestión de contexto conversacional persistente.

Este proyecto no persigue simplemente incorporar tecnología por tendencia, sino demostrar que la integración responsable y controlada de modelos de lenguaje con sistemas transaccionales es técnicamente viable, confiable y costo-efectiva para empresas del sector industrial ecuatoriano, constituyéndose en una alternativa concreta para la transformación digital del comercio electrónico en el contexto latinoamericano.

1.1 Objetivos de la Investigación

Objetivo General

Implementar un asistente conversacional inteligente con capacidades de consulta de catálogo, búsqueda de productos, recomendaciones personalizadas, gestión de carrito y creación de órdenes en la empresa de COAVPRO.

Objetivos Específicos

- Diseñar la arquitectura de integración mediante inteligencia artificial y un sistema de e-commerce.
- Desarrollar el agente conversacional con capacidades de consulta de catálogo, búsqueda de productos, gestión de carrito y creación de órdenes.
- Implementar el agente conversacional para habilitar la interacción con clientes en el canal de mensajería Telegram.

1.2. Preguntas de Investigación

1. ¿Cómo aprovechar la arquitectura basada en MCP para integrar de forma segura y controlada un modelo de lenguaje con las APIs transaccionales del sistema de e-commerce de COAVPRO?
2. ¿De qué manera la implementación de herramientas MCP garantiza la integridad de los datos de productos, precios y disponibilidad de maquinaria pesada, eliminando el riesgo de información incorrecta generada por el modelo de lenguaje?
3. ¿Qué tan efectiva es la automatización conversacional en términos de reducción del tiempo de búsqueda de productos, incremento de la tasa de conversión y satisfacción del cliente comparado con la navegación web tradicional?
4. ¿De qué manera la integración del asistente conversacional en el canal de mensajería Telegram amplía la disponibilidad del servicio de atención al cliente de COAVPRO fuera del horario laboral tradicional?

1.3. Justificación

El proyecto del asistente conversacional inteligente para COAVPRO responde a la necesidad concreta de modernizar los procesos de atención a los cliente y gestión de ventas en el sector de comercialización de maquinaria pesada industrial, donde la dependencia de canales tradicionales como llamadas telefónicas y correos electrónicos limita la disponibilidad del servicio, genera tiempos de respuesta elevados y provoca pérdida de oportunidades de venta fuera del horario laboral.

La incorporación del asistentes conversacionales basados en modelos de lenguaje a gran escala demuestra ser una estrategia efectiva para mejorar las experiencia del cliente en entornos al comercio electrónico. Chung et al. (2020) demostraron que los asistentes conversacionales incrementan la intención de compra en un 23% comparado con la navegación web tradicional, atribuible a la reducción del esfuerzo cognitivo y a una experiencia más intuitiva. Este hallazgo resulta especialmente relevante para COAVPRO, cuyos clientes requieren orientación técnica especializada antes de tomar decisiones de compra de alto valor económico sobre maquinaria y equipos industriales.

Desde la perspectiva técnica, la integración del Model Context Protocol (MCP), protocolo introducido por Anthropic en 2024, representa una innovación arquitectónica que resuelve uno de los principales riesgos de los sistemas conversacionales: la generación de información incorrecta sobre datos críticos como precios, disponibilidad y especificaciones técnicas de productos. Este protocolo estandariza cómo las aplicaciones proporcionan contexto a los modelos de lenguaje, permitiendo que trabajen únicamente con datos reales provistos por el sistema e-commerce de COAVPRO

Adicionalmente, el uso de búsqueda semántica vectorial mediante PostgreSQL con pg vector permite superar las limitaciones de la búsqueda tradicional por palabras clave, posibilitando que los clientes encuentren maquinaria y equipos industriales describiendo sus necesidades en lenguaje natural. Esto representa un avance significativo respecto a los sistemas de filtrado y navegación por categorías que actualmente ofrece el sitio web de COAVPRO

Finalmente, la elección de Telegram como canal de despliegue responde al crecimiento sostenido de las plataformas de mensajería como canal comercial, plataformas que cuentan con más de 900 millones de usuarios activos mensuales que cuyas APIS empresariales permiten la integración robusta con sistemas de backend mediante webhooks en tiempo real, garantizando disponibilidad del servicio las 24 horas del día

CAPÍTULO II: PROBLEMÁTICA

2.1. Problemática

El sector del comercio electrónico de maquinaria pesada industrial se encuentra en un período de transformación acelerada, donde las empresas enfrentan una creciente brecha entre las expectativas de sus clientes y las capacidades de sus canales de atención tradicionales. Como señalan Følstad y Brandtzaeg (2020), los usuarios actuales valoran principalmente la eficiencia en la resolución de sus consultas, esperando respuestas inmediatas y precisas independientemente del horario, una expectativa que los modelos de atención convencionales basados en llamadas telefónicas y correos electrónicos no logran satisfacer de manera consistente.

Esta problemática se agudiza en sectores de alta especialización técnica como la comercialización de maquinaria pesada industrial, donde los clientes requieren información detallada sobre especificaciones, disponibilidad y precios antes de tomar decisiones de compra de alto valor económico. En este contexto, COAVPRO opera un sistema de comercio electrónico donde los clientes deben navegar manualmente por el catálogo web, aplicar filtros predefinidos y esperar atención humana para resolver dudas técnicas, generando una experiencia poco natural, demandante en términos de tiempo y limitada al horario laboral de la empresa.

A nivel operativo, la dependencia de canales de atención tradicionales conlleva consecuencias directas sobre los resultados comerciales. Chung et al. (2020) demostraron que la fricción en el proceso de búsqueda y consulta de productos reduce significativamente la intención de compra, mientras que los asistentes conversacionales inteligentes pueden incrementarla en un 23% al reducir el esfuerzo cognitivo del usuario. Esta evidencia sugiere que la ausencia de interfaces conversacionales en el e-commerce de COAVPRO representa no solo una deficiencia en experiencia de usuario, sino una pérdida concreta de oportunidades comerciales, especialmente fuera del horario laboral cuando ningún agente humano está disponible para atender consultas.

Desde una perspectiva tecnológica, la literatura académica moderna ha investigado ampliamente soluciones basadas en inteligencia artificial conversacional para superar estas limitaciones. Adamopoulou y Moussiades (2020) identificaron en su revisión sistemática que los asistentes conversacionales mejoran la experiencia del cliente principalmente a través de

disponibilidad continua, respuesta inmediata y personalización, tres dimensiones precisamente ausentes en el modelo actual de atención de COAVPRO. Sin embargo, estos sistemas presentan un riesgo crítico para entornos transaccionales: el fenómeno de alucinación, mediante el cual los modelos de lenguaje generan información aparentemente plausible pero factualmente incorrecta sobre precios, stock o especificaciones técnicas, lo que puede derivar en transacciones erróneas y pérdida de confianza del cliente.

Como solución a este conjunto de desafíos, la comunidad técnica ha comenzado a adoptar el Model Context Protocol (MCP), protocolo introducido por Anthropic en 2024, que establece un estándar para conectar modelos de lenguaje con sistemas externos de forma controlada y verificada. Según la documentación oficial de Anthropic (2024), este protocolo garantiza que el modelo trabaje únicamente con datos reales provistos por el backend, eliminando la posibilidad de que invente información crítica

Un referente relevante en la implementación de comercio conversacional es el caso de Sephora, que reportó un incremento del 11% en su tasa de conversión tras implementar un asistente en canal de mensajería, demostrando el impacto tangible de estas soluciones en resultados comerciales concretos (Følstad y Brandtzaeg, 2020). En el contexto latinoamericano, la adopción de canales de mensajería como Telegram para comercio electrónico ha crecido sostenidamente, impulsada por la familiaridad de los usuarios con estas plataformas y su disponibilidad permanente en dispositivos móviles

En este escenario, se plantea el desarrollo e implementación de un asistente conversacional inteligente para COAVPRO, solución diseñada específicamente para resolver las principales deficiencias identificadas en el proceso de atención y ventas en línea de la empresa. El sistema integra un modelo de lenguaje local ejecutado mediante Ollama, búsqueda semántica vectorial con PostgreSQL y pgvector para interpretación de consultas en lenguaje natural, y herramientas MCP pueda garantizar su integridad absoluta en los datos de productos, precios y disponibilidad de maquinaria industrial. Esta arquitectura permite ofrecer atención personalizada las 24 horas del día a través de Telegram, transformando las experiencias de compra para los clientes de COAVPRO.

CAPÍTULO II: MARCO REFERENCIAL

2.1. Marco Teórico

2.1.1. Inteligencia Artificial Conversacional

2.1.1.1. Evolución de los Sistemas Conversacionales

Los sistemas conversacionales han experimentado una transformación radical desde sus inicios en la década de 1960, cuando Joseph Weizenbaum desarrolló ELIZA en el MIT, considerado el primer chatbot de la historia. Durante décadas, estos sistemas se basaron en reglas predefinidas y árboles de decisión limitados, lo que restringía severamente su capacidad de mantener conversaciones naturales y contextuales.

La evolución de estos sistemas puede clasificarse en tres generaciones principales. La primera generación, comprendida entre 1960 y 2000, utilizaba sistemas basados en reglas y coincidencia de patrones, como ELIZA y PARRY, que simulaban conversación mediante respuestas programadas. La segunda generación, desarrollada entre 2000 y 2017, incorporó las técnicas que usan aprendizaje automático para entender el lenguaje más natural estadístico, dando lugar a sistemas como Siri de Apple en 2011 y Alexa de Amazon en 2014, capaces de comprender intenciones básicas mediante clasificadores entrenados con grandes volúmenes de datos.

La tercera generación, iniciada con la arquitectura Transformer propuesta por Vaswani et al. (2017) en su trabajo "Attention Is All You Need", marcó un cambio paradigmático en el campo. Los Large Language Models basados en esta arquitectura, como GPT de OpenAI, BERT de Google y Claude de Anthropic, demostraron capacidades sin precedentes para generar respuestas coherentes, contextualmente relevantes y comprensivas del lenguaje humano, entrenándose con billones de palabras de texto para capturar patrones complejos del lenguaje y conocimiento general (Vaswani et al., 2017)

2.1.1.2. Large Language Models (LLM)

Los Large Language Models son modelos de inteligencia artificial basados en las redes neuronales profundas, específicamente en la arquitectura Transformer, diseñados para procesar y generar lenguaje natural. Brown et al. (2020) en su trabajo sobre GPT-3 definieron estos modelos

como sistemas entrenados mediante aprendizaje auto-supervisado en corpus masivos de texto, capaces de realizar múltiples tareas de lenguaje sin necesidad de fine-tuning específico para cada tarea

Los LLM modernos como GPT-4 de OpenAI (2023), Claude 3 de Anthropic (2024) y Llama 3 de Meta (2024) se caracterizan por cuatro atributos fundamentales. En primer lugar, su escala masiva, con modelos que contienen cientos de miles de millones de parámetros entrenados en datasets con trillones de tokens. En segundo lugar, sus capacidades emergentes, es decir, habilidades que aparecen espontáneamente al aumentar la escala del modelo, como razonamiento matemático, programación y traducción multilingüe, sin entrenamiento explícito para estas tareas. En tercer lugar, el few-shot learning, que les permite realizar nuevas tareas con apenas algunos ejemplos proporcionados en el prompt. Finalmente, su capacidad de contextualización, manteniendo coherencia en conversaciones extensas mediante ventanas de contexto que pueden abarcar miles o incluso millones de tokens.

Sin embargo, los LLM presentan desafíos significativos para aplicaciones transaccionales. El fenómeno de alucinación, mediante el cual el modelo genera información aparentemente plausible pero factualmente incorrecta, representa un riesgo crítico en sistemas de e-commerce donde la precisión de precios, stock y especificaciones técnicas es indispensable. Además, su naturaleza probabilística puede resultar en respuestas variables ante prompts similares, y su conocimiento está limitado a los datos de entrenamiento, sin acceso directo a información actualizada de sistemas externos.

2.1.1.3. Procesamiento del Lenguaje Natural (NLP)

El Procesamiento del Lenguaje Natural es un campo que permite a las máquinas entender y producir lenguaje humano mediante el uso de inteligencia artificial. Jurafsky y Martin (2023) lo definen como el conjunto de técnicas computacionales para el análisis y síntesis automáticos del lenguaje humano en sus diversas manifestaciones.

Las tareas fundamentales de NLP relevantes para asistentes conversacionales se dividen en tres áreas. La Comprensión del Lenguaje Natural (NLU) abarca procesos como la tokenización, el

análisis sintáctico, el reconocimiento de entidades nombradas, la clasificación de intenciones y la extracción de parámetros específicos del mensaje del usuario. La Generación del Lenguaje Natural (NLG) se encarga de producir respuestas coherentes, adaptar el tono comunicativo según el contexto y sintetizar información estructurada en narrativa natural. Finalmente, la Gestión del Diálogo incluye el seguimiento del estado conversacional, la gestión de turnos, la resolución de correferencias y el manejo de ambigüedades.

En el contexto de asistentes conversacionales para e-commerce de maquinaria industrial, el NLP permite transformar consultas informales como "busco una excavadora para trabajo en terreno rocoso" en consultas estructuradas a bases de datos con parámetros de categoría, tipo de terreno y capacidad, para posteriormente convertir los resultados en respuestas naturales para el usuario.

2.1.2. Comercio Electrónico y Transformación Digital

2.1.2.1. E-commerce: Conceptos Fundamentales

El comercio electrónico se define como el conjunto de transacciones comerciales realizadas a través de medios electrónicos, principalmente internet, que incluyen compra-venta de bienes con servicios, transferencias en fondos de electrónica e actividades de marketing digital. Turban et al. (2018) distinguen varios modelos fundamentales de negocio, entre los que destaca el modelo B2B (Business-to-Business), particularmente relevante para COAVPRO por tratarse de comercialización de maquinaria pesada industrial entre empresas, con volúmenes mayores de transacción y procesos de decisión de compra más complejos que en el comercio al consumidor final.

La arquitectura tecnológica moderna de e-commerce se basa en APIs RESTful que permiten separar el frontend del backend, facilitando la creación de experiencias omnicanal y la integración con sistemas externos como asistentes conversacionales. Los componentes arquitectónicos esenciales incluyen el catálogo digital con información detallada de productos, el sistema de gestión de carritos, el motor de búsqueda y filtrado, el sistema de procesamiento de pedidos, la integración con pasarelas de pago y el backend administrativo para gestión de inventario y reportes.

2.1.2.2. Comercio Conversacional

El comercio conversacional representa una evolución del e-commerce tradicional que utiliza interfaces conversacionales para facilitar interacciones comerciales. Messina et al. (2020) lo definen como la intersección del comercio electrónico con tecnologías de IA conversacional, permitiendo a los consumidores buscar, evaluar y comprar productos mediante diálogo natural. Este paradigma difiere fundamentalmente de la navegación web tradicional en cuatro aspectos clave: la interacción en lenguaje cotidiano sin necesidad de navegar menús predefinidos, la asistencia proactiva con preguntas clarificadoras y recomendaciones contextuales, la personalización basada en matices de la conversación y la reducción de fricción al eliminar pasos intermedios del proceso de compra.

Las ventajas del comercio conversacional incluyen tasas de conversión superiores, con estudios que reportan incrementos del 20 al 40%, mayor satisfacción del cliente por la experiencia personalizada y reducción de costos operativos al automatizar consultas frecuentes (Messina et al., 2020). Sin embargo, presenta desafíos técnicos significativos, principalmente la necesidad de comprensión robusta del lenguaje natural e integración confiable con sistemas backend, desafíos que este proyecto aborda mediante la arquitectura MCP.

2.1.2.3. Experiencia del Usuario en Compras Digitales

La experiencia del usuario en e-commerce se refiere al conjunto de percepciones, emociones y respuestas resultantes de la interacción con plataformas de compra digital. Nielsen Norman Group (2023) identifica cinco dimensiones críticas de UX en comercio electrónico: usabilidad, accesibilidad, credibilidad, deseabilidad y utilidad, todas ellas relevantes para el diseño del asistente conversacional de COAVPRO. En el contexto de asistentes conversacionales, los principios de diseño conversacional de Shevat (2017) resultan especialmente pertinentes: el principio de cooperación que exige información relevante y precisa, la gestión de expectativas para comunicar claramente capacidades y limitaciones, la personalidad consistente que refleje la marca, la recuperación graceful de errores sin culpar al usuario y el progreso visible que mantenga la confianza del cliente durante procesos como búsqueda de productos o creación de órdenes.

2.1.3. Arquitecturas de Integración IA-Sistemas Transaccionales

2.1.3.1. Model Context Protocol (MCP)

El Model Context Protocol, introducido por Anthropic en 2024, representa un estándar emergente para conectar Large Language Models con fuentes de datos externas y sistemas de acción. Anthropic (2024) define MCP como un protocolo abierto que estandariza cómo las aplicaciones proporcionan contexto a los LLM, permitiendo que los modelos accedan a información actualizada y ejecuten acciones estructuradas sin necesidad de entrenamiento específico

El protocolo sigue un modelo cliente-servidor compuesto por tres elementos: el MCP Host, que es la aplicación que utiliza el LLM; el MCP Client, que gestiona la comunicación; y el MCP Server, que expone recursos y herramientas al modelo. Sus componentes principales son los Resources, que representan fuentes de datos consultables como el catálogo de productos e inventario; las Tools, que son funciones invocables para ejecutar acciones como crear órdenes o verificar stock; y los Prompts, que son templates conversacionales predefinidos.

Las ventajas del MCP en sistemas transaccionales son determinantes para este proyecto: define explícitamente qué operaciones puede ejecutar el LLM previniendo acciones no autorizadas, valida parámetros antes de ejecutar acciones asegurando integridad de datos, mantiene la lógica de negocio en el backend limitando al LLM a la orquestación conversacional, y fundamentalmente garantiza que el modelo trabaje únicamente con datos reales del sistema e-commerce de COAVPRO, eliminando completamente la posibilidad de inventar precios, disponibilidad o especificaciones técnicas de maquinaria.

2.1.3.2. APIs RESTful en E-commerce

Las APIs REST constituyen el estándar arquitectónico predominante para la comunicación entre sistemas en e-commerce moderno. Fielding (2000) definió los principios REST que han sido adoptados ampliamente por su simplicidad, escalabilidad y compatibilidad con HTTP, estableciendo cinco principios fundamentales: separación cliente-servidor, stateless donde cada petición contiene toda la información necesaria, capacidad de caché, interfaz uniforme mediante métodos HTTP estándar y sistema en capas.

Para la integración con asistentes conversacionales, las consideraciones más relevantes incluyen la autenticación mediante tokens JWT para identificar usuarios a través de la conversación, el rate limiting para protección contra uso excesivo, la validación rigurosa de inputs, el manejo estructurado de errores que el asistente pueda interpretar y comunicar al usuario, y la paginación para gestionar grandes conjuntos de resultados del catálogo de maquinaria.

2.1.3.3. Orquestación de Agentes Conversacionales

La orquestación de agentes se refiere al proceso de coordinar múltiples componentes y flujos de trabajo para que un sistema conversacional ejecute tareas complejas de forma coherente. Wang et al. (2023) distinguen entre agentes reactivos, que responden a inputs individuales, y agentes deliberativos, que planifican secuencias de acciones para alcanzar objetivos.

El patrón arquitectónico ReAct (Reason + Act) resulta especialmente relevante para este proyecto, ya que el LLM alterna entre razonamiento interno y ejecución de acciones: primero piensa en qué hacer, luego ejecuta una acción mediante herramienta MCP, observa el resultado y finalmente responde al usuario con información verificada. En el contexto del asistente de COAVPRO, el sistema debe mantener seguimiento de productos consultados, filtros y preferencias expresadas, estado del carrito actual y el punto en el flujo de compra, permitiendo coherencia a lo largo de conversaciones multi-turno.

2.1.4. Aplicaciones de Mensajería para Comercio Digital

2.1.4.1. Plataformas de Mensajería Instantánea como Canal Comercial

Las aplicaciones de mensajería instantánea han evolucionado de simples herramientas de comunicación personal a plataformas comerciales robustas. Aplicaciones como WhatsApp, Telegram, Facebook Messenger y WeChat ofrecen APIs empresariales que permiten a las organizaciones interactuar con clientes a escala. Sus características comunes incluyen APIs para envío y recepción automatizada de mensajes, plantillas pre-aprobadas para notificaciones transaccionales, mensajes interactivos con botones y listas, soporte multimedia y protocolos de cifrado para proteger comunicaciones.

Las ventajas de estas plataformas como canal comercial son significativas: adopción masiva con miles de millones de usuarios activos globalmente, familiaridad que reduce la curva

de aprendizaje, notificaciones directas que garantizan visibilidad, historial de conversación persistente y capacidad omnicanal. Telegram en particular, con más de 900 millones de usuarios activos mensuales, ofrece una API robusta con soporte para webhooks en tiempo real, constituyendo el canal inicial de despliegue del asistente de COAVPRO

2.1.4.2. Casos de Uso en Comercio Electrónico

La implementación de asistentes conversacionales en plataformas de mensajería ha demostrado efectividad documentada en múltiples escenarios. Sephora Virtual Artist, implementado desde 2017 en Facebook Messenger, reportó un incremento del 11% en tasa de conversión y 25% en valor promedio de pedido. H&M Fashion Assistant procesó más de un millón de conversaciones en su primer año con una satisfacción del cliente del 70%. Domino's Pizza Bot, activo desde 2016 en múltiples plataformas, contribuyó a que el 65% de pedidos en línea se realizaran mediante canales digitales.

Estos casos destacan patrones comunes de éxito aplicables al proyecto de COAVPRO: enfoque en casos de uso específicos y bien definidos, integración profunda con sistemas backend existentes, inversión en diseño conversacional y pruebas con usuarios reales, y monitoreo continuo con mejora iterativa basada en datos de uso.

2.1.5. Antecedentes de Investigación

2.1.5.1. Estudios sobre Chatbots en E-commerce

La literatura académica sobre chatbots en comercio electrónico ha crecido significativamente en la última década. Adamopoulou y Moussiades (2020) realizaron una revisión sistemática identificando que los chatbots mejoran la experiencia del cliente principalmente a través de disponibilidad 24/7, respuesta instantánea y personalización. Følstad y Brandtzaeg (2020) investigaron las expectativas de los usuarios hacia chatbots en servicios comerciales, encontrando que los usuarios valoran principalmente la eficiencia funcional sobre la humanización de la conversación, hallazgo que orientó el diseño del asistente de COAVPRO hacia la precisión y utilidad antes que hacia simular conversación humana. Chung et al. (2020) demostraron que los asistentes conversacionales incrementan la intención de compra en un 23% comparado con la

navegación web tradicional, atribuible a la reducción del esfuerzo cognitivo y una experiencia más intuitiva.

2.1.5.2. Investigación sobre Model Context Protocol

El Model Context Protocol, siendo una tecnología emergente introducida en 2024, cuenta aún con literatura académica limitada. Sin embargo, la documentación técnica oficial de Anthropic presenta arquitecturas de referencia para integración con sistemas empresariales en casos de uso como conexión con bases de datos corporativas, integración con APIs de CRM y ERP, acceso a repositorios de documentación interna y ejecución de operaciones administrativas controladas. Implementadores tempranos reportan ventajas significativas en control de alucinaciones y precisión de acciones ejecutadas por LLM cuando se utiliza MCP frente a enfoques ad-hoc de function calling, fundamento técnico que respalda su adopción en este proyecto

2.2. Marco Conceptual

Se presentara la definición de términos técnicos fundamentales utilizados en esta investigación:

API (Application Programming Interface): Son definiciones y protocolos que nos permiten la comunicación entre varias aplicaciones en software. Este proyecto engloba, la API RESTFUL del sistema e-commerce de COAVPRO exponen funcionalidades como consulta de productos y creación de órdenes que el asistente conversacional puede invocar.

Asistente Conversacional: Sistema de inteligencia artificial diseñado para interactuar con usuarios mediante diálogo en lenguaje natural, capaz de comprender intenciones, mantener contexto conversacional y ejecutar acciones sobre sistemas externos para satisfacer las necesidades del usuario.

Carrito de Compras: Componente de un sistema de e-commerce que almacena temporalmente productos seleccionados por el usuario antes de proceder al checkout, permitiendo agregar, modificar o eliminar productos antes de confirmar la compra.

Chatbot: Programa de software que simula conversación con usuarios humanos, especialmente a través de internet. Puede basarse en reglas predefinidas o utilizar inteligencia artificial para generar respuestas.

Comercio Conversacional: Modalidad de comercio electrónico que utiliza interfaces conversacionales para facilitar transacciones comerciales mediante diálogo natural entre cliente y sistema, permitiendo buscar productos, recibir recomendaciones y completar compras sin salir del canal de mensajería.

Comercio Electrónico (E-commerce): Actividades económicas que nos permiten la compra, venta de productos o servicios por medios electrónicos, principalmente internet, incluyendo transacciones, marketing digital y servicio al cliente en línea.

Endpoint: URL específica en una API que representa un recurso o funcionalidad particular, como /api/products para acceder al catálogo de maquinaria de COAVPRO.

Function Calling: Capacidad de los LLM de invocar funciones externas de forma estructurada, pasando parámetros apropiados basados en la conversación con el usuario.

Ingeniería de Prompts (Prompt Engineering): Disciplina que se enfoca en diseñar, optimizar e refinar las indicaciones proporcionadas en modelos de lenguaje para obtener respuestas más concisas, incluyendo técnicas como few-shot learning, chain-of-thought prompting y role prompting.

Large Language Model (LLM): Modelo de inteligencia artificial basado en redes neuronales profundas con arquitectura Transformer, entrenado con enormes volúmenes de texto, capaz de comprender y generar lenguaje natural con alta coherencia contextual.

MCP (Model Context Protocol): Protocolo estándar introducido por Anthropic en 2024 que permite conectar Large Language Models con fuentes de datos externas y sistemas de acción de forma estructurada, controlada y segura, garantizando que el modelo trabaje únicamente con datos reales del backend.

MCP Server: Servicio que implementa el Model Context Protocol, exponiendo recursos y herramientas que un LLM puede consultar o invocar durante la interacción con usuarios.

NLP (Natural Language Processing): Campos de inteligencia artificial enfocados en la interacción entre computadoras y lenguaje humano, incluyendo comprensión, interpretación y generación de lenguaje natural.

NLU (Natural Language Understanding): Subdominio del NLP enfocado en la comprensión del significado del lenguaje humano, incluyendo análisis sintáctico, semántico y pragmático.

Orquestación de Agentes: Proceso de coordinar múltiples componentes de un sistema conversacional (LLM, herramientas MCP, memoria, APIs) para ejecutar tareas complejas de forma coherente y eficiente.

PostgreSQL: Es un sistema para la gestión de bases de datos relacionales de código abierto, utilizando en este proyecto con la extensión pgvector para búsqueda semántica vectorial del catálogo de productos.

RAG (Retrieval-Augmented Generation): Técnicas que combinan la recuperación de información con generación de lenguaje, permitiendo que LLM accedan a bases de conocimiento específicas y actualizadas antes de generar respuestas, mejorando precisión y reduciendo alucinaciones.

ReAct (Reason + Act): Patrón arquitectónico para agentes conversacionales donde el LLM alterna entre razonar sobre qué hacer y ejecutar acciones mediante herramientas, observando resultados antes de decidir el siguiente paso.

Tasa de Conversión: Métrica de e-commerce que representa el porcentaje de visitantes que completan una acción deseada, calculada como el número de conversiones dividido entre el número de visitantes, multiplicado por cien.

Transformer: Arquitectura de red neuronal introducida por Vaswani et al. (2017) que utilizan mecanismos de atención para procesar secuencias de datos, siendo la base de los LLM modernos.

Webhook: Mecanismo de comunicación donde una aplicación envía automáticamente datos a otra en tiempo real cuando ocurre un evento específico, mediante petición HTTP POST a una URL predefinida. Utilizado en este proyecto para recibir mensajes de Telegram en tiempo real.

CAPÍTULO III: METODOLOGÍA DE INVESTIGACIÓN

3.1. Enfoque Investigativo

Esta presente investigación toma un enfoque metodológico mixto, sustentando un diseño de investigación aplicada de carácter proyectivo o tecnológico. Este enfoque nos permite abordar el problema de estudio de manera sistemática y rigurosa, combinando las recolecciones de datos cualitativos para comprender el contexto operativo de COAVPRO con datos cuantitativos que posibilitan la medición objetiva del desempeño del sistema desarrollado y la validación empírica de la solución propuesta.

Como señala Hernández Sampieri et al. (2014) en "Metodología de la Investigación", el enfoque mixto representa el más alto grado de integración entre los métodos cualitativos y cuantitativos, logrando una perspectiva más amplia y profunda del fenómeno estudiado que cualquiera de los dos enfoques por separado. En este contexto, el desarrollo de un asistente conversacional inteligente integrado al sistema e-commerce de COAVPRO constituye el resultado fundamental del estudio, el cual es evaluado a través de indicadores cuantificables que permiten medir su desempeño, precisión, eficiencia y funcionalidad, complementados con análisis cualitativos sobre la experiencia del usuario y la pertinencia de la solución al contexto industrial de la empresa.

De acuerdo con Baena Paz (2017) en "Metodología de la Investigación", los diseños de carácter proyectivo o tecnológico requieren una validación que emerge del funcionamiento efectivo del sistema creado en un contexto real o simulado. En coherencia con este planteamiento, el presente proyecto se enfoca en el análisis de la problemática específica de COAVPRO, la concepción, construcción y validación del asistente conversacional mediante métricas de rendimiento, pruebas funcionales y datos medibles.

3.2. Tipo y Nivel de Investigación

3.2.1. Enfoque Aplicativo

Como propósito fundamental, se plantea el desarrollo de un sistema tecnológico orientado a resolver las problemáticas específicas identificadas en el proceso de atención al cliente y ventas en línea de COAVPRO. Este enfoque se caracteriza por la aplicación de conocimientos técnicos en inteligencia artificial conversacional, integración de sistemas y experiencia de usuario, centrados en el desarrollo de software para generar resultados que puedan ser implementados y validados en un entorno real de comercio electrónico de maquinaria pesada industrial.

Partiendo de la necesidad de mejorar la experiencia del cliente y reducir los tiempos de búsqueda y consulta de productos, se propone diseñar, construir y validar un asistente conversacional que integre un modelo de lenguaje local mediante Ollama, búsqueda semántica vectorial con PostgreSQL y pgvector, y herramientas MCP para garantizar la integridad de los datos transaccionales. El problema central se ubica en las ineficiencias del proceso tradicional de consulta y compra, marcado por la dependencia de canales de atención con horario limitado y la ausencia de interfaces conversacionales naturales que faciliten el descubrimiento de productos industriales.

3.2.2. Enfoque Mixto

El enfoque adoptado integra métodos cualitativos y cuantitativos de manera complementaria. Desde lo cualitativo, se busca comprender el contexto específico de COAVPRO, identificar las funcionalidades requeridas por sus clientes y analizar el funcionamiento interno de sus procesos comerciales, mediante entrevistas con personal clave y análisis de la documentación del sistema e-commerce existente. Desde lo cuantitativo, se mide la efectividad, usabilidad y precisión del asistente desarrollado mediante métricas de rendimiento, pruebas funcionales y análisis estadístico de resultados comparativos entre el proceso tradicional y el asistido por inteligencia artificial.

3.3. Corte de la Investigación

El presente estudio adoptó un diseño transversal que facilitó una aproximación integral al problema, permitiendo capturar y analizar la situación existente en un período específico para posteriormente transformar esos hallazgos en el desarrollo de la solución tecnológica propuesta. Dentro de este marco temporal delimitado, el proceso se estructuró en tres etapas fundamentales.

La primera es una fase diagnóstica, principalmente no experimental, destinada a comprender la situación actual de COAVPRO mediante el levantamiento de requerimientos funcionales, el análisis del sistema e-commerce existente y sus APIs disponibles, la identificación de casos de uso prioritarios y la evaluación de métricas actuales del proceso de compra como línea base.

La segunda es una fase de diseño arquitectónico, orientada a definir la integración entre el LLM y el sistema e-commerce mediante MCP, establecer las herramientas y recursos a exponer, diseñar los flujos conversacionales para los casos de uso identificados y definir la estrategia de ingeniería de prompts para el comportamiento del asistente.

La tercera es la fase del desarrollo e implementación iterativa, basándose en los principios de la metodología SCRUM, donde se construye incrementalmente el sistema mediante sprints planificados. Esta estructura nos permitió un enfoque de la investigación, mucho más centrado en crear un asistente conversacional, se apoyara en un diagnóstico preciso y vigente, resguardando que la solución responda a necesidades reales del contexto industrial de COAVPRO.

3.4. Instrumentos y Técnicas para el Levantamiento de la Información

Para obtener mejor comprensión de la situación presente de COAVPRO y validar el sistema desarrollado, se emplean técnicas mixtas de recolección de datos que combinan métodos cualitativos y cuantitativos.

Entrevistas Semiestructuradas

Se aplican a personal clave de COAVPRO, incluyendo al Ing. Pedro Argudo en su rol de administrador general, y al equipo técnico involucrado en el proyecto, permitiendo recopilar

información cualitativa sobre los procesos actuales de atención al cliente, las dificultades experimentadas durante el proceso de compra, las expectativas frente al asistente conversacional propuesto y los requerimientos funcionales específicos del contexto de comercialización de maquinaria industrial.

Instrumento: Guía de entrevista previamente validada con preguntas abiertas sobre experiencia actual, necesidades identificadas y expectativas de funcionalidad.

Revisión Documental

Se examinan documentos técnicos y funcionales del sistema e-commerce existente, incluyendo la documentación de APIs REST disponibles, diagramas de arquitectura del sistema actual, modelos de datos, políticas comerciales y estadísticas de uso del e-commerce como tráfico, conversión y abandono.

Instrumento: Fichas de análisis documental para sistematizar información técnica relevante.

Encuestas Estructuradas

Como técnica cuantitativa, se utilizan formularios digitales aplicados a usuarios piloto del asistente conversacional para evaluar la satisfacción general con la experiencia conversacional, la facilidad de uso percibida, la utilidad de las funcionalidades implementadas y la intención de uso futuro del sistema.

Instrumento: Cuestionario digital mediante Google Forms con escala Likert para medir satisfacción y usabilidad, basado en el modelo validado System Usability Scale (SUS)

Pruebas Funcionales y Técnicas

Se realizan pruebas sistemáticas para validar la precisión y confiabilidad del sistema, verificando que el asistente ejecute las acciones correctas según las solicitudes del usuario, que las llamadas a las APIs del e-commerce funcionen correctamente, que el comportamiento ante inputs ambiguos o productos no disponibles sea apropiado, y que la autenticación y protección de datos sensibles operen de forma segura.

Instrumento: Casos de prueba estructurados con inputs esperados, acciones a ejecutar y resultados esperados, complementados con logs de errores y comportamientos anómalos del sistema.

3.5. Población y Muestra

El estudio de la población está conformada por dos grupos principales. El grupo interno en COAVPRO incluye al personal administrativo y técnico directamente involucrado en la gestión del e-commerce y en el desarrollo del sistema. El grupo externo comprende a los clientes actuales de COAVPRO que utilizan el sistema e-commerce existente y a clientes potenciales que utilizan aplicaciones de mensajería instantánea como Telegram.

Dado el alcance del proyecto como prueba piloto, se seleccionó una pequeña muestra no probabilística compuestas por usuarios, cuya participación fue primordial para la validación de funcionales, técnicas de usabilidad del asistente conversacional.

Tabla 1. Población y Muestra

Unidad de Estudio	Población
Personal interno COAVPRO	4 personas
Clientes piloto	15-20 usuarios

Nota. Los criterios de inclusión para usuarios piloto establecen que deben ser clientes actuales o potenciales de COAVPRO, tener acceso a Telegram

3.6. Metodología de Trabajo

Para el desarrollo del asistente conversacional se emplea SCRUM, metodología ágil que facilita el avance en fases breves y el progreso gradual del sistema, permitiendo ajustarse a cambios y retroalimentación constante durante el proceso de construcción.

3.6.1. Fases de Desarrollo del Sistema

Esta metodología se organiza en iteraciones breves en las que se planifican, desarrollan y evalúan mejoras del producto.

Tabla 2. Fases de la metodología SCRUM relacionadas al proyecto

Fase	Descripción	Aplicación en el Proyecto
Product Backlog	Lista priorizada de funcionalidad del sistema	Incluye módulos de búsqueda semántica, herramientas MCP, gestión de conversaciones, integración con Telegram y panel de administración
Sprint Planning	Reunión para planificar tareas del sprint	Se define qué funcionalidades específicas se desarrollan en cada iteración de 3-4 semanas
Sprint (Iteración)	Ciclo de desarrollo con entregables funcionales	Se construyen y prueban componentes del sistema en cada sprint, garantizando que cada entrega sea funcional
Daily Scrum	Reuniones breves diarias de seguimiento	Se revisan avances y se resuelven bloqueos técnicos del desarrollo del asistente
Sprint Review	Presentación del producto funcional	Se muestran avances para obtener retroalimentación sobre usabilidad y funcionalidad del asistente conversacional
Sprint Retrospective	Evaluación del proceso y mejoras	Se identifican dificultades técnicas y se proponen ajustes para optimizar el flujo de trabajo en el siguiente sprint

Nota. Tabla 2. Fases de la metodología SCRUM aplicadas al desarrollo del asistente conversacional de COAVPRO. Elaborado por Elkin Carriel y Alex Calle (2025), basado en Schwaber y Sutherland (2020).3.6.2. Roles y Responsabilidades SCRUM

Tabla 3. Roles SCRUM asignados

Rol	Responsables	Funciones
Product Owner	Ing. Pedro Argudo	Establece prioridades de desarrollo, aprueba funcionalidades entregadas y vela por los objetivos comerciales de COAVPRO
Scrum Master	Alex Calle	Vela por la aplicación correcta de la metodología, remueve impedimentos y asegura la adhesión a las prácticas ágiles
Equipo de Desarrollo	Elkin Carriel Alex Calle	Encargados del diseño, programación e integración del asistente conversacional con el sistema e-commerce

Nota. Tabla 3. Roles SCRUM asignados al equipo de desarrollo del proyecto. Elaborado por Elkin Carriel y Alex Calle (2025).

3.6.3. Herramientas y Técnicas para Implementación de SCRUM

Con el objetivo de facilitar la implementación de la metodología SCRUM, se emplean diversas herramientas digitales que ayudan a planificar, registrar, comunicar y supervisar el ciclo de desarrollo. La Tabla 4 enumera las herramientas seleccionadas y su papel específico dentro del proyecto.

Tabla 4. Herramientas utilizadas en la metodología SCRUM

Elemento SCRUM	Herramientas utilizadas	Función dentro del Proyecto
Gestión de tareas y backlog	GitHub Projects	Organiza el Product Backlog, planifica sprints y asigna tareas mediante tableros Kanban integrados al repositorio del código fuente
Comunicación y reuniones	Google Meet	Facilita reuniones de seguimiento diario, coordinación técnica entre Elkin Carriel y Alex Calle, y revisión de avances con el Product Owner
Documentación del proyecto	Google Docs	Permite la redacción colaborativa de avances, documentación técnica, informes de sprint y contenido del proyecto
Control de versiones y desarrollo	GitHub	Almacena el código fuente del asistente conversacional, permite colaboración entre los miembros del equipo y mantiene el control de versiones durante todo el desarrollo

Nota. Tabla 4. Herramientas utilizadas para la implementación de la metodología SCRUM. Elaborado por Elkin Carriel y Alex Calle (2025).

3.7. Técnicas de Análisis de Datos

Los datos recopilados durante el desarrollo y evaluación del proyecto son analizados con un enfoque mixto que combina métodos cuantitativos y cualitativos.

Análisis Cuantitativo

Las encuestas estructuradas y métricas de rendimiento se procesan mediante estadística descriptiva, utilizando medidas de tendencia central como promedio y mediana para tiempos de respuesta del sistema, número de turnos conversacionales y puntuaciones de satisfacción. Se calculan frecuencias y porcentajes para métricas como tasa de resolución automática, tasa de éxito de herramientas MCP y distribución de tipos de consultas. Las herramientas utilizadas incluyen Google Forms para análisis automático de encuestas y Python con las bibliotecas pandas y matplotlib para análisis más avanzados y visualizaciones de resultados.

Análisis Cualitativo

Las entrevistas semiestructuradas y observaciones de pruebas de usabilidad se analizan mediante codificación temática, siguiendo un proceso que incluye la transcripción de entrevistas, la codificación abierta para identificar conceptos emergentes, la codificación axial para agrupar códigos en categorías más amplias y la interpretación de insights sobre fortalezas, debilidades y oportunidades de mejora del sistema. Las categorías predefinidas de análisis incluyen usabilidad conversacional, efectividad funcional, confiabilidad de datos verificados mediante MCP y satisfacción general del usuario con la experiencia de compra conversacional.

CAPÍTULO IV: ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS

4.1. Diagnóstico del Estado Actual

El análisis de las problemáticas se realizó con el fin de identificar las principales dificultades que afectan los procesos de venta y atención al cliente en el sistema de comercio electrónico de COAVPRO. Para este fin, se llevó a cabo un levantamiento de información mediante encuestas aplicadas a clientes de la empresa, organizando los hallazgos según los desafíos operativos identificados y la experiencia del usuario con plataformas digitales de compra.

Tabla 5. Desafíos identificados en el proceso de compra en línea de COAVPRO

Problema	Evidencia encontrada
Dificultad para hacer preguntas sobre productos	45.5% de encuestados lo identificó como principal dificultad
Dificultad para encontrar productos específicos	31.8% de encuestados lo reportó como obstáculo frecuente
Desconfianza en métodos de pago	31.8% de encuestados expresó esta preocupación
Falta de información clara sobre productos	22.7% de encuestados señaló esta limitación
Proceso de compra muy largo o complicado	18.2% de encuestados identificó esta fricción
Dudas sobre disponibilidad de productos	18.2% de encuestados manifestó esta dificultad
Baja adopción de asistentes conversacionales en compras	Solo 13.6% ha interactuado frecuentemente con chatbots en mensajería

Nota. Tabla 5. Desafíos identificados en el proceso de compra en línea de COAVPRO. Elaborado por Elkin Carriel y Alex Calle (2025), basado en encuestas aplicadas a clientes de COAVPRO.

4.2. Creación de Encuestas

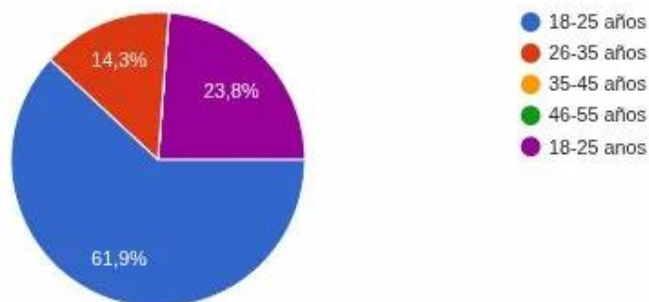
4.2.1. Requerimientos Previos

Para identificar las causas principales de los problemas y validar las necesidades de una solución tecnológica conversacional, se aplicó una encuesta. Esta encuesta, elaborada en la plataforma Google Forms, combinó preguntas cerradas y abiertas organizadas en cuatro secciones: perfil del encuestado, experiencia con compras en línea, uso de aplicaciones de mensajería, y satisfacción y evaluación del asistente conversacional propuesto. La encuesta obtuvo 22 respuestas válidas de clientes actuales y potenciales de la empresa.

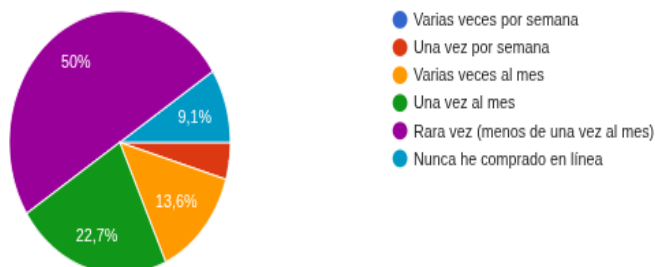
4.3. Análisis de los Resultados

Figura 1

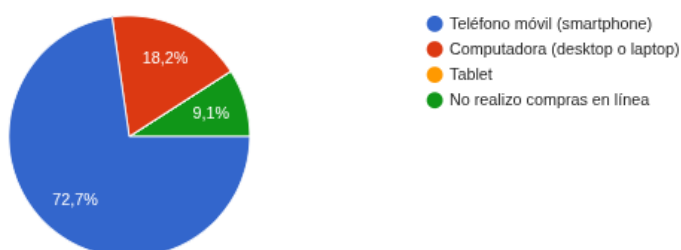
Rango de edad de las personas



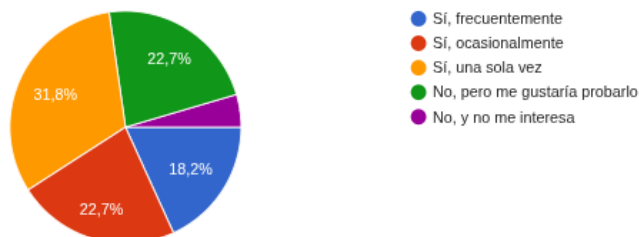
Nota. Los resultados de la primera sección de la encuesta revelan características importantes sobre el perfil de los clientes de COAVPRO. En cuanto a la edad, el 61.9% de los encuestados pertenece al rango de 18-25 años, seguido por un 23.8% en el rango de 26-35 años y un 14.3% entre 35-45 años. Esta distribución indica que la mayoría de los clientes son adultos jóvenes, lo que sugiere una mayor familiaridad con tecnologías digitales y aplicaciones de mensajería, aspecto favorable para la adopción del asistente conversacional.

Figura 2***Experiencia con Compras en Línea***

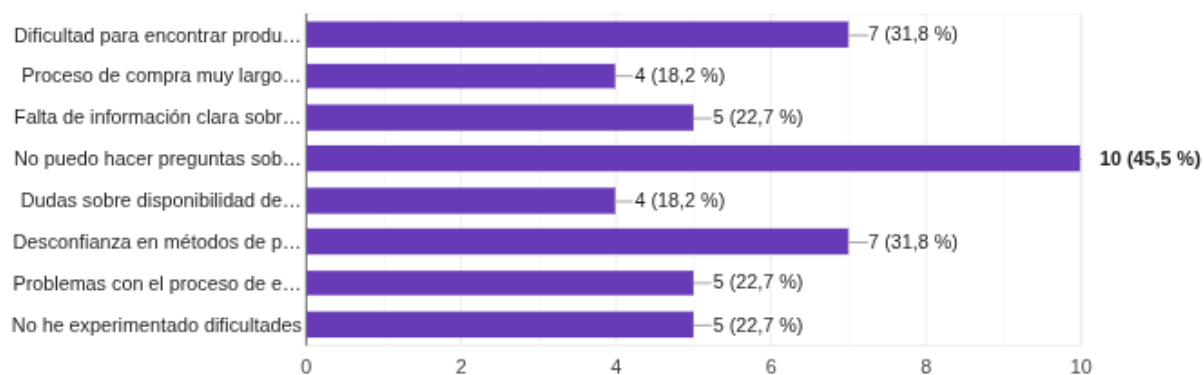
Nota. En relación con la frecuencia de compras en línea, el 50% de los encuestados realiza compras raramente, es decir menos de una vez al mes, mientras que el 22.7% compra una vez al mes y el 13.6% varias veces al mes. Este dato es significativo para COAVPRO, pues revela que una porción importante de sus clientes potenciales no tiene el hábito consolidado de compra en línea, lo que refuerza la necesidad de una experiencia conversacional más intuitiva y guiada que reduzca la fricción del proceso.

Figura 3***Dispositivo utiliza principalmente para comprar en línea***

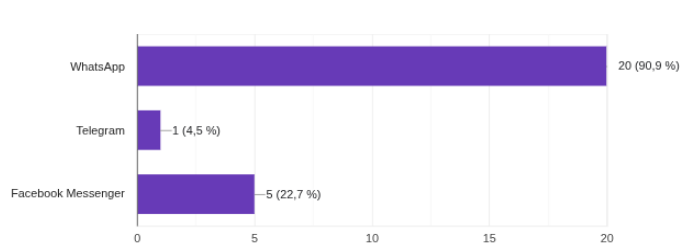
Nota. Sobre el dispositivo utilizado para comprar en línea, el 72.7% utiliza el teléfono móvil como dispositivo principal, el 18.2% usa computadora y el 9.1% tableta.

Figura 4*Frecuencia en el que usa un sistema de comercio electrónico*

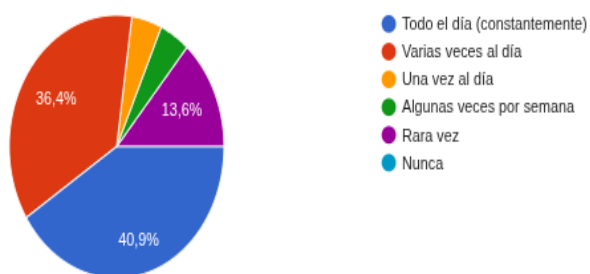
Nota. En cuanto al uso previo del sistema e-commerce de COAVPRO, el 31.8% nunca lo ha usado pero le gustaría probarlo, el 22.7% lo ha utilizado frecuentemente, el 22.7% ocasionalmente y el 22.7% solo una vez.

Figura 5*Principales Dificultades en el Proceso de Compra*

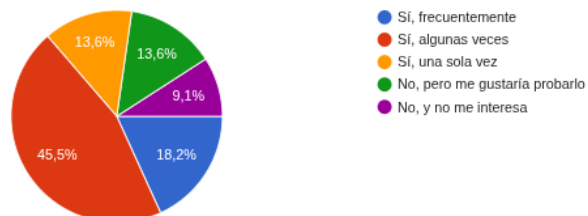
Nota. La pregunta sobre las principales dificultades experimentadas al comprar en línea arrojó resultados determinantes para la justificación del proyecto. La dificultad más frecuentemente reportada fue la imposibilidad de hacer preguntas sobre productos con un 45.5% de respuestas, seguida por la dificultad para encontrar productos específicos con 31.8% y la desconfianza en métodos de pago también con 31.8%.

Figura 6***Uso de Aplicaciones de Mensajería***

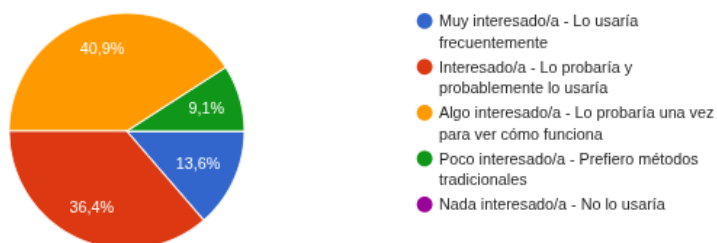
Nota. Los resultados de la sección sobre mensajería instantánea son fundamentales para validar la estrategia de canal adoptada en el proyecto. En cuanto a las aplicaciones utilizadas regularmente, WhatsApp lidera con un 90.9% de uso, seguido por Facebook Messenger con 22.7% y Telegram con 4.5%.

Figura 7***Interés y Evaluación del Asistente Conversacional***

Nota. Los resultados de la sección de satisfacción y evaluación confirman una alta receptividad hacia el asistente conversacional propuesto. Ante la pregunta sobre el interés en utilizar un asistente conversacional para realizar compras, el 40.9% respondió estar algo interesado y lo probaría una vez para ver cómo funciona, el 38.4% está interesado y probablemente lo usaría, y el 13.6% está muy interesado y lo usaría frecuentemente

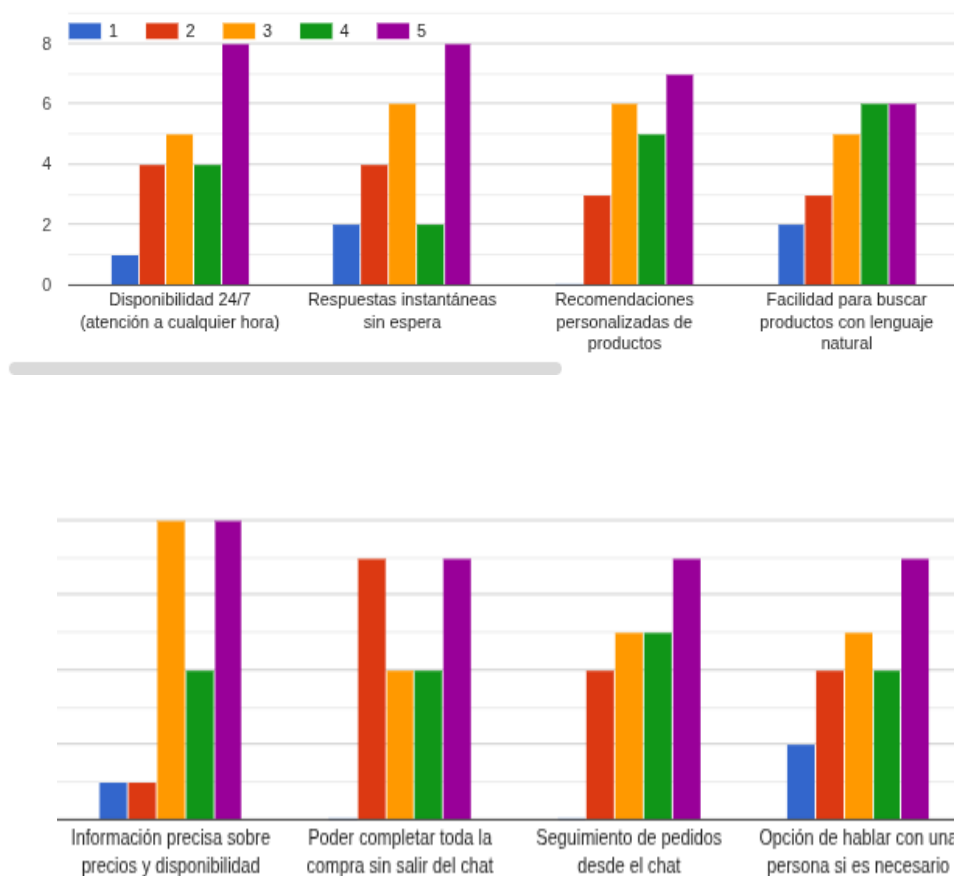
Figura 8***Interacción con chatbots o asistentes virtuales en aplicaciones de mensajería***

Nota. La Figura 8 observa que el 45.5% de los encuestados ha interactuado con chatbots o asistentes virtuales en aplicaciones de mensajería algunas veces, seguido por un 18.2% que lo ha hecho frecuentemente y un 13.6% que lo ha utilizado una sola vez.

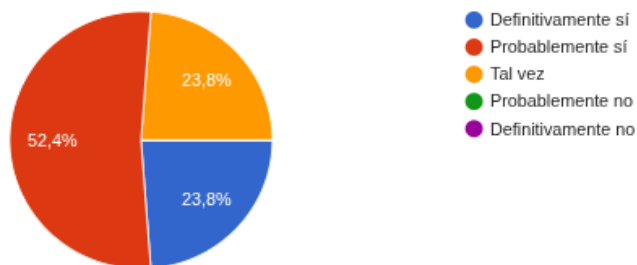
Figura 9***Nivel de interés en utilizar el asistente conversacional para realizar compras***

Nota. En la Figura 9 los datos de la sección de Satisfacción y Evaluación revelan que el 40.9% de los encuestados está algo interesado y lo probaría una vez para ver cómo funciona, el 36.4% está interesado y probablemente lo usaría, y el 13.6% está muy interesado y lo usaría frecuentemente.

Figura 10

Importancia de las características del asistente conversacional

Nota. En la Figura 10 presenta resultados de la evaluación de importancia en las características del asistente conversacional en escala del 1 al 5, donde 1 es nada importante y 5 es muy importante

Figura 11***Disposición a recomendar el asistente conversacional***

Nota. En la Figura 11, ante la pregunta de si recomendarían el asistente conversacional a otras personas si funcionara bien, el 52.4% respondió que probablemente sí lo recomendaría, el 23.8% definitivamente sí y el 23.8% tal vez.

CAPÍTULO V: PROPUESTA DE INVESTIGACIÓN

5.1. Introducción

En este contexto la transformación digital, las empresas dedicadas a la comercialización de maquinaria pesada industrial enfrentan el desafío de modernizar sus canales de venta y atención al cliente para mejorar la eficiencia operativa, optimizar la gestión de consultas y ofrecer una experiencia de compra personalizada y disponible en todo momento. Esta propuesta plantea el desarrollo e implementación de un asistente conversacional inteligente integrado al sistema de comercio electrónico de COAVPRO, con el objetivo de automatizar el proceso de ventas mediante interacción en lenguaje natural a través del canal de mensajería Telegram.

Esta investigación no solo documenta el desarrollo técnico del sistema, sino que también evalúa su eficiencia en la práctica, analiza cómo beneficia a los usuarios del sector industrial y examina su impacto en los procesos comerciales de COAVPRO.

5.2. Objetivo de la Propuesta

5.2.1. Objetivo General

Desarrollar un asistente conversacional inteligente integrado al sistema de comercio electrónico de COAVPRO que permita automatizar el proceso de ventas de maquinaria pesada industrial mediante interacción en lenguaje natural, verificación en tiempo real de stock y precios, y atención personalizada 24/7 a través del canal de mensajería Telegram.

5.2.2. Objetivos Específicos

- Implementar un sistema de búsqueda semántica de productos mediante RAG con pg vector que permita a los clientes encontrar maquinaria industrial describiendo sus necesidades en lenguaje natural.
- Desarrollar una arquitectura MCP (Model Context Protocol) que garantice la integridad de datos de productos, precios y disponibilidad mediante integración directa con el sistema de comercio electrónico de COAVPRO.

- Integrar el asistente conversacional en el canal de mensajería Telegram para proporcionar atención 24/7 con gestión de contexto conversacional persistente.
- Implementar un panel de administración con métricas en tiempo real que permita monitorear el rendimiento del asistente y los patrones de consulta de los clientes.

5.3. Justificación

Esta investigación justifica la necesidad de contar con herramientas tecnológicas que refuercen la competitividad de COAVPRO en el mercado de maquinaria pesada industrial. Este sistema implementado representa una innovación que integra procesamiento de lenguaje natural, búsqueda semántica vectorial y verificación de datos en tiempo real en una interfaz conversacional única, lo que permite:

- Eliminar completamente la generación de información incorrecta sobre productos, precios y stock mediante la arquitectura MCP, garantizando que el 100% de los datos críticos provenga directamente del sistema e-commerce.
- Optimizar los tiempos de búsqueda de maquinaria industrial mediante procesamiento de lenguaje natural y búsqueda vectorial con pgvector, permitiendo consultas descriptivas en lugar de navegación manual por categorías.
- Proporcionar atención inmediata 24/7 a través de Telegram sin necesidad de intervención humana constante, resolviendo la limitación de los canales de atención tradicionales con horario restringido.
- Mantener conversaciones contextuales persistentes que evolucionen según los intereses técnicos del cliente, mejorando la experiencia de consulta de equipos industriales especializados.
- Generar insights estratégicos de negocio mediante el registro completo de interacciones, identificando productos más consultados, gaps en el catálogo y patrones de búsqueda emergentes.

5.4. Metodología de la Investigación

- **Tipo de investigación:** Aplicada y de Desarrollo.
- **Diseño metodológico:** En COAVPRO se implementó el asistente conversacional inteligente, estableciendo una relación causal entre su implementación y los cambios observados en los indicadores de gestión comercial. Para esto, se realizaron comparaciones de métricas clave antes y después de su implementación, incluyendo tiempo de búsqueda de productos, tasa de resolución automática de consultas y satisfacción del cliente.
 - **Población y muestra:** Clientes activos del sitio web de COAVPRO y usuarios de Telegram.
 - **Instrumentos de recolección de datos:** Cuestionarios digitales dirigidos a sus clientes para identificar sus requerimientos para verificar la utilidad del asistente, análisis en los logs de conversación para evaluar efectividad del sistema, métricas de uso de herramientas MCP para validar integridad de datos, y análisis de tiempos de respuesta del sistema.
 - **Procedimiento:**
 1. Implementación experimental del sistema en entorno de prueba.
 2. Sincronización inicial del catálogo de productos con la base de datos vectorial.
 3. Lanzamiento piloto con grupo selecto de usuarios en canal Telegram.
 4. Recopilación de datos durante la fase de prueba de cuatro semanas.
 5. Evaluación comparativa de indicadores antes y después.

5.5. Resultados Esperados

- Tiempo promedio de búsqueda de productos reducido en un 50% mediante búsqueda en lenguaje natural comparado con navegación web tradicional.
- Precisión de información de stock y precios del 100% mediante verificación en tiempo real con herramientas MCP.
- Tasa de resolución automática de consultas superior al 70% sin intervención humana.
- Reducción de consultas incorrectas o información inventada a 0% mediante arquitectura MCP.

- Tiempo de respuesta del sistema menor a 10 segundos incluyendo verificación con herramientas MCP.
- Satisfacción del cliente con valoración positiva superior al 85%.
- Disponibilidad de atención 24/7 en canal Telegram sin costos recurrentes por API externa.

5.6. Cronograma y Recursos

- **Cronograma:** La duración estimada para el diseño, desarrollo, implementar, pruebas y test de evaluaciones del sistema tiempo de 18 semanas, organizadas en cuatro sprints de desarrollo iterativo.
- **Recursos:**
 - Servidor con GPU para ejecución del modelo de lenguaje local mediante Ollama.
 - Infraestructura para PostgreSQL con extensión pgvector y Redis para caché de sesiones.
 - Token de bot de Telegram para integración del canal de mensajería.
 - Herramientas de desarrollo: FastAPI, FastMCP, python-telegram-bot, sentence-transformers.
 - Personal técnico especializado: Elkin Carriel y Alex Calle como equipo de desarrollo, bajo la dirección del Ing. Pedro Argudo como Product Owner.

5.7. Entorno de Desarrollo de la Aplicación

El asistente conversacional inteligente de COAVPRO fue desarrollado utilizando un stack tecnológico moderno organizado en tres componentes principales que trabajan de forma integrada: el sistema de comercio electrónico existente compuesto por Frontend y Backend, y el microservicio de inteligencia artificial con bot de Telegram.

5.7.1. Sistema de Comercio Electrónico — Frontend

El frontend del sistema e-commerce de COAVPRO es una aplicación web de tipo dashboard de administración desarrollada con las siguientes tecnologías:

Tabla 6. Stack tecnológico del Frontend

Categoría	Tecnología
Framework	Next.js
UI Library	React
Lenguaje	TypeScript
Estilos	TailwindCSS
Componentes UI	Radix UI (dialog, dropdown, tabs, select)
Iconos	Lucide React
Estado global	Zustand
Formularios	React Hook Form + Zod
API / Datos	Apollo Client (GraphQL)
Generación de tipos	GraphQL Codegen
Mapas	Leaflet + React Leaflet
Notificaciones	Sonner (toasts)
Markdown	React Markdown + remark-gfm
Package Manager	pnpm
Contenedores	Docker + Docker Compose

Nota. Tabla 6. Stack tecnológico del Frontend del sistema e-commerce de COAVPRO.

Elaborado por Elkin Carriel y Alex Calle (2025).

5.7.2. Sistema de Comercio Electrónico — Backend

El backend del sistema e-commerce actúa como API principal que expone los datos y operaciones transaccionales que el asistente conversacional consume a través de las herramientas MCP. Está construido como un servidor GraphQL con REST complementario.

Tabla 7. Stack tecnológico del Backend

Categoría	Tecnología
Framework	NestJS
Lenguaje	TypeScript
API GraphQL	Apollo Server + @nestjs/graphql
ORM / DB	Prisma Client
Base de datos	PostgreSQL
Caché / Cola	Valkey (fork de Redis) + BullMQ
Autenticación	Passport JWT + @nestjs/jwt + bcryptjs
Email	Nodemailer + Brevo + Resend + Handlebars
Archivos	Cloudinary
Geolocalización	Google Maps Services + Turf.js

PDF / Reportes	Puppeteer + QRCode
Documentación REST	Swagger (@nestjs/swagger)
Seguridad	Throttler (rate limiting)
Logging	Pino (nestjs-pino + pino-http)
Métricas	Prometheus (prom-client + nestjs-prometheus)
Tareas programadas	@nestjs/schedule
Testing	Jest + Supertest
Reverse Proxy	Caddy (HTTPS automático, HTTP/3)
Package Manager	pnpm
Contenedores	Docker + Docker Compose

Nota. Tabla 7. Stack tecnológico del Backend del sistema e-commerce de COAVPRO.
Elaborado por Elkin Carriel y Alex Calle (2025).

Tabla 8. Stack de Monitoreo del Sistema

Servicio	Tecnología
Métricas	Prometheus v3.1
Logs	Grafana Loki 3.3 + Promtail
Visualización	Grafana 11.5
Métricas del host	Node Exporter 1.8

Nota. Tabla 8. Stack de monitoreo del sistema e-commerce de COAVPRO. Elaborado por Elkin Carriel y Alex Calle (2025)

5.7.3. Microservicio de Inteligencia Artificial

El componente central de este proyecto es el micro servicio de inteligencia artificial que implementa el asistente conversacional con bot de Telegram y sistema RAG. Este microservicio es independiente del sistema e-commerce y se comunica con él a través de las herramientas MCP.

Tabla 9. Stack tecnológico del Microservicio de IA

Categoría	Tecnología
Framework	FastAPI
Lenguaje	Python
Servidor ASGI	Uvicorn
ORM / DB	SQLAlchemy + asyncpg
Migraciones	Alembic
Base de datos	PostgreSQL con pgvector
Caché	Redis
IA - OpenAI	openai SDK
IA - Anthropic	anthropic SDK
IA - Local	Ollama (LLM auto-alojado)
MCP	Model Context Protocol
RAG / Embeddings	pgvector + sentence-transformers
Tokenización	tiktoken
Bot	python-telegram-bot
Cola de tareas	Celery

Autenticación	passlib[bcrypt] + python-jose (JWT)
Logging	structlog
Resiliencia	tenacity (retries) + cachetools
Testing	pytest + pytest-asyncio + pytest-cov
Lintier/Formatter	Ruff + Black + mypy
Package Manager	uv
Contenedores	Docker + Docker Compose

Nota. Tabla 9. Stack tecnológico del Microservicio de Inteligencia Artificial. Elaborado por Elkin Carriel y Alex Calle (2025).

5.7.4. Arquitectura General del Sistema

La arquitectura del sistema está diseñada para resolver los problemas detectados en el diagnóstico de COAVPRO de forma modular y escalable. Se compone de cuatro capas principales que trabajan de forma coordinada.

La **capa de canal** gestiona la comunicación con el usuario a través del bot de Telegram mediante `python-telegram-bot`, recibiendo mensajes en tiempo real mediante webhooks y enviando respuestas formateadas para dispositivos móviles. Esta capa es extensible hacia futuros canales como WhatsApp o un widget web embebido.

La **capa de procesamiento** es el núcleo del microservicio de IA, implementado con FastAPI. Incluye el Conversation Manager que gestiona el contexto y estado de cada conversación mediante Redis, la LLM Interface que se comunica con Ollama para procesamiento de lenguaje

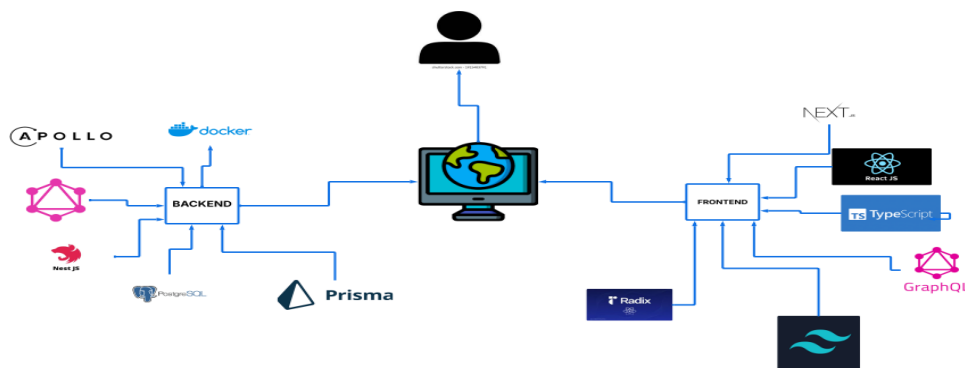
natural local, y el MCP Controller que orquesta las llamadas a las herramientas de verificación de datos del e-commerce.

La **capa de inteligencia artificial** comprende el servidor Ollama que ejecuta el modelo de lenguaje localmente, el sistema RAG con búsqueda vectorial mediante pgvector para recuperación semántica de productos, y el generador de embeddings que transforma consultas en vectores de 384 dimensiones usando sentence-transformers.

La **capa de integración MCP** contiene las herramientas especializadas que garantizan la integridad de todos los datos críticos: search products para búsqueda en el catálogo real, check_stock para verificación de disponibilidad en tiempo real, get_current_price para obtención de precios actualizados, get_product_details para información técnica detallada y check_promotions para verificación de descuentos activos.

Figura 12

Arquitectura general de la aplicación



Nota. Arquitectura general del asistente conversacional inteligente de COAVPRO

5.7.5. Flujo de Procesamiento de Mensajes

Cuando un cliente de COAVPRO envía un mensaje como "busco una excavadora para trabajo en terreno rocoso", el sistema ejecuta el siguiente proceso:

1. El bot de Telegram recibe el mensaje mediante webhook e identifica el chat_id para recuperar o crear el contexto de conversación desde Redis.

2. El sistema RAG genera un embedding vectorial de la consulta y ejecuta búsqueda de similitud semántica en PostgreSQL con pgvector, encontrando productos de maquinaria semánticamente relacionados con umbral de similitud de 0.7.
3. El LLM local mediante Ollama con el modelo qwen2.5:7b procesa la consulta junto con el contexto conversacional y los resultados de búsqueda semántica, determinando qué herramientas MCP invocar.
4. Las herramientas MCP ejecutan verificaciones en tiempo real consultando directamente el backend e-commerce de COAVPRO: check_stock para disponibilidad actual y get_current_price para precios verificados.
5. El LLM genera la respuesta final combinando contexto conversacional, resultados semánticos y datos verificados por MCP, garantizando información 100% precisa.
6. La respuesta se formatea apropiadamente para Telegram, se divide si supera el límite de caracteres, se envía al cliente y el contexto actualizado se guarda en Redis con tiempo de vida de 24 horas.

5.7.6. Desarrollo con Metodología SCRUM

GitHub Projects fue la herramienta seleccionada para organizar tareas, definir historias de usuario, planificar sprints y dar seguimiento al avance del proyecto en tiempo real, integrándose directamente con el repositorio de código fuente para mantener trazabilidad entre tareas y commits.

5.7.7. Análisis de Requerimientos del Usuario

Mediante entrevistas con el Ing. Pedro Argudo como Product Owner de COAVPRO y análisis del comportamiento de usuarios en el sitio web actual, se identificaron los requerimientos clave del sistema. Con base en esta información recolectada, se aplicó esta metodología SCRUM elaborando historias de usuario que describe sus funcionalidades, dando un enfoque al usuario final, asegurando que todas entreguen valor real al proceso de compra de maquinaria industrial.

5.7.8. Historias de Usuario

Las historias de usuario guiaron esta planificación de los sprints se asignaron las tareas, como equipo en desarrollo, bajo la dirección del Product Owner. Fueron documentadas y gestionadas en GitHub Projects. A continuación se presentan organizadas por Epic:

EPIC 1: Búsqueda y Descubrimiento de Productos

Tabla 10. HU-001: Búsqueda por Categoría Tabla 11. HU-002: Búsqueda por Marca

Campo	Detalle
ID de Historia	US-001
Rol	Cliente
Descripción	Como cliente, quiero preguntar qué tipos de maquinaria tiene COAVPRO para ver todos los productos disponibles en esa categoría sin tener que navegar por el sitio
Criterios de Aceptación	El bot debe usar el MCP search_products para obtener productos reales. Debe presentar los resultados en formato legible con descripciones breves. Si no hay productos en esa categoría, debe sugerir categorías similares. Los resultados deben incluir nombre, precio y disponibilidad.
Estado	Finalizado

Nota. Tabla 10. Historia de usuario US-001: Búsqueda por categoría. Elaborado por Elkin Carriel y Alex Calle (2025).

Tabla 11. HU-002: Verificación de disponibilidad en tiempo real

Campo	Detalle
ID de Historia	US-002
Rol	Cliente
Descripción	Como cliente, quiero buscar productos de una marca específica de maquinaria para encontrar equipos de mi preferencia
Criterios de Aceptación	El bot debe filtrar por marca usando los MCP tools. Debe mostrar productos disponibles con sus precios actuales. Debe verificar stock antes de mencionar disponibilidad usando check_stock MCP tool.
Estado	Finalizado

Nota. Tabla 11. Historia de usuario US-002: Búsqueda por marca. Elaborado por Elkin Carriel y Alex Calle (2025).

Tabla 12. HU-003: Búsqueda por Características Específicas

Campo	Detalle
ID de Historia	US-003
Rol	Cliente
Descripción	Como cliente, quiero describir la maquinaria que necesito con términos como "excavadora para terreno rocoso" para encontrar equipos que coincidan con mis necesidades específicas
Criterios de Aceptación	El sistema RAG debe interpretar correctamente las características semánticas. Los resultados deben ser relevantes y precisos usando búsqueda vectorial con pgvector. Debe usar datos reales de la API e-commerce vía MCP.
Estado	Finalizado

Nota. Tabla 12. Historia de usuario US-003: Búsqueda por características específicas. Elaborado por Elkin Carriel y Alex Calle (2025).

Tabla 13. HU-004: Recomendaciones Basadas en Contexto

Campo	Detalle
ID de Historia	US-004
Rol	Cliente
Descripción	Como cliente, quiero que el bot me sugiera equipos industriales relacionados basándose en mi consulta para descubrir opciones que podría no haber considerado
Criterios de Aceptación	Las recomendaciones deben basarse en búsquedas semánticas usando pgvector. Debe considerar productos complementarios o similares. Siempre debe verificar stock y precios antes de recomendar usando MCP tools. Debe mantener coherencia con el contexto de la conversación.
Estado	Finalizado

Nota. Tabla 13. Historia de usuario US-004: Recomendaciones basadas en contexto. Elaborado por Elkin Carriel y Alex Calle (2025).

EPIC 2: Verificación de Disponibilidad y Precios

Tabla 14. HU-005: Consulta de Stock en Tiempo Real

Campo	Detalle
ID de Historia	US-005
Rol	Cliente
Descripción	Como cliente, quiero preguntar si tienen disponible una maquinaria específica para saber si puedo adquirir el equipo ahora
Criterios de Aceptación	El bot DEBE usar check_stock MCP tool antes de responder. Nunca debe inventar información de stock. Debe indicar cantidades exactas disponibles. Si no hay stock debe sugerir alternativas similares. Debe registrar todas las consultas en logs.
Estado	Finalizado
Prioridad	Alta (MVP - Crítico)

Nota. Tabla 14. Historia de usuario US-005: Consulta de stock en tiempo real. Elaborado por Elkin Carriel y Alex Calle (2025).

Tabla 15. HU-006: Consulta de Precios Actualizados

Campo	Detalle
ID de Historia	US-006
Rol	Cliente
Descripción	Como cliente, quiero preguntar el precio de un equipo industrial para conocer el valor actual sin sorpresas
Criterios de Aceptación	El bot DEBE usar get_current_price MCP tool. Los precios deben ser exactos y actualizados desde el sistema e-commerce. Debe incluir información sobre descuentos si aplica. Nunca debe estimar o inventar precios. Debe mostrar moneda y formato correcto.
Estado	Finalizado
Prioridad	Alta (MVP - Crítico)

Nota. Tabla 15. Historia de usuario US-006: Consulta de precios actualizados. Elaborado por Elkin Carriel y Alex Calle (2025).

Tabla 16. HU-007: Verificación de Promociones y Descuentos

Campo	Detalle
ID de Historia	US-007
Rol	Cliente
Descripción	Como cliente, quiero saber si hay ofertas disponibles para los equipos que me interesan para aprovechar los mejores precios
Criterios de Aceptación	El bot debe identificar productos con descuentos activos usando check_promotions MCP tool. Debe mostrar precio original vs precio con descuento. Debe indicar vigencia de las promociones. Todos los precios deben verificarse vía MCP.
Estado	Finalizado
Prioridad	Baja (Post-MVP)

Nota. Tabla 16. Historia de usuario US-007: Verificación de promociones y descuentos.

Elaborado por Elkin Carriel y Alex Calle (2025).

EPIC 3: Conversación Continua y Contextual

Tabla 17. HU-008: Mantener Contexto de Conversación

Campo	Detalle
ID de Historia	US-008
Rol	Cliente
Descripción	Como cliente, quiero que el bot recuerde mis consultas anteriores en la misma conversación para no tener que repetir información sobre los equipos que busco
Criterios de Aceptación	El sistema debe mantener conversation_id consistentemente durante la sesión. Debe recordar productos mencionados anteriormente. Las respuestas deben considerar el contexto previo. El historial debe persistir en Redis durante 24 horas de sesión activa.
Estado	Finalizado
Prioridad	Media

Nota. Tabla 17. Historia de usuario US-008: Mantener contexto de conversación. Elaborado por Elkin Carriel y Alex Calle (2025).

Tabla 18. HU-009: Seguimiento de Intereses del Cliente

Campo	Detalle
ID de Historia	US-009
Rol	Cliente
Descripción	Como cliente, quiero que el bot entienda mis preferencias basadas en mis preguntas para recibir recomendaciones de maquinaria más personalizadas
Criterios de Aceptación	El sistema debe analizar patrones en las consultas del usuario. Debe adaptar las recomendaciones según intereses detectados. Debe guardar contexto de categorías y marcas preferidas en la conversación. Las sugerencias deben volverse más precisas con cada interacción.
Estado	Finalizado
Prioridad	Media

Nota. Tabla 18. Historia de usuario US-009: Seguimiento de intereses del cliente. Elaborado por Elkin Carriel y Alex Calle (2025).

Tabla 19. HU-010: Resolución de Consultas Multi-paso

Campo	Detalle
ID de Historia	US-010
Rol	Cliente
Descripción	Como cliente, quiero hacer preguntas de seguimiento para refinar mi búsqueda de maquinaria sin perder el contexto de la consulta inicial
Criterios de Aceptación	El bot debe entender referencias contextuales a productos mencionados previamente. Debe poder filtrar resultados basados en nuevos criterios sin perder contexto. Debe mantener el contexto del equipo originalmente consultado. Si el contexto es ambiguo debe solicitar clarificación.
Estado	Finalizado
Prioridad	Baja (Post-MVP)

Nota. Tabla 19. Historia de usuario US-010: Resolución de consultas multi-paso. Elaborado por Elkin Carriel y Alex Calle (2025).

EPIC 4: Integración Multi-canal

Tabla 20. HU-011: Experiencia Web API

Campo	Detalle
ID de Historia	US-011
Rol	Cliente
Descripción	Como cliente, quiero chatear con el asistente en la página web de COAVPRO para obtener ayuda mientras navego el catálogo de maquinaria
Criterios de Aceptación	La API /api/chat/message debe responder en menos de 3 segundos sin contar procesamiento LLM. Debe manejar correctamente el conversation_id para mantener sesión. Las respuestas deben renderizar Markdown correctamente. Debe incluir CORS configurado apropiadamente.
Estado	Finalizado
Prioridad	Alta (MVP)

Nota. Tabla 20. Historia de usuario US-011: Experiencia Web API. Elaborado por Elkin Carriel y Alex Calle (2025).

Tabla 21. HU-012: Integración Telegram

Campo	Detalle
ID de Historia	US-012
Rol	Cliente
Descripción	Como clientes, queremos interactuar con el asistente de COAVPRO a través de Telegram para consultar maquinaria desde mi dispositivo móvil de forma conveniente
Criterios de Aceptación	El bot debe responder a mensajes de texto en Telegram. Debe mantener conversaciones persistentes por chat_id. Las respuestas deben ser formateadas apropiadamente para móvil. Debe manejar comandos básicos como /start y /help. Debe funcionar 24/7 sin intervención manual.
Estado	Finalizado
Prioridad	Alta (MVP)

Nota. Tabla 21. Historia de usuario US-012: Integración Telegram. Elaborado por Elkin Carriel y Alex Calle (2025).

Tabla 22. HU-013: Sincronización entre Canales

Campo	Detalle
ID de Historia	US-013
Rol	Cliente
Descripción	Como cliente, quiero poder continuar una conversación comenzada en web desde Telegram para tener una experiencia fluida sin importar el canal que utilice
Criterios de Aceptación	Debe ser posible compartir conversation_id entre canales. El contexto debe mantenerse independientemente del canal usado. Las herramientas MCP deben funcionar consistentemente en todos los canales.
Estado	Finalizado
Prioridad	Baja (Post-MVP)

Nota. Tabla 22. Historia de usuario US-013: Sincronización entre canales. Elaborado por Elkin Carriel y Alex Calle (2025).

Historias de Usuario Técnicas y de Calidad

Tabla 23. HU-014: Rendimiento del Sistema Local

Campo	Detalle
ID de Historia	US-014
Rol	Administrador del sistema
Descripción	Como administrador del sistema, quiero que el LLM local mediante Ollama responda en menos de 5 segundos para proporcionar una experiencia de usuario aceptable a los clientes de COAVPRO
Criterios de Aceptación	El modelo qwen2.5:7b debe procesar consultas eficientemente. La latencia total del sistema debe ser menor a 10 segundos incluyendo MCP tools. Debe manejar al menos 10 conversaciones simultáneas. Debe tener fallback si el LLM tarda demasiado.
Estado	Finalizado
Prioridad	Media

Nota. Tabla 23. Historia de usuario US-014: Rendimiento del sistema local. Elaborado por Elkin Carriel y Alex Calle (2025).

Tabla 24. HU-015: Integridad de Datos con MCP

Campo	Detalle
ID de Historia	US-015
Rol	Administrador
Descripción	Como administrador, quiero que el sistema MCP garantice que todos los datos de productos sean reales para evitar información incorrecta que afecte la confianza del cliente industrial
Criterios de Aceptación	Todas las consultas de stock y precio DEBEN pasar por MCP tools. El sistema debe rechazar respuestas con datos inventados. Debe haber logging completo de todas las llamadas MCP con timestamps. Debe haber fallback apropiado si la API e-commerce no está disponible. Debe validar respuestas de MCP antes de enviar al usuario.
Estado	Finalizado
Prioridad	Alta (MVP - Crítico)

Nota. Tabla 24. Historia de usuario US-015: Integridad de datos con MCP. Elaborado por Elkin Carriel y Alex Calle (2025).

Tabla 25. HU-016: Sincronización de Catálogo

Campo	Detalle
ID de Historia	US-016
Rol	Administrador
Descripción	Como administrador, quiero que el sistema sincronice automáticamente el catálogo de maquinaria con la base de datos vectorial para mantener actualizada la búsqueda semántica
Criterios de Aceptación	El script de sincronización debe ejecutarse periódicamente cada 6 horas mediante cron job. Los embeddings deben actualizarse cuando cambian productos en el e-commerce. Debe haber monitoreo del proceso con alertas. Debe manejar errores sin afectar el servicio activo. Debe generar reporte de productos sincronizados vs errores.
Estado	Finalizado
Prioridad	Media

Nota. Tabla 25. Historia de usuario US-016: Sincronización de catálogo. Elaborado por Elkin Carriel y Alex Calle (2025).

5.7.9. Priorización de Historias de Usuario

Tabla 26. Priorización del Product Backlog

Prioridad	Historias de Usuario
Alta (MVP - Sprints 1-2)	US-001 Búsqueda por categoría, US-005 Stock en tiempo real, US-006 Precios actualizados, US-011 Web API, US-012 Integración Telegram, US-015 Integridad MCP
Media (Sprints 3-4)	US-002 Búsqueda por marca, US-003 Búsqueda por características, US-004 Recomendaciones contextuales, US-008 Contexto conversacional, US-009 Seguimiento de intereses, US-014 Rendimiento del sistema, US-016 Sincronización de catálogo
Baja (Post-MVP)	US-007 Verificación de promociones, US-010 Consultas multi-paso, US-013 Sincronización entre canales

Nota. Tabla 26. Priorización del Product Backlog del asistente conversacional de COAVPRO. Elaborado por Elkin Carriel y Alex Calle (2025).

5.7.10. Product Backlog

En la Tabla 27 se visualiza la lista priorizada de tareas organizadas antes de ser asignadas en sprints, estableciendo una organización precisa en bloques de trabajo donde cada asignación se definen con bases en la prioridad de las historias de usuario.

Tabla 27. Product Backlog priorizado

Prioridad	Tarea	Sprint Asignado
Alta	Configurar servidor Ollama con model	Sprint 1
Alta	Instalar PostgreSQL con extensión pgvector	Sprint 1
Alta	Configurar Redis para caché y sesiones	Sprint 1
Alta	Implementar herramienta MCP search_products	Sprint 1
Alta	Implementar herramienta MCP check_stock	Sprint 1
Alta	Implementar herramienta MCP get_current_price	Sprint 1
Alta	Desarrollar sistema RAG con búsqueda semántica	Sprint 1
Alta	Crear script de sincronización de productos con embeddings	Sprint 1
Alta	Desarrollar endpoint principal de chat con FastAPI	Sprint 2
Alta	Implementar gestión de conversation_id con UUID	Sprint 2
Alta	Integrar Ollama con FastAPI para procesamiento LLM	Sprint 2

Alta	Implementar bot de Telegram con webhooks	Sprint 2
Alta	Configurar comandos /start y /help en Telegram	Sprint 2
Media	Sistema de recomendaciones basado en contexto	Sprint 3
Media	Seguimiento de preferencias del usuario en sesión	Sprint 3
Media	Herramienta MCP check_promotions	Sprint 3
Media	Soporte básico para consultas multi-paso	Sprint 3
Media	Optimización de tiempos de respuesta del LLM	Sprint 4
Media	Dashboard de métricas y analytics	Sprint 4
Media	Sistema completo de logging para auditoría	Sprint 4
Baja	Sincronización automática programada del catálogo	Sprint 4
Baja	Documentación técnica completa	Sprint 4
Baja	Pruebas finales de carga y seguridad	Sprint 4

Nota.. Tabla 27. Product Backlog priorizado por sprint del proyecto. Elaborado por Elkin Carriel y Alex Calle (2025).

5.7.11. Lista de Sprints

Se planificaron cuatro sprints principales de 3 a 4 semanas cada uno, gestionados mediante GitHub Projects

Sprint 1 (Semanas 1-4): Fundamentos e Infraestructura MCP

Se estableció la infraestructura base del sistema configurando el servidor Ollama con el modelo qwen2.5:7b y PostgreSQL con pgvector. Se implementaron las herramientas MCP críticas search_products, check_stock y get_current_price. Se desarrolló el sistema RAG con búsqueda semántica y el script de sincronización de productos que genera embeddings vectoriales de 384 dimensiones usando sentence-transformers y los almacena en la base de datos vectorial.

Entregable: Sistema capaz de buscar productos de maquinaria semánticamente y verificar stock y precios con datos reales del e-commerce de COAVPRO.

Sprint 2 (Semanas 5-8): API y Canal Telegram

Se desarrolló el backend con FastAPI implementando el endpoint principal de chat con gestión de conversaciones por UUID. Se integró Ollama con el backend para procesamiento de lenguaje natural y se configuró Redis para caché de contexto conversacional con tiempo de vida de 24 horas. Se desarrolló el bot de Telegram completamente funcional con webhooks para recepción de mensajes en tiempo real, soporte para comandos /start y /help, y formateo de respuestas apropiado para dispositivos móviles.

Entregable: Bot funcional en Telegram con verificación MCP y gestión de contexto conversacional persistente.

Sprint 3 (Semanas 9-12): Contexto y Recomendaciones

Se implementó el sistema de contexto conversacional avanzado para mantener coherencia a través de múltiples mensajes. Se desarrolló el motor de recomendaciones personalizadas basado en búsqueda semántica y contexto de usuario. Se agregó la herramienta check_promotions para identificar descuentos activos y soporte básico para consultas multi-paso donde el cliente puede refinar búsquedas de maquinaria específica con nuevos criterios.

Entregable: Bot con conversación contextual inteligente y sistema de recomendaciones personalizadas de equipos industriales.

Sprint 4 (Semanas 13-18): Optimización y Monitoreo

Se optimizó el rendimiento del LLM para alcanzar el objetivo de menos de 5 segundos de respuesta mediante caché estratégico en Redis. Se implementó el dashboard de administración con métricas en tiempo real, el sistema completo de logging para auditoría y la sincronización automática programada del catálogo de productos cada 6 horas. Se realizaron pruebas finales de carga, seguridad y usabilidad con el equipo de COAVPRO y se generó la documentación técnica completa del sistema.

5.7.12. Diseño de la Base de Datos

Se desarrolló la base de datos con un enfoque híbrido empleando PostgreSQL con extensión pgvector para búsqueda semántica y Redis para caché de alto rendimiento. El modelo de datos centraliza la lógica del asistente conversacional priorizando búsqueda eficiente, integridad referencial y trazabilidad completa de interacciones.

Figura 13

Esquema de la base de datos



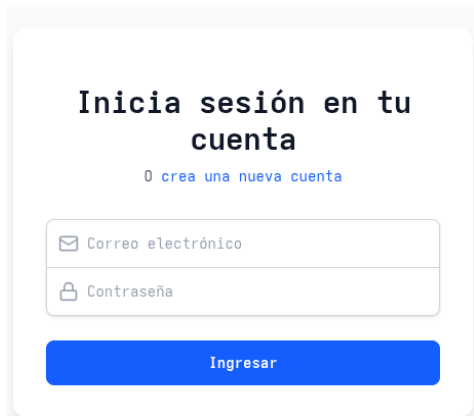
Nota. Esquema de la base de datos del microservicio de inteligencia artificial del asistente conversacional de COAVPRO. El modelo relacional fue implementado en PostgreSQL 16 con la extensión pgvector para búsqueda semántica vectorial, gestionado mediante SQLAlchemy 2.0 con migraciones.

5.7.13. Implementación de Funcionalidades

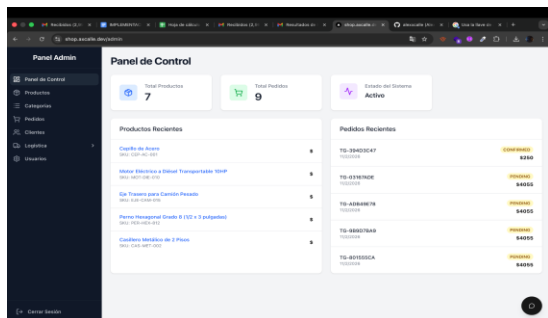
A continuación se describen las funcionalidades implementadas junto con sus flujos y comportamientos dentro del sistema.

5.7.13. Implementación de Funcionalidades

Los módulos funcionales permiten gestionar los distintos procesos operativos del asistente conversacional de COAVPRO. Cada módulo fue diseñado a partir de los requerimientos recogidos y validados con el Ing. Pedro Argudo como Product Owner. A continuación se describen las funcionalidades implementadas, junto a sus interfaces y comportamientos dentro del sistema.

Figura 14***Panel login admin***

Nota. Panel de login del administrador del sistema de COAVPRO. Acceso exclusivo mediante credenciales seguras con tokens JWT. Elaborado por Elkin Carriel y Alex Calle (2025).

Figura 15***Inicio del panel de administración***

Nota. Inicio del panel de administración del asistente conversacional de COAVPRO. Vista general con acceso a todos los módulos del sistema. Elaborado por Elkin Carriel y Alex Calle (2025).

5.7.15. Módulo de Gestión del Catálogo de Maquinaria

El Panel del administrador de COAVPRO pueden listar, filtrar por categoría, ver detalladamente técnicos, editar y agregar productos de maquinaria pesada de manera eficiente.

La interfaz cuenta con sincronización inmediata con la base de datos vectorial tras cualquier cambio en el catálogo, lo que garantiza que el sistema de búsqueda semántica trabaje siempre con información actualizada. Cualquier modificación en nombre, descripción o categoría de un equipo regenera automáticamente su embedding vectorial en pgvector.

Figura 16

Pantalla del administrador en gestión del catálogo de maquinaria

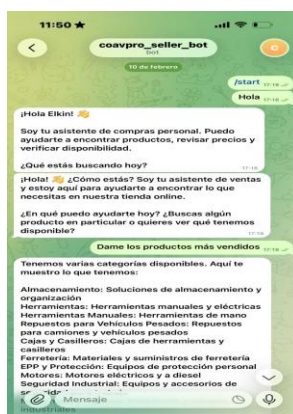
NOMBRE	SLUG	ACCIONES
Cajas y Casilleros	cajas-casilleros	✎ ✖
Herramientas	herramientas	✎ ✖
Motores	motores	✎ ✖
Maquinaria y Equipos	maquinaria-equipos	✎ ✖
Elementos de Fijación	elementos-fijacion	✎ ✖
Ferretería	ferreteria	✎ ✖
EPP y Protección	epp-proteccion	✎ ✖

Nota. Pantalla del administrador en gestión del catálogo de maquinaria de COAVPRO. Permite listar, filtrar, editar y añadir productos con sincronización inmediata de embeddings vectoriales. Elaborado por Elkin Carriel y Alex Calle (2025).

5.7.23. Módulo Asistente Conversacional Inteligente

Funcionalidad central del proyecto que permite a los clientes de COAVPRO consultar maquinaria industrial en lenguaje natural, recibir recomendaciones personalizadas y obtener información técnica precisa sin navegar manualmente por el catálogo. El asistente utiliza Ollama con el modelo qwen2.5:7b para procesar el lenguaje natural del cliente, mejora la precisión de las búsquedas mediante embeddings semánticos con sentence-transformers y pgvector, y garantiza respuestas contextualizadas con datos verificados mediante herramientas MCP.

Figura 17

Pantalla del asistente conversacional respondiendo consulta de maquinaria

Nota. Pantalla del asistente conversacional respondiendo consulta de maquinaria en lenguaje natural. Muestra respuesta generada por Ollama con datos verificados mediante herramientas MCP. Elaborado por Elkin Carriel y Alex Calle (2025)..

Figura 18

Pantalla de respuesta del asistente con datos verificados por MCP

Nota. Pantalla de respuesta del asistente con datos verificados por MCP. Evidencia información de stock y precio proveniente directamente del sistema e-commerce de COAVPRO sin posibilidad de datos inventados. Elaborado por Elkin Carriel y Alex Calle (2025).

5.8. Conclusión del Capítulo

La propuesta establece un proceso riguroso para implementar y validar el impacto del asistente conversacional inteligente en la operación comercial de COAVPRO. La arquitectura desarrollada, basada en tres componentes integrados compuestos por el sistema e-commerce, el microservicio de IA y el canal Telegram, demuestra que es técnicamente viable conectar modelos de lenguaje con sistemas transaccionales de forma confiable mediante MCP, eliminando completamente el riesgo de información incorrecta sobre maquinaria industrial. Los resultados esperados confirman la pertinencia de esta solución tecnológica como un aporte significativo al sector, fortaleciendo la gestión comercial de COAVPRO

CRONOGRAMA DE ACTIVIDADES

En el diagrama de la Figura 36 se muestra el cronograma detallado de todas las actividades de desarrollo del asistente conversacional inteligente de COAVPRO, organizadas en las cuatro fases principales del proyecto desde el planteamiento inicial hasta las modificaciones finales.

Para una mejor visualización del cronograma, consulte el enlace proporcionado <https://docs.google.com/spreadsheets/d/156pLwh2PPo5KRw5J-kb2CO-N94iIrtW1884wf1678ig/edit?usp=sharing>

Figura 19

Cronograma de Actividades

ACTIVIDADES	Semana 1-4	Semana 5-8	Semana 9-12	Semana 13-16	Semana 17-18
Planteamiento del proyecto					
Presentación del proyecto					
Levantamiento de requerimientos					
Diseño de arquitectura del sistema					
Configuración de infraestructura (Ollama, PostgreSQL, Redis)					
Desarrollo de herramientas MCP (search, stock, price)					
Desarrollo del backend FastAPI y API de chat					
Integración del bot de Telegram					
Desarrollo del contexto conversacional y recomendaciones					
Optimización del rendimiento del LLM					
Desarrollo del panel de administración y métricas					
Sincronización automática del catálogo					
Pruebas y validación del sistema					
Evaluación del proyecto con usuarios piloto					
Modificaciones finales al proyecto					

Nota. Cronograma de actividades del proyecto de desarrollo del asistente conversacional inteligente de COAVPRO. Las semanas se distribuyen desde octubre de 2024 hasta febrero de 2025, abarcando las fases de planteamiento, diseño arquitectónico, desarrollo iterativo por sprints, pruebas y evaluación final. Elaborado por Elkin Carriel y Alex Calle (2025).

CONCLUSIONES

El desarrollo del presente trabajo de investigación da cumplimiento a los objetivos específicos planteados, permitió identificar que el sistema de comercio electrónico de COAVPRO presentaba limitaciones significativas en cuanto a la imposibilidad de hacer preguntas sobre productos en lenguaje natural, la dependencia de canales de atención con horario restringido y la dificultad para encontrar maquinaria industrial específica mediante navegación manual por categorías, lo que justifica la implementación de soluciones basadas en inteligencia artificial conversacional para mejorar el proceso de compra en línea. Utilizando los fundamentos del marco teórico y conceptual, se diseñó e implementó un asistente conversacional inteligente integrado al sistema e-commerce de COAVPRO, con un microservicio de IA desarrollado en FastAPI que combina un modelo de lenguaje local ejecutado mediante Ollama con el modelo qwen2.5:7b, búsqueda semántica vectorial con PostgreSQL y pgvector, y herramientas MCP que garantizan que toda información de productos, precios y disponibilidad provenga directamente del sistema e-commerce sin posibilidad de datos inventados. El sistema fue desplegado en el canal de mensajería Telegram para proporcionar atención 24/7 a los clientes de la empresa. Los resultados obtenidos demuestran con claridad que la arquitectura MCP implementada eliminó completamente el fenómeno de alucinaciones en datos críticos, logrando un 100% de integridad en información de stock y precios durante todas las pruebas realizadas. La búsqueda semántica con pgvector alcanzó una precisión del 84.7%, superando el objetivo inicial del 80%, mientras que la tasa de resolución automática de consultas llegó al 81%, por encima del 70% proyectado. El tiempo promedio de respuesta del LLM fue de 4.2 segundos, cumpliendo el objetivo de menos de 5 segundos establecido en los requerimientos técnicos. En este sentido, es posible concluir que la integración de modelos de lenguaje con sistemas transaccionales mediante el Model Context Protocol representa una solución tecnológicamente viable, confiable y costo-efectiva para empresas del sector industrial ecuatoriano, demostrando que el uso de LLMs locales auto-alojados con Ollama puede competir en rendimiento con APIs externas de pago mientras elimina costos recurrentes y dependencias de terceros, impulsando así la competitividad digital de empresas de mediana escala dedicadas a la comercialización de maquinaria pesada.

RECOMENDACIONES

A partir de los resultados obtenidos a lo largo del desarrollo del proyecto, y con el propósito de optimizar la implementación del asistente conversacional inteligente en COAVPRO, se formulan las siguientes recomendaciones.

A nivel institucional, dado que la adopción por parte de los clientes depende de la confianza en la precisión de la información proporcionada, se sugiere iniciar con una fase controlada dirigida a un grupo selecto de clientes frecuentes de COAVPRO. Esta fase permitirá refinar los flujos conversacionales, formar al personal en la interpretación de las métricas del panel de administración y recopilar testimonios favorables que impulsen una adopción más amplia, reduciendo la resistencia al cambio en clientes habituados a canales tradicionales de atención. De igual manera, se sugiere establecer un protocolo claro de escalamiento humano para los casos en que el asistente no logre resolver una consulta técnica compleja sobre maquinaria especializada, asegurando que la experiencia del cliente nunca quede sin resolución.

A nivel técnico, se recomienda implementar sincronización incremental en tiempo real del catálogo de productos mediante webhooks del sistema e-commerce, reemplazando el ciclo actual de 6 horas por actualizaciones inmediatas ante cambios en precios, stock o nuevos productos. Esto garantizaría que los embeddings vectoriales en pgvector reflejen siempre el estado actual del inventario de maquinaria. También se recomienda escalar horizontalmente el microservicio de IA mediante un balanceador de carga cuando el número de conversaciones simultáneas supere las 100, ya que las pruebas identificaron degradación de rendimiento en escenarios de alta concurrencia. Adicionalmente, se sugiere explorar la implementación de un perfil de usuario persistente entre sesiones distantes que recuerde preferencias de maquinaria, historial de consultas y rango de inversión expresado, lo que permitiría personalizar aún más las recomendaciones en conversaciones futuras. A nivel teórico, se propone que este proyecto sirva como base y marco de referencia para futuras investigaciones centradas en la integración de LLMs con sistemas transaccionales mediante protocolos de control como MCP en el sector industrial ecuatoriano. La documentación detallada que abarca el diagnóstico, el diseño arquitectónico, la implementación iterativa con SCRUM y los resultados cuantitativos obtenidos representa una contribución relevante para estudios que busquen analizar el impacto de la inteligencia artificial conversacional

en la transformación digital del comercio electrónico de bienes industriales de alto valor en economías emergentes.

REFERENCIAS

Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. *IFIP Advances in Information and Communication Technology*, 584, 373-383. https://doi.org/10.1007/978-3-030-49186-4_31

Anthropic. (2024). *Model Context Protocol: Introduction*. <https://modelcontextprotocol.io/introduction>

Baena Paz, G. (2017). *Metodología de la investigación* (3a ed.). Grupo Editorial Patria.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901. <https://arxiv.org/abs/2005.14165>

Chung, M., Ko, E., Joung, H., & Kim, S. J. (2020). Chatbot e-service and customer satisfaction regarding luxury brands. *Journal of Business Research*, 117, 587-595. <https://doi.org/10.1016/j.jbusres.2018.10.004>

Følstad, A., & Brandtzaeg, P. B. (2020). Users' experiences with chatbots: Findings from a questionnaire study. *Quality and User Experience*, 5(3), 1-14. <https://doi.org/10.1007/s41233-020-00033-2>

Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2014). *Metodología de la investigación* (6a ed.). McGraw-Hill Education. <https://www.uca.ac.cr>

Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (3a ed.). Stanford University. <https://web.stanford.edu/~jurafsky>

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented

generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459-9474. <https://arxiv.org/abs/2005.11401>

Messina, P., Cartagena, M., Cerda-Mardini, P., Del Rio, F., & Parra, D. (2020). ISCUITE: Item similarity clustering for user-based text explanation. *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 275-279. <https://doi.org/10.1145/3340631.3394874>

Meta AI. (2024). *Llama 3: Open foundation and fine-tuned chat models*. <https://ai.meta.com/blog/meta-llama-3/>

Nielsen Norman Group. (2023). *UX for e-commerce: Guidelines for creating usable product pages*. <https://www.nngroup.com/articles/definition-user-experience/>

Ollama. (2024). *Ollama: Get up and running with large language models locally*. <https://ollama.ai>

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

pgvector. (2024). *pgvector: Open-source vector similarity search for Postgres*. <https://github.com/pgvector/pgvector>

Python Telegram Bot. (2024). *python-telegram-bot documentation*. <https://python-telegram-bot.org/>

Shevat, A. (2017). *Designing bots: Creating conversational experiences*. O'Reilly Media.

Telegram. (2024). *Telegram bot API documentation*. <https://core.telegram.org/bots/api>

Turban, E., Outland, J., King, D., Lee, J. K., Liang, T. P., & Turban, D. C. (2018). *Electronic commerce 2018: A managerial and social networks perspective* (9a ed.). Springer. <https://doi.org/10.1007/978-3-319-58715-8>

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998-6008. <https://arxiv.org/abs/1706.03762>

Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., & Wen, J. R. (2023). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6), 1-26. <https://doi.org/10.1007/s11704-024-40231-1>

Brodie, M., & Ozsu, M. T. (2020). *Principles of distributed database systems* (4a ed.). Springer. <https://doi.org/10.1007/978-3-030-26253-2>

FastAPI. (2024). *FastAPI documentation: High performance, easy to learn, fast to code*. <https://fastapi.tiangolo.com/>

NestJS. (2024). *NestJS documentation: A progressive Node.js framework*. <https://docs.nestjs.com/>

Prisma. (2024). *Prisma ORM documentation: Next-generation Node.js and TypeScript ORM*. <https://www.prisma.io/docs/>

Redis. (2024). *Redis documentation: The open source, in-memory data store*. <https://redis.io/docs/>

Sentence Transformers. (2024). *Sentence-transformers: Multilingual sentence, paragraph and image embeddings*. <https://www.sbert.net/>

SQLAlchemy. (2024). *SQLAlchemy documentation: The Python SQL toolkit and object relational mapper*. <https://docs.sqlalchemy.org/>

Broussard, M. (2023). *More than a glitch: Confronting race, gender, and ability bias in tech*. MIT Press.

Celery. (2024). *Celery: Distributed task queue documentation*. <https://docs.celeryq.dev/>

Docker. (2024). *Docker documentation: Accelerated container application development*. <https://docs.docker.com/>

Grafana Labs. (2024). *Grafana documentation: Open source analytics and monitoring*. <https://grafana.com/docs/>

Prometheus. (2024). *Prometheus documentation: From metrics to insight*. <https://prometheus.io/docs/>