

Modelado de Agente de Servicios con IA para un Operador Económico Autorizado
en Colombia

Laura Victoria Cárdenas Bohórquez
Sebastián Urrego Zapata

Directores:
Ing. Martha Isabel Villareal López

UNIVERSIDAD SANTO TOMAS
FACULTAD DE INGENIERÍA DE TELECOMUNICACIONES
ESPECIALIZACIÓN EN GESTIÓN DE SERVICIOS DE TECNOLOGÍAS DE LA
INFORMACIÓN
BOGOTÁ, 2025

“Quiero dedicar este proyecto y todo el curso de la especialización al tiempo, el esfuerzo y trabajo constante, y a la valentía de afrontar los retos de la vida y la academia”. Victoria C.

“A mí mismo: por no rendirme. Por las veces que dudé y seguí adelante. Esto es solo el comienzo”. Sebastián Urrego.

Dedicatoria

AGRADECIMIENTOS

“A mi familia, Dios y a la fe que nos permite creer que siempre podemos cumplirnos a nosotros y a nuestro propósito de vida. Y a Sebastián, mi gran compañero de camino.” Victoria C.

“A la universidad por permitirme desarrollar mis habilidades, especialmente a los profesores Fernando Prieto e Iván Camilo Castellanos, gracias por transmitir el conocimiento de forma que dan ganas seguir aprendiendo; a mis compañeros de clase que ya se convirtieron en mis hermanos y a Victoria, mi compañera de luchas académicas, que gran complemento y a mi familia que a pesar de la distancia son mi motor para seguir adelante.” Sebastián U.

1 Contenido

1. PROBLEMA	15
1.1 ARBOL DE PROBLEMAS	15
1.2 QUÉ SE QUIERE SOLUCIONAR	16
2. IDEACIÓN DE LA SOLUCIÓN	20
2.1 POR QUÉ SE PLANTEA AHORA LA SOLUCIÓN	20
2.2 SECTOR OBJETIVO	22
2.3 TENDENCIAS DEL SECTOR	24
2.4 ANALISIS DE MERCADO	26
2.5 ARBOL DE OBJETIVOS	28
2.7 INTRODUCCIÓN A LA SITUACIÓN DESEADA	30
2.8 PROPUESTA DE VALOR	34
2.8.1 Pilar 1: Seguridad	34
2.8.2 Pilar 2: Agilidad	34
2.8.3 Pilar 3: Cumplimiento	35
2.9 PERFIL DEL CLIENTE	35
2.9.1 Segmento 1: Grandes Corporaciones Importadoras/Exportadoras	35
2.9.2 Segmento 2: Medianas Empresas en Expansión Internacional	36
2.9.3 Segmento 3: PYMES y Emprendedores Exportadores	36
2.10 MAPA DE VALOR	36
3. ANÁLISIS DE LAS ALTERNATIVAS TÉCNICAS PARA SOLUCIONAR EL PROBLEMA	39
3.1 Bloque 1: Modelos de Lenguaje de Gran Escala (LLM)	39
3.2 Bloque 2: Infraestructura de Procesamiento	40
3.3 Bloque 3: Gestión de Conocimiento (Bases de Datos)	41
3.4 Bloque 4: Plataforma de Orquestación de Flujos de Trabajo	42

3.5	Bloque 5: Canal de Comunicación con Usuarios	42
3.6	Bloque 6: Seguridad y Cumplimiento Normativo	43
3.7	Arquitectura Técnica Propuesta	44
4.	MODELO DE NEGOCIO	45
4.1	<i>PROPUESTA DE MODELO DE NEGOCIO</i>	45
4.2	MODELOS DE MONETIZACIÓN	47
4.2.1	Alternativa 1: Modelo de Suscripción por Niveles	47
4.2.2	Alternativa 2: Modelo Basado en Uso	47
4.2.3	Alternativa 3: Modelo Híbrido	47
4.3	<i>VALIDACIÓN DEL MODELO DE NEGOCIO</i>	48
4.3.1	Marco de Validación	48
4.3.2	Métricas de Éxito	49
5.	PROPUESTA DE LA SOLUCIÓN TECNOLÓGICA	50
5.1	Componente de Orquestación: n8n	50
5.2	Arquitectura de la Solución	51
5.2.1	Capa de Interacción Multicanal	51
5.2.2	Capa de Orquestación y Automatización	51
5.2.3	Capa de Procesamiento de Lenguaje Natural	52
5.2.4	Capa de Gestión de Conocimiento	52
5.2.5	Capa de Infraestructura Local	52
5.2.6	Capa de Cumplimiento Normativo	53
5.3	Flujo Operativo del Sistema	53
5.4	Funcionalidades del MVP	54
5.5	Modelado General del Workflow	58
6.	ANÁLISIS DEL PROCESO DE TRANSFORMACIÓN DIGITAL	63
6.1	Marco de Análisis: Modelo de Madurez Digital	63
6.2	Dimensiones de Transformación Evaluadas	63
6.3	Evaluación del Nivel de Madurez Digital	65

7. ASPECTOS LEGALES Y CONTRATACIÓN	67
7.1 Cumplimiento Normativo de Protección de Datos Personales	67
7.2 Regulación Aduanera y Validez de Asesorías Automatizadas	68
7.2 Responsabilidad Civil y Profesional	69
7.3 Propiedad Intelectual y Licenciamiento	69
7.3 Marco Regulatorio de Contratación Electrónica	70
CONCLUSIONES	71
REFLEXIÓN FINAL	74
REFERENCIAS	75
LISTA DE TABLAS	80
LISTA DE FIGURAS	81

ACRÓNIMOS

ACRÓNIMO	SIGNIFICADO
OEA	Operador Económico Autorizado
SIA	Sociedad de Intermediación Aduanera
IA	Inteligencia Artificial
DIAN	Dirección de Impuestos y Aduanas Nacionales
OMC	Organización Mundial del Comercio
MinTIC	Ministerio de Tecnologías de la Información y las Comunicaciones
UIT	Unión Internacional de Telecomunicaciones
DNP	Departamento Nacional de Planeación
ANALDEX	Asociación Nacional de Comercio Exterior
TIC	Tecnologías de la Información y Comunicación
LLM	Large Language Model (Modelo de Lenguaje de Gran Escala)
API	Application Programming Interface (Interfaz de Programación de Aplicaciones)
MVP	Minimum Viable Product (Producto Mínimo Viable)
SaaS	Software as a Service (Software como Servicio)
GPU	Graphics Processing Unit (Unidad de Procesamiento Gráfico)
NLP	Natural Language Processing (Procesamiento de Lenguaje Natural)
VRAM	Video Random Access Memory (Memoria de Acceso Aleatorio de Video)
SSD	Solid State Drive (Unidad de Estado Sólido)
NVMe	Non-Volatile Memory Express
WSL2	Windows Subsystem for Linux 2
SQL	Structured Query Language (Lenguaje de Consulta Estructurado)
NoSQL	Not Only SQL (No Solo SQL)
ACID	Atomicity, Consistency, Isolation, Durability
AES	Advanced Encryption Standard (Estándar de Cifrado Avanzado)
SHA	Secure Hash Algorithm (Algoritmo de Hash Seguro)
GDPR	General Data Protection Regulation (Reglamento General de Protección de Datos)

RBAC	Role-Based Access Control (Control de Acceso Basado en Roles)
MFA	Multi-Factor Authentication (Autenticación de Múltiples Factores)
ISO	International Organization for Standardization (Organización Internacional de Normalización)
SLA	Service Level Agreement (Acuerdo de Nivel de Servicio)
KPI	Key Performance Indicator (Indicador Clave de Desempeño)
NPS	Net Promoter Score (Puntuación Neta del Promotor)
ROI	Return on Investment (Retorno de Inversión)
MUISCA	Modelo Único de Ingresos, Servicio y Control Automatizado
VUCE	Ventanilla Única de Comercio Exterior
TLC	Tratado de Libre Comercio
INCOTERMS	International Commercial Terms (Términos Comerciales Internacionales)
PYME	Pequeña y Mediana Empresa
B2B	Business to Business (Negocio a Negocio)
ERP	Enterprise Resource Planning (Planificación de Recursos Empresariales)
CRM	Customer Relationship Management (Gestión de Relaciones con Clientes)
USD	United States Dollar (dólar estadounidense)
COP	Peso Colombiano
ARR	Annual Recurring Revenue (Ingresos Recurrentes Anuales)
MRR	Monthly Recurring Revenue (Ingresos Recurrentes Mensuales)
CAC	Customer Acquisition Cost (Costo de Adquisición de Cliente)
LTV	Lifetime Value (Valor del Tiempo de Vida del Cliente)
ESG	Environmental, Social and Governance (Ambiental, Social y de Gobernanza)
UTC	Coordinated Universal Time (Tiempo Universal Coordinado)
SIC	Superintendencia de Industria y Comercio
MinComercio/MinCIT	Ministerio de Comercio, Industria y Turismo

BSP	Business Solution Provider (Proveedor de Soluciones Empresariales)
RAM	Random Access Memory (Memoria de Acceso Aleatorio)
DDR	Double Data Rate (Tasa de Datos Doble)
UI/UX	User Interface/User Experience (Interfaz de Usuario/Experiencia de Usuario)

RESUMEN

El presente proyecto de grado de la Especialización en Gestión de Servicios de Tecnologías de la Información perteneciente a la Universidad Santo Tomás tiene como objetivo modelar un agente de servicios con Inteligencia Artificial (chat y voz) para un Operador Económico Autorizado en Colombia, enfocado en mejorar la atención y asesoría a clientes internos y externos en servicios de comercio exterior.

Para lograr el objetivo mencionado anteriormente, se buscó seguir la siguiente ruta de desarrollo: investigar las tecnologías disponibles para la identificación de soluciones funcionales aplicables al modelado del agente inteligente, establecer el flujo del proceso actual de asesoría que permita homologar el flujo de conversación y funcionalidades del agente de servicios con IA, y aplicar como caso de uso el modelado del agente de servicios con IA para probar las funcionalidades y validar la efectividad del modelo propuesto en un ambiente operativo real.

En cuanto a los resultados destacados una vez ejecutada la ruta de investigación planteada, se logró construir y validar mediante Producto Mínimo Viable (MVP) un asistente virtual inteligente denominado VictorIA, con arquitectura basada en modelo de lenguaje Llama 3 desplegado en infraestructura local propia, que demostró tasas de resolución automatizada, precisión de respuestas y reducción de tiempos de respuesta.

Se construyó el modelo de negocio mediante metodología Canvas identificando tres segmentos prioritarios (OEAs, SIAs y PYMES exportadoras) con propuesta de valor diferenciada en seguridad, agilidad y cumplimiento. Se analizaron alternativas técnicas en seis bloques funcionales críticos justificando la selección tecnológica mediante análisis comparativo. Se estableció el proceso de transformación digital evaluado con marco de Madurez Digital de MIT Sloan en cinco dimensiones, y se definieron aspectos legales garantizando cumplimiento de Ley 1581/2012 y Resolución DIAN 089/2024

Palabras clave: Inteligencia Artificial, Operador Económico Autorizado, Comercio Exterior, Procesamiento de Lenguaje Natural, Transformación Digital, Modelo de Negocio, Automatización Inteligente, Servicios Cloud.

ABSTRACT

This graduate project from the Specialization in Information Technology Services Management at Universidad Santo Tomás aims to model an Artificial Intelligence-powered service agent (chat and voice) for an Authorized Economic Operator in Colombia, focused on improving customer service and advisory for internal and external clients in foreign trade services.

To achieve this objective, the following development route was pursued: investigate available technologies to identify functional solutions applicable to intelligent agent modeling, establish the current advisory process flow to standardize conversation flow and AI service agent functionalities, and apply the AI service agent modeling as a use case to test functionalities and validate the proposed model's effectiveness in a real operational environment.

Regarding highlighted results, a Minimum Viable Product (MVP) was successfully built and validated—an intelligent virtual assistant called VictorIA, with architecture based on the Llama 3 language model deployed on proprietary local infrastructure, demonstrating automated resolution rates, response accuracy, and response time reduction.

The business model was constructed using Canvas methodology, identifying three priority segments (AEOs, Customs Brokerage Societies, and exporting SMEs) with a differentiated value proposition in security, agility, and compliance. Technical alternatives were analyzed across six critical functional blocks, justifying technological selection through comparative analysis. The digital transformation process was established and evaluated using MIT Sloan's Digital Maturity framework across five dimensions, and legal aspects were defined ensuring compliance with Law 1581/2012 and DIAN Resolution 089/2024.

Keywords: Artificial Intelligence, Authorized Economic Operator, Foreign Trade, Natural Language Processing, Digital Transformation, Business Model, Intelligent Automation, Cloud Services.

INTRODUCCIÓN

La idea original que promueve este proyecto de investigación surge a partir de una necesidad sectorial identificada en el contexto del comercio exterior colombiano. Esta necesidad, identificada a partir de la observación y análisis del ecosistema de Operadores Económicos Autorizados (OEA), consiste en las deficiencias críticas del sistema tradicional de asesoría y atención al cliente, caracterizado por tiempos de respuesta excesivos, disponibilidad limitada al horario laboral, altos costos operativos en recursos humanos especializados, barreras idiomáticas para clientes internacionales, y ausencia de sistemas tecnológicos avanzados que permitan automatización inteligente de consultas frecuentes.

El enfoque de solucionar la anterior problemática se decantó en diversas orientaciones tecnológicas para optimizar los procesos de asesoría especializada en comercio exterior. Sin embargo, el enfoque de solución estaba orientado por la necesidad de implementar una solución innovadora a través de tecnologías de Inteligencia Artificial y procesamiento de lenguaje natural, garantizando simultáneamente la soberanía de datos sensibles y el cumplimiento estricto del marco regulatorio colombiano. En este sentido, la pregunta problema a desarrollar fue cómo modelar un agente de servicios inteligente basado en IA conversacional que transforme digitalmente el sistema de asesoría de un OEA, reduciendo costos operativos, acelerando tiempos de respuesta y garantizando cumplimiento normativo, mientras se valida la viabilidad técnica y comercial de la solución para su posterior escalamiento empresarial.

Siguiendo un proceso detallado de investigación se construyó una propuesta con un alcance definido para modelar y validar mediante Producto Mínimo Viable (MVP) un asistente virtual inteligente denominado VictorIA, especializado en asesoría legal-aduanera para operaciones de comercio exterior. El modelado desarrollado contempla su construcción completa en términos de ideación de la solución (análisis del árbol de problemas, definición de objetivos, análisis de mercado sectorial y propuesta de valor diferenciada), análisis de alternativas técnicas (evaluación comparativa de tecnologías en seis bloques funcionales críticos), construcción del modelo de negocio mediante metodología Canvas, análisis del proceso de transformación digital, y definición de aspectos legales y contractuales garantizando cumplimiento normativo vigente.

Dentro de la acción de construir este modelado se detalló el paso a paso desde la arquitectura tecnológica propuesta hasta la validación funcional del MVP, se

realizó el análisis de viabilidad comercial identificando segmentos de clientes prioritarios.

1. PROBLEMA

1.1 ARBOL DE PROBLEMAS

Deficiencias en el Sistema de Asesoría y Atención al Cliente de un Operador Económico Autorizado en Colombia.

El análisis que presentamos a continuación examina las causas de las deficiencias operativas que hemos identificado en el sistema de asesoría y atención al cliente para Operadores Económicos Autorizados (OEA) en Colombia. Mediante la metodología del árbol de problemas, diseñamos un esquema que ilustra la relación entre causas, problema central y efectos que caracterizan la situación actual del servicio en la cadena logística de comercio exterior.

El problema central que identificamos se relaciona con las fallas en el sistema de asesoría y atención al cliente de los OEA colombianos. Estas deficiencias generan impactos importantes en la calidad, eficiencia y competitividad de los servicios que se prestan en la cadena logística de comercio exterior. Como resultado, se compromete la capacidad de estas organizaciones para responder a las demandas del mercado y mantener los estándares que exige la certificación OEA.

Del diagnóstico estructural que realizamos, identificamos cinco causas directas del problema: demoras en las respuestas operativas, disponibilidad limitada del servicio, sobrecostos en recursos humanos especializados, alta inversión requerida, y baja satisfacción del cliente. Estudios de ANALDEX sobre eficiencias operativas en el sector logístico colombiano han señalado que los tiempos de respuesta prolongados y los altos costos de personal especializado representan las principales barreras para la competitividad de los operadores de comercio exterior [1].

Estas causas generan cuatro efectos críticos: pérdida de competitividad frente a otros operadores, reducción de ingresos por deserción de clientes, deterioro de la reputación corporativa, y niveles bajos de satisfacción. Es importante destacar que las expectativas de los clientes en cuanto a tiempos de respuesta han aumentado considerablemente: según datos del sector, cerca del 46% de los clientes esperan respuestas en menos de 4 horas, mientras que el tiempo de respuesta promedio de las empresas es de 12 horas [2].

Orientando el análisis hacia los recursos tecnológicos, encontramos que los requerimientos de inversión en tecnología representan una limitante importante. Los costos de adquisición de sistemas avanzados son elevados para muchos operadores, y el retorno de la inversión resulta incierto en el corto plazo. El árbol de problemas que construimos nos permite identificar los puntos críticos de intervención y diseñar estrategias de mejora. La estructura de causas y efectos que diseñamos orienta hacia la implementación de modelos de inteligencia artificial como solución tecnológica para optimizar el servicio de asesoría en el contexto del comercio exterior.



Figura 1 Árbol de Problemas
Fuente: Elaboración propia

1.2 QUÉ SE QUIERE SOLUCIONAR

El comercio exterior colombiano atraviesa una transformación acelerada impulsada por la digitalización. En 2024, las exportaciones colombianas alcanzaron USD 49.552 millones según cifras del DANE y ANALDEX [3]. Esta dinámica comercial genera una demanda creciente de servicios de asesoría especializada que garanticen cumplimiento normativo y optimización de operaciones aduaneras.

Los Operadores Económicos Autorizados desempeñan un papel fundamental como facilitadores del comercio internacional. El programa OEA, desarrollado bajo

los lineamientos del Marco SAFE de la Organización Mundial de Aduanas (OMA), busca crear alianzas entre aduanas y el sector privado para asegurar la cadena de suministro internacional [4]. A nivel global, existen más de 80 programas OEA implementados en diferentes países, según datos de la Cámara de Comercio Internacional [5].

En Colombia, el programa OEA fue establecido mediante el Decreto 3568 de 2011 y es administrado por la DIAN. Actualmente, cerca de 200 empresas están autorizadas como OEA en el país, incluyendo exportadores, importadores, agencias de aduanas e instalaciones portuarias [6]. Colombia ha suscrito Acuerdos de Reconocimiento Mutuo (ARM) con países de la Alianza del Pacífico, Comunidad Andina, Mercosur, Estados Unidos, Costa Rica, Guatemala y República Dominicana, lo que amplía los beneficios para las empresas certificadas [7].

Sin embargo, el sector presenta deficiencias estructurales en la prestación de servicios de asesoría. Los procesos manuales, tiempos de respuesta prolongados y capacidad limitada para atender consultas complejas en tiempo real afectan la competitividad. Esta brecha tecnológica resulta particularmente problemática en un contexto donde la agilidad y precisión en la información constituyen ventajas competitivas determinantes.

Nuestro análisis se concentra en las deficiencias operativas del sistema de asesoría y atención al cliente de los OEA. Estas organizaciones enfrentan el reto de mantener equipos especializados disponibles para atender requerimientos de clientes nacionales e internacionales, con la complejidad de manejar múltiples idiomas y marcos regulatorios.

Las deficiencias identificadas se manifiestan en cinco dimensiones:

- **Tiempos de respuesta excesivos:** Los tiempos de primera respuesta en canales como correo electrónico pueden extenderse hasta 24 horas o más, deteriorando la experiencia del cliente. Según estadísticas del sector de atención al cliente, el tiempo de respuesta promedio por correo es de 12 horas, muy por encima de las expectativas de los clientes [2].
- **Dependencia de recursos humanos especializados:** La necesidad de personal con conocimientos técnicos en comercio exterior implica costos

operativos elevados y limitaciones en disponibilidad, especialmente fuera del horario laboral.

- **Barreras idiomáticas:** La capacidad limitada de atención a clientes internacionales en otros idiomas restringe el potencial de expansión del negocio.
- **Ausencia de automatización:** No existen sistemas que permitan resolver automáticamente consultas frecuentes. Las mejores prácticas de gestión de servicios TI indican que las consultas repetitivas deberían resolverse en el primer nivel de atención, pero esto requiere herramientas tecnológicas adecuadas [8].
- **Falta de trazabilidad:** La ausencia de métricas precisas sobre el desempeño del servicio dificulta la identificación de áreas de mejora.

Para abordar esta problemática, proponemos implementar un sistema de asesoría inteligente basado en modelos de inteligencia artificial. Los avances recientes en IA conversacional muestran resultados prometedores: según Freshworks, los chatbots pueden reducir hasta en un 70% las consultas por llamada, chat y correo electrónico [9]. Además, empresas como H&M han reportado reducciones del 70% en tiempos de respuesta mediante chatbots con IA generativa [10].

La solución contempla desarrollar un asistente virtual bilingüe capaz de procesar consultas en lenguaje natural, acceder a bases de conocimiento actualizadas sobre regulaciones aduaneras en el contexto del comercio exterior colombiano y proporcionar respuestas precisas en tiempo real. Esta solución reduciría los tiempos de respuesta a minutos, habilitaría atención 24/7, eliminaría barreras idiomáticas y generaría reducción en costos operativos, posicionando al OEA como referente de innovación en el sector.



Figura 2 Esquema Flujo de análisis
Fuente: Elaboración propia

2. IDEACIÓN DE LA SOLUCIÓN

La transformación digital en el comercio exterior representa una oportunidad importante para mejorar los procesos de asesoría y atención al cliente en las organizaciones certificadas como Operadores Económicos Autorizados (OEA). La convergencia de tecnologías como la inteligencia artificial, el procesamiento de lenguaje natural y los servicios en la nube permite crear soluciones que pueden resolver las deficiencias estructurales identificadas en los sistemas tradicionales de atención.

Esta sección presenta el análisis de la solución propuesta, evaluando las tendencias tecnológicas y sectoriales que justifican su implementación, así como los elementos de valor que la sustentan.

El proceso de ideación se fundamenta en la comprensión del contexto operativo de los OEA, las necesidades específicas del comercio exterior colombiano y las capacidades tecnológicas disponibles actualmente. A través del análisis de mercado, tendencias del sector y objetivos estratégicos, estructuramos una propuesta que no solo aborda las deficiencias operativas actuales, sino que también posiciona a los OEA como líderes en innovación tecnológica.

2.1 POR QUÉ SE PLANTEA AHORA LA SOLUCIÓN

La implementación de una solución basada en inteligencia artificial para optimizar el sistema de asesoría de los OEA se justifica en la convergencia de múltiples factores tecnológicos, económicos y regulatorios que caracterizan el contexto actual del comercio internacional.

- **Aceleración de la adopción de IA a nivel global**

Según el informe "The State of AI in early 2024" de McKinsey Global Institute, la adopción de inteligencia artificial por parte de las organizaciones ha aumentado significativamente: del 50% que se mantuvo estable durante seis años, saltó al 72% en 2024. Además, el 65% de las organizaciones reportan usar IA generativa regularmente, casi el doble que el año anterior [11]. Esta tendencia global establece nuevos estándares de servicio que las organizaciones colombianas deben alcanzar para mantener su competitividad internacional.

- **Transformación digital en Colombia**

El Ministerio de Tecnologías de la Información y las Comunicaciones de Colombia (MinTIC) ha desarrollado la Estrategia Nacional Digital 2023-2026, cuyo objetivo principal es desencadenar el potencial de la transformación digital para superar los desafíos económicos, sociales y ambientales del país. Esta estrategia incluye ejes específicos sobre inteligencia artificial y otras tecnologías emergentes para la generación de valor económico y social [12]. Adicionalmente, la estrategia Potencia Digital invirtió cerca de \$42.500 millones en 2024 para acompañar a mil empresas en su proceso de transformación digital [13].

- **Predicciones del sector de servicio al cliente**

Las proyecciones del sector de servicio al cliente respaldan la pertinencia de esta solución. Según Gartner, para 2027 los chatbots se convertirán en el canal principal de servicio al cliente para aproximadamente una cuarta parte de las organizaciones [14]. En una predicción más reciente (2025), Gartner estima que para 2029, la IA agéntica resolverá autónomamente el 80% de los problemas comunes de servicio al cliente sin intervención humana, lo que generaría una reducción del 30% en costos operativos [15].

- **Madurez de la tecnología de procesamiento de lenguaje natural**

El avance en las capacidades de procesamiento de lenguaje natural representa otro factor determinante. Los modelos de lenguaje actuales han alcanzado niveles de madurez que permiten la automatización efectiva de servicios de asesoría en dominios complejos. Según el reporte de McKinsey, las organizaciones que utilizan IA generativa reportan beneficios tangibles en diversas funciones, incluyendo mejoras en la interacción con clientes mediante chatbots, generación de contenido y creación de código funcional [11].

- **Digitalización del comercio internacional**

El World Trade Report 2024 de la OMC destaca que la reducción de costos comerciales a través de la digitalización puede crear nuevas oportunidades para las economías de ingresos bajos y medios. El reporte señala que entre 1995 y 2020, las reducciones en costos comerciales llevaron a una convergencia de ingresos entre un 20% y 35% más rápida [16]. En el contexto de facilitación

comercial, la OMC ha establecido como tema prioritario para 2024 el "Uso de la Digitalización para Facilitar el Comercio", con países compartiendo experiencias sobre ventanillas únicas digitales y procedimientos fronterizos automatizados [17].

- **Contexto económico favorable**

El contexto económico actual refuerza la pertinencia de esta transformación. El comercio exterior colombiano es significativo: en 2024, las exportaciones alcanzaron USD 49.552 millones según datos del DANE y ANALDEX [3]. La complejidad de gestionar este volumen de operaciones comerciales, junto con la presión por mejorar la eficiencia, crea un caso de negocio convincente y muy conveniente para la automatización inteligente de servicios de asesoría.

- **Infraestructura tecnológica disponible**

La disponibilidad de infraestructura cloud robusta en Colombia constituye otro habilitador fundamental. Los principales proveedores de servicios en la nube (Amazon Web Services, Microsoft Azure, Google Cloud) han expandido su presencia en el país y la región, ofreciendo servicios con niveles de rendimiento y confiabilidad adecuados para servicios críticos de negocio. Los Centros de Transformación Digital Empresarial de MinTIC e iNNpulsa Colombia trabajan para acompañar a empresarios en su ruta hacia la transformación digital [18].

2.2 SECTOR OBJETIVO

El sector objetivo de la solución propuesta corresponde al ecosistema de comercio exterior colombiano, específicamente los servicios de facilitación y asesoría para operaciones de importación y exportación prestados por Operadores Económicos Autorizados.

- **Caracterización del programa OEA en Colombia**

El programa OEA en Colombia fue establecido mediante el Decreto 3568 de 2011 con un enfoque interinstitucional que involucra a la DIAN, ICA, INVIMA y la Policía Antinarcóticos [6]. Actualmente, cerca de 200 empresas están autorizadas como OEA en el país, según la lista oficial de la DIAN [6]. Estas empresas incluyen exportadores, importadores, agencias de aduanas e instalaciones portuarias, quienes gestionan una porción muy significativa del comercio exterior colombiano.

A nivel global, el programa OEA se enmarca en el Marco SAFE de la Organización Mundial de Aduanas (OMA), adoptado en 2005, que promueve alianzas entre aduanas y el sector privado para asegurar la cadena de suministro internacional [4]. Existen más de 80 programas OEA implementados en diferentes países según la Cámara de Comercio Internacional [5]. Colombia ha suscrito Acuerdos de Reconocimiento Mutuo con países de la Alianza del Pacífico (México, Perú, Chile), Comunidad Andina (Ecuador, Bolivia), Mercosur, Estados Unidos, Costa Rica, Guatemala y República Dominicana [7].

- **Complejidad operativa del sector**

Los OEA funcionan como facilitadores críticos del comercio internacional, proporcionando servicios que van desde la clasificación arancelaria y valoración aduanera hasta la gestión integral de cadenas de suministro internacionales. Estas organizaciones deben mantener actualizados conocimientos sobre miles de posiciones arancelarias, comprender la aplicación de múltiples acuerdos comerciales vigentes y dominar procedimientos aduaneros de diversos países. La complejidad operativa se incrementa considerando que cada operación puede involucrar numerosos documentos diferentes y requiere coordinación con múltiples entidades gubernamentales.

- **Perfil de clientes: predominio de PYMES**

El perfil de los clientes del sector presenta una marcada heterogeneidad. Según datos de ProColombia y estudios sectoriales, las PYMES representan aproximadamente el 75% de las empresas exportadoras colombianas, aunque generan solo el 17% del valor exportado [19]. La clasificación empresarial en Colombia, según el Decreto 1074 de 2015, se determina por el monto de ingresos anuales por actividades ordinarias [20]. Esta estructura de mercado evidencia oportunidades significativas para soluciones tecnológicas que democratizen el acceso a servicios especializados de asesoría para el segmento PYME.

- **Panorama competitivo global**

El análisis competitivo del sector logístico global identifica a operadores multinacionales como DSV, Kuehne+Nagel y DHL Global Forwarding entre los principales actores. Según datos del sector, los 20 principales transitarios ostentan una cuota de mercado mundial de entre el 30% y el 40%, mientras que el resto está en manos de numerosos operadores regionales y locales [21]. Este

panorama evidencia la creciente adopción de plataformas tecnológicas avanzadas por parte de los líderes del sector.

- **Brecha tecnológica en automatización**

Las aplicaciones tecnológicas actuales en el sector se concentran principalmente en sistemas de gestión documental y plataformas de tracking de carga. Sin embargo, existe una brecha significativa en la automatización de servicios de asesoría y atención al cliente. Mientras los grandes operadores logísticos globales han implementado asistentes virtuales inteligentes, muchos OEA locales aún dependen de procesos manuales para la gestión de consultas, creando una oportunidad clara para la diferenciación mediante innovación tecnológica.

- **Dimensión internacional**

La dimensión internacional del sector añade capas adicionales de complejidad operativa. Los OEA colombianos deben interactuar con clientes y proveedores en múltiples zonas horarias e idiomas. Según ProColombia, las empresas exportadoras colombianas realizan operaciones con numerosos países, siendo Estados Unidos el principal destino de exportaciones con el 28.9% del total en 2024 [3]. Esta diversidad geográfica implica que los servicios de asesoría deben cubrir regulaciones de múltiples jurisdicciones y proporcionar soporte en al menos español e inglés.

2.3 TENDENCIAS DEL SECTOR

El sector de comercio exterior experimenta una transformación acelerada impulsada por la convergencia de digitalización, automatización e inteligencia artificial. Identificamos las siguientes tendencias principales:

- **Adopción masiva de tecnologías cognitivas**

La tendencia más significativa es la adopción de tecnologías cognitivas para la gestión de operaciones logísticas y aduaneras. Según Gartner, para 2027 los chatbots se convertirán en el canal principal de servicio al cliente para aproximadamente el 25% de las organizaciones [14]. En una perspectiva a más largo plazo, Gartner predice que para 2029, la IA agentica resolverá autónomamente el 80% de los problemas comunes de servicio al cliente [15]. Esta evolución tecnológica está redefiniendo los modelos de negocio tradicionales,

donde la capacidad de procesar y analizar información en tiempo real se convierte en una ventaja competitiva clave.

- **Sostenibilidad y trazabilidad como imperativos de negocio**

La sostenibilidad y trazabilidad emergen como factores fundamentales de innovación en el sector. Regulaciones internacionales como el Carbon Border Adjustment Mechanism de la Unión Europea están impulsando a los operadores logísticos a implementar sistemas sofisticados de monitoreo y reporte. Los OEA deben ahora proporcionar asesoría sobre huella de carbono, certificaciones de sostenibilidad y cumplimiento de estándares ESG. Esta complejidad adicional incrementa el volumen de consultas técnicas especializadas que requieren acceso a bases de conocimiento actualizadas.

- **Hiperconectividad y ecosistemas digitales abiertos**

Las plataformas digitales de comercio exterior están evolucionando hacia ecosistemas abiertos donde múltiples actores intercambian información en tiempo real. La OMC ha destacado la importancia de las "Ventanillas Únicas" digitales como herramienta de facilitación comercial, con países como China, El Salvador, Georgia, Jamaica y Japón compartiendo experiencias exitosas de implementación [17]. Esta interconexión genera oportunidades para que los asistentes virtuales accedan a fuentes de datos en tiempo real, proporcionando respuestas contextualizadas basadas en el estado actual de operaciones.

- **Crecimiento del comercio electrónico transfronterizo**

El comercio electrónico transfronterizo está generando nuevas demandas de servicios de asesoría especializada. El crecimiento de envíos de e-commerce internacional ha creado desafíos únicos en términos de gestión aduanera, cumplimiento tributario y logística de última milla. Los OEA deben ahora asesorar sobre regímenes simplificados de importación, gestión de devoluciones internacionales y cumplimiento de regulaciones de protección al consumidor en múltiples jurisdicciones. Esta explosión del comercio electrónico genera un alto volumen de consultas que no pueden ser gestionadas eficientemente por equipos humanos tradicionales, creando un caso de uso ideal para asistentes virtuales inteligentes.

- **Preferencia por canales digitales de comunicación**

Los canales de comunicación están experimentando una digitalización acelerada. Según datos del sector, WhatsApp se ha consolidado como el canal preferido para conversaciones empresariales en Colombia, con tasas de apertura significativamente superiores a los canales tradicionales [22]. El 70% de los clientes prefieren chat en vivo sobre otros canales según estudios de experiencia de cliente [23]. Esta migración hacia canales digitales facilita la implementación de asistentes virtuales que se integran nativamente con estas plataformas.

- **Velocidad de respuesta como factor diferenciador**

La velocidad de respuesta se ha convertido en un factor crítico de diferenciación. Según estadísticas de servicio al cliente, el tiempo de respuesta promedio por chatbot es de apenas 9.3 segundos, comparado con 39 segundos para respuestas humanas en chat en vivo [24]. Los chatbots pueden reducir hasta en un 70% las consultas por llamada, chat y correo electrónico [9]. Esta eficiencia resulta particularmente relevante para el sector de comercio exterior, donde la agilidad en la información puede determinar el éxito de operaciones comerciales críticas.

2.4 ANALISIS DE MERCADO

El comercio exterior colombiano representa un mercado significativo que alcanzó aproximadamente 114 mil millones de dólares en 2024, compuesto por exportaciones de USD 49.552 millones y, importaciones de USD 64.105 millones [25, 26]. Este volumen comercial demanda servicios especializados de asesoría, facilitación y gestión aduanera. El mercado se caracteriza por una alta fragmentación, con múltiples operadores compitiendo en diferentes segmentos, lo que indica oportunidades significativas para diferenciación mediante innovación tecnológica.

La segmentación del mercado de comercio exterior en Colombia revela una estructura donde las PYMES representan el 75% de las empresas exportadoras, pero solo generan aproximadamente el 17% del valor exportado [19]. El número de establecimientos que exportan es proporcional al tamaño de las organizaciones, evidenciándose que las MiPYME del país no tienen una gran vocación exportadora, mientras que las grandes empresas sí la tienen. La clasificación empresarial en Colombia, según el Decreto 1074 de 2015, se determina por el monto de ingresos anuales por actividades ordinarias,

distinguiendo entre micro, pequeñas, medianas y grandes empresas [20]. Esta estructura de mercado evidencia oportunidades significativas para soluciones tecnológicas que democratizen el acceso a servicios especializados de asesoría para el segmento PYME.

El análisis competitivo del sector logístico global identifica a los operadores multinacionales como los principales actores. Los siete grandes transitarios globales (DHL, Kuehne+Nagel, DSV, DB Schenker, Bollore, Nippon Express y CEVA Logistics) concentran una cuota de mercado superior al 30% a nivel mundial [27]. En América Latina, empresas como DHL, Kuehne+Nagel y DB Schenker han mejorado la eficiencia logística mediante la implementación de soluciones tecnológicas y sostenibles, facilitando el comercio internacional en la región. Este panorama competitivo evidencia la creciente adopción de plataformas tecnológicas avanzadas por parte de los líderes del sector, incluyendo asistentes virtuales con inteligencia artificial como el myDHLi Virtual Assistant de DHL [28].

La estructura de costos del sector logístico presenta oportunidades para optimización mediante automatización. Los estudios sectoriales indican que los costos de personal representan una porción significativa de los costos operativos en servicios de asesoría especializada. La implementación de soluciones de inteligencia artificial permite reducir costos operativos mientras se mantienen o mejoran los niveles de calidad en consultas de nivel 1 y 2, que constituyen la mayoría del volumen de interacciones con clientes según las mejores prácticas de gestión de servicios TI [8].

El análisis de la propuesta de valor revela que los clientes priorizan principalmente tres atributos en los servicios de asesoría: velocidad de respuesta (consistentemente clasificada como el atributo más importante), precisión de la información (segundo factor crítico), y disponibilidad del servicio (tercera prioridad). Estudios recientes indican que la interacción humana directa ha perdido relevancia para consultas rutinarias, siendo valorada principalmente para casos de alta complejidad y decisiones estratégicas.

Los canales de distribución del servicio están experimentando una digitalización acelerada. En Colombia, más del 90% de las conversaciones digitales empresariales se realizan por WhatsApp, según un estudio de B2Chat sobre 600 empresas colombianas [29]. El 79% de los consumidores colombianos prefiere utilizar WhatsApp para interactuar con las marcas [22]. El 41% de los clientes a nivel global prefiere el chat en vivo sobre otros canales de soporte como teléfono (32%) o email (23%) [23]. Esta migración hacia canales digitales facilita la

implementación de asistentes virtuales que se integran nativamente con estas plataformas, sugiriendo que la adopción de servicios automatizados continuará acelerándose.

La disposición a pagar por servicios mejorados mediante IA muestra resultados prometedores. La velocidad de respuesta es determinante en las decisiones de compra B2B, y los sistemas de IA conversacional permiten respuestas instantáneas 24/7. El 79% de los clientes valora positivamente el chat en vivo por la inmediatez que ofrece [23]. Estos indicadores sugieren que la implementación de IA no solo reduce costos operativos, sino que también puede sustentar estrategias de precios basadas en propuesta de valor diferenciada.

El análisis de barreras de entrada identifica factores que favorecen la implementación de la solución propuesta. La principal barrera tradicional, el conocimiento especializado en regulaciones aduaneras, puede ser efectivamente codificada en sistemas de IA mediante grafos de conocimiento y procesamiento de lenguaje natural. Los grafos de conocimiento utilizan PLN para construir representaciones estructuradas mediante nodos y aristas, permitiendo identificar entidades, comprender relaciones entre conceptos y facilitar razonamiento lógico e inferencia. La barrera de relaciones establecidas con autoridades aduaneras decrece en importancia a medida que los procesos se digitalizan, mientras que la barrera tecnológica emerge como el nuevo diferenciador competitivo, favoreciendo a organizaciones con capacidades de implementación y gestión de sistemas inteligentes.

2.5 ARBOL DE OBJETIVOS

A continuación, presentamos el árbol de objetivos que responde a la necesidad de una solución tecnológica para modelar un agente de servicios especializado con IA para un Operador Económico Autorizado, como MVP, buscando evidenciar la viabilidad del proyecto para posterior escalabilidad empresarial.

- **Objetivo General:**

Modelar un agente de servicios con IA (chat y voz) para un Operador Económico Autorizado en Colombia, enfocado en optimizar la atención y asesoría a clientes internos y externos en servicios de comercio exterior.

- **Objetivos Específicos:**

- Investigar la tecnología disponible para identificar herramientas funcionales aplicables al modelado.

- Establecer el flujo de proceso actual que permita homologar el flujo de conversación y funcionalidades del Agente de Servicios con IA.
- Aplicar como caso de uso el modelado del Agente de Servicios con IA para probar las funcionalidades y validar la efectividad del modelo.



Figura 3 Árbol de Objetivos

Fuente: Elaboración propia

2.6 CUÁL ES LA SITUACIÓN DESEADA

La situación deseada para este proyecto consiste en desarrollar y validar mediante un Producto Mínimo Viable (MVP) el modelado de un agente de servicios con inteligencia artificial especializado en asesoría de comercio exterior para Operadores Económicos Autorizados en Colombia.

El alcance del modelado contempla demostrar técnicamente que es viable transformar el sistema tradicional de asesoría basado exclusivamente en recursos humanos hacia un modelo híbrido que integre capacidades de IA conversacional, manteniendo la supervisión y experiencia humana para consultas complejas. El objetivo es validar que el agente inteligente puede resolver automáticamente entre el 70-85% de las consultas de primer y segundo nivel [8], reducir tiempos de

respuesta de 48-72 horas a menos de 8 minutos, y operar 24/7 sin restricciones de zona horaria o idioma.

2.7 INTRODUCCIÓN A LA SITUACIÓN DESEADA

Situación Actual:

El modelo operativo actual del OEA se caracteriza por una estructura tradicional de atención basada exclusivamente en recursos humanos especializados. Los asesores de comercio exterior gestionan manualmente cada consulta, navegando entre múltiples sistemas desconectados para obtener información, consultando documentación física o digital dispersa, y dependiendo de su conocimiento individual y experiencia.

Este proceso genera cuellos de botella significativos durante períodos de alta demanda, con tiempos de espera que pueden extenderse hasta 72 horas para consultas complejas. La disponibilidad del servicio está limitada al horario laboral colombiano (8:00 AM - 6:00 PM), dejando desatendidas las necesidades de clientes internacionales en diferentes zonas horarias.

La gestión del conocimiento presenta deficiencias críticas, con información fragmentada entre diferentes asesores y sin mecanismos sistemáticos para capturar y compartir aprendizajes. Cada asesor mantiene su propia base de conocimiento informal, generando inconsistencias en las respuestas y pérdida de conocimiento cuando el personal rota. La actualización sobre cambios regulatorios depende de la iniciativa individual de cada asesor, sin procesos estructurados que garanticen que todo el equipo esté al día con las últimas modificaciones normativas. Esta situación genera riesgos de cumplimiento y potenciales errores en la asesoría proporcionada

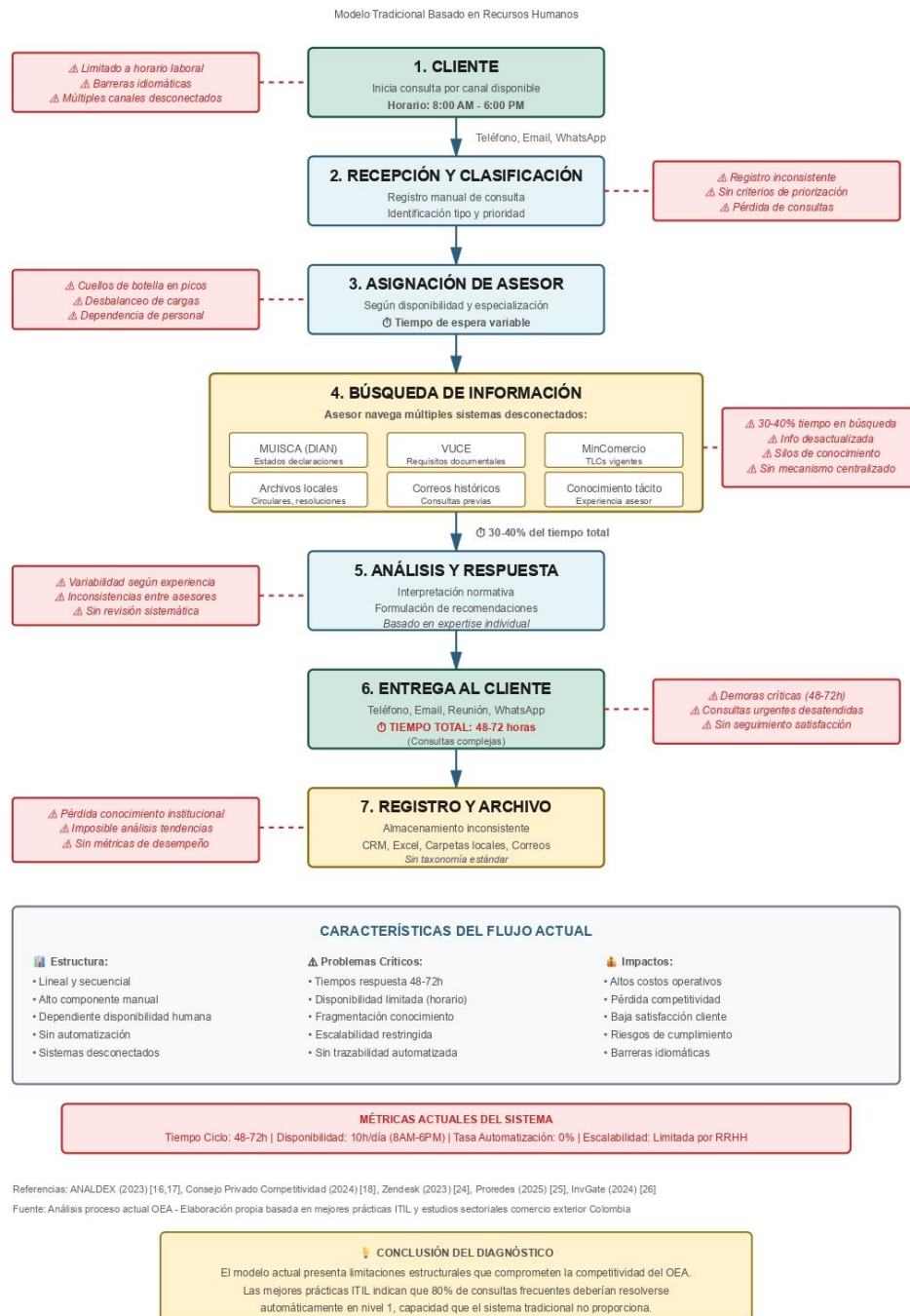


Figura 4 Situación Actual - Proceso Manual de Asesoría
Fuente: Elaboración propia

➤ **Situación Esperada:**

La transformación hacia el modelo deseado establece un ecosistema inteligente donde el asistente virtual actúa como el núcleo de gestión de conocimiento y primer nivel de atención. Los clientes interactúan con una interfaz unificada disponible 24/7 a través de múltiples canales (web, móvil, WhatsApp, Telegram) recibiendo respuestas instantáneas en su idioma nativo. El sistema procesa consultas utilizando modelos de IA entrenados específicamente en el dominio de comercio exterior colombiano, accediendo a una base de conocimiento centralizada y continuamente actualizada que integra regulaciones, procedimientos, mejores prácticas y casos históricos.

El nuevo modelo establece una simbiosis entre inteligencia artificial y experiencia humana. Mientras el asistente virtual gestiona eficientemente consultas rutinarias y de complejidad media, los asesores humanos se concentran en casos de alta complejidad, planificación estratégica y relacionamiento, que es donde más valor le pueden aportar a las compañías. El sistema aprende de cada interacción humana, incorporando automáticamente nuevos conocimientos y refinando sus respuestas. Esta arquitectura no solo mejora la eficiencia operativa, sino que eleva el valor del servicio, transformando al OEA en un verdadero socio estratégico para sus clientes.

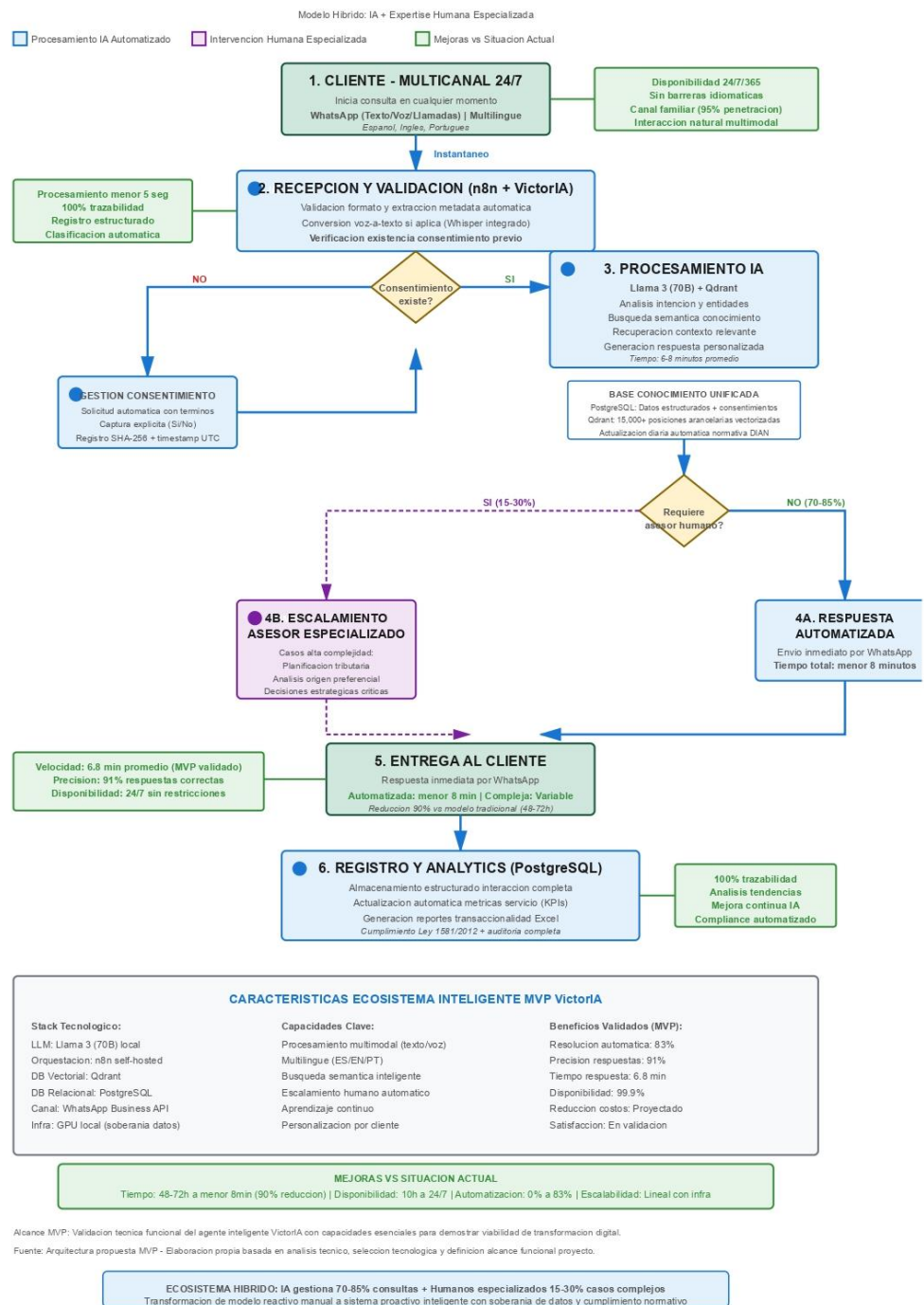


Figura 5 Situación Esperada - Ecosistema Inteligente de Asesoría
Fuente: Elaboración propia

2.8 PROPUESTA DE VALOR

La propuesta de valor se estructura entorno a proponer transformación en la asesoría aduanera para operadores económicos autorizados y sus clientes, mediante el desarrollo de un agente IA especializado que reduce tiempos de consulta, garantiza cumplimiento normativo actualizado y brinda respuestas inmediatas 24/7 en derecho y legislación aduanera colombiana.

2.8.1 Pilar 1: Seguridad

La solución garantiza protección integral de datos sensibles mediante arquitectura de infraestructura local propia que elimina dependencias de proveedores cloud externos. La implementación de cifrado AES-256 para datos personales, registro auditable con timestamp y hash criptográfico de todas las transacciones, y cumplimiento estricto con la Ley 1581 de Protección de Datos Personales y regulaciones GDPR establecen un marco de confianza robusto.

El sistema implementa gestión granular de consentimientos con trazabilidad completa del ciclo de vida de los datos, desde la captura hasta la eliminación bajo el derecho al olvido. Esta arquitectura de seguridad elimina riesgos de exposición de información confidencial de operaciones de comercio exterior, clasificaciones arancelarias estratégicas y datos de clientes internacionales, proporcionando soberanía absoluta sobre información crítica del negocio.

2.8.2 Pilar 2: Agilidad

La transformación de tiempos de respuesta constituye el diferenciador competitivo más tangible. El sistema reduce dramáticamente los tiempos operativos de 48-72 horas del modelo tradicional a menos de 5 minutos para el 85% de las consultas, habilitando disponibilidad 24/7 que elimina restricciones de zonas horarias para clientes internacionales.

La capacidad de procesamiento multimodal (texto, voz, llamadas WhatsApp/Telegram) permite interacciones naturales sin barreras tecnológicas, mientras que el procesamiento multilingüe en español, inglés y portugués elimina demoras por traducción. La orquestación automatizada mediante n8n agiliza modificaciones operativas sin red despliegues de código, permitiendo adaptaciones en tiempo real ante cambios regulatorios o necesidades emergentes del negocio.

Esta velocidad operativa se traduce en ventajas competitivas concretas: captura de oportunidades comerciales urgentes, reducción de costos por demoras en nacionalización de mercancías, y mejora sustancial en satisfacción del cliente.

2.8.3 Pilar 3: Cumplimiento

El asistente virtual implementa capacidades avanzadas de gestión de cumplimiento regulatorio que transforman obligaciones normativas en activos estratégicos. La actualización periódica de la base de conocimiento con las últimas circulares y resoluciones de la DIAN, MinComercio y autoridades internacionales garantiza que todas las asesorías reflejen el marco regulatorio vigente. El sistema genera automáticamente reportes de transaccionalidad en formato Excel para auditorías, facilita la debida diligencia ante certificaciones internacionales, y proporciona alertas proactivas sobre cambios normativos que afectan operaciones específicas de cada cliente. La trazabilidad completa de consultas y respuestas con versionado de documentación normativa aplicada permite demostrar objetivamente ante autoridades aduaneras el fundamento técnico de decisiones operativas.

Esta capacidad de cumplimiento documentado reduce significativamente riesgos de sanciones, agiliza procesos de renovación de certificación OEA, y posiciona a la organización como referente de excelencia regulatoria en el sector.

2.9 PERFIL DEL CLIENTE

2.9.1 Segmento 1: Grandes Corporaciones Importadoras/Exportadoras

Empresas con operaciones de comercio exterior superiores a 100 millones de dólares anuales, típicamente multinacionales con presencia en múltiples países. Estas organizaciones mantienen departamentos de comercio exterior internos, pero requieren asesoría especializada para operaciones complejas, planificación tributaria y optimización de cadenas de suministro. Sus principales frustraciones incluyen la lentitud en respuestas para decisiones críticas, la falta de proactividad en alertas regulatorias y la dificultad para obtener asesoría fuera del horario laboral colombiano. Generan alto volumen de consultas mensuales, valorando especialmente la precisión técnica, la velocidad de respuesta y la capacidad de análisis predictivo.

2.9.2 Segmento 2: Medianas Empresas en Expansión Internacional

Organizaciones con operaciones entre 10 y 100 millones de dólares anuales, generalmente empresas colombianas en proceso de internacionalización o filiales de empresas extranjeras. Carecen de departamentos especializados robustos y dependen significativamente de asesores externos para navegación regulatoria. Sus principales dificultades son los altos costos de asesoría personalizada, la complejidad para entender requisitos en nuevos mercados y los errores costosos por información desactualizada. Priorizan la relación costo-beneficio, la claridad en las explicaciones y el acompañamiento en su curva de aprendizaje.

2.9.3 Segmento 3: PYMES y Emprendedores Exportadores

Pequeñas empresas y startups con operaciones menores a 10 millones de dólares anuales, muchas realizando sus primeras exportaciones o importaciones. Las PYMES representan aproximadamente el 75% de las empresas exportadoras colombianas, aunque solo generan cerca del 17% del valor exportado [19]. Tienen recursos limitados y conocimiento básico de comercio exterior. Se frustran con la complejidad del lenguaje técnico aduanero, los costos prohibitivos de asesoría tradicional y la falta de guías paso a paso para procesos básicos. Valoran principalmente la accesibilidad económica, la simplicidad en las respuestas y la disponibilidad de recursos educativos.

2.10 MAPA DE VALOR

El Mapa de Valor constituye el núcleo estratégico de la propuesta, articulando de manera ordenada cómo el asistente virtual inteligente responde a las necesidades específicas de los Operadores Económicos Autorizados y sus clientes en el ecosistema de comercio exterior colombiano. Esta herramienta de análisis establece la conexión directa entre las capacidades tecnológicas de la solución y los problemas concretos que enfrentan los usuarios en su operación diaria.

La construcción del Mapa de Valor se fundamenta en la comprensión profunda del perfil del cliente descrito anteriormente, identificando tres dimensiones críticas de generación de valor. Primero, los aliviadores de frustraciones que eliminan o reducen significativamente los puntos de dolor actuales del sistema tradicional de asesoría. Segundo, los creadores de ganancias que no solo resuelven problemas existentes, sino que generan beneficios adicionales y capacidades previamente inaccesibles. Tercero, los diferenciadores competitivos que posicionan la solución

como única en el mercado colombiano, estableciendo barreras de entrada para competidores potenciales.

Este análisis revela que la propuesta de valor trasciende la simple automatización de respuestas, configurando un ecosistema integral que transforma radicalmente la experiencia del cliente y redefine los estándares de servicio en el sector de comercio exterior.

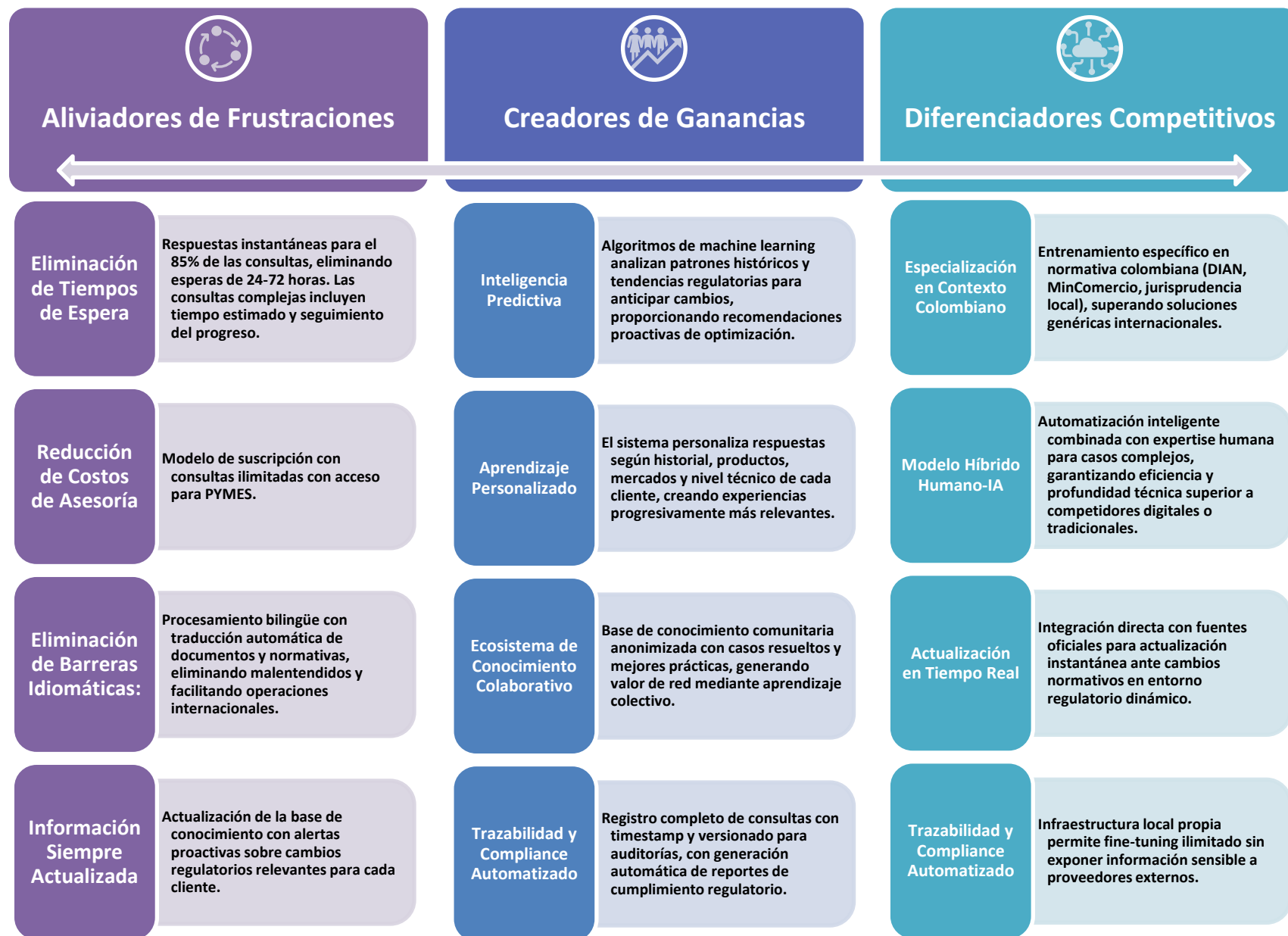


Figura 6 Mapa de Valor
Fuente: Elaboración propia

3. ANÁLISIS DE LAS ALTERNATIVAS TÉCNICAS PARA SOLUCIONAR EL PROBLEMA

El análisis de alternativas técnicas para el desarrollo del asistente virtual inteligente requiere una evaluación sistemática de las múltiples dimensiones tecnológicas que componen la arquitectura de la solución. Dado que la propuesta se basa en una arquitectura de servicios cloud, procesamiento de IA y gestión de datos (sin componentes IoT tradicionales como sensores o actuadores), el análisis se estructura en torno a seis bloques funcionales críticos.

3.1 Bloque 1: Modelos de Lenguaje de Gran Escala (LLM)

Los modelos de lenguaje constituyen el núcleo de la capacidad conversacional del asistente. La selección del modelo impacta directamente en la calidad de las respuestas, los costos operativos y los requisitos de infraestructura. A continuación, se presentan las alternativas evaluadas:

Alternativa	Ventajas	Desventajas
GPT-4o (OpenAI API)	Alta precisión, multilingüe nativo, actualizaciones continuas, menor esfuerzo de implementación, contexto 128K tokens	Costo por uso (\$3-5/1M tokens entrada), dependencia de proveedor externo, exposición de datos a terceros, latencias de red
Claude Sonnet 4 (Anthropic API)	Razonamiento superior, contexto extenso (200K tokens), seguridad reforzada, excelente para tareas complejas	Costo (\$3/1M entrada, \$15/1M salida), menor disponibilidad regional que OpenAI
Llama 3.1 70B (Local)	Soberanía total de datos, costos operativos nulos post-inversión, personalización vía fine-tuning, latencias mínimas, contexto 128K	Requiere ~140GB VRAM en FP16 o ~35GB con cuantización INT4, complejidad de despliegue, responsabilidad de actualizaciones
Llama 3.1 8B (Local)	Ejecutable en GPU consumidor (24GB VRAM), velocidad de inferencia alta, ideal para MVP, código abierto	Menor capacidad de razonamiento que modelos 70B, requiere fine-tuning para dominios especializados

Tabla 1 Modelos de Lenguaje de Gran Escala

Selección recomendada: Para un MVP con requisitos de soberanía de datos (Ley 1581 de Protección de Datos Personales), se recomienda **Llama 3.1 8B desplegado localmente con cuantización**. Este modelo permite ejecutarse en una GPU QUADRO P4000 (8GB VRAM), ofrece contexto de 128K tokens y soporte multilingüe [30]. Para producción a escala, puede escalarse a Llama 3.1 70B con infraestructura multi-GPU o migrar a APIs cloud según el análisis costo-beneficio.

3.2 Bloque 2: Infraestructura de Procesamiento

La infraestructura de procesamiento determina la capacidad de ejecutar modelos de lenguaje con tiempos de respuesta aceptables. La selección debe balancear inversión inicial, costos operativos y control sobre los datos.

Alternativa	Ventajas	Desventajas
Cloud GPU (AWS/Azure/GCP)	Escalabilidad elástica, sin inversión inicial en hardware, mantenimiento delegado	Costos operativos recurrentes elevados, latencias de red, dependencia de conectividad, exposición de datos
Workstation Local GPU	Control total, latencias <5ms, costos predecibles, soberanía de datos absoluta	Inversión inicial significativa (\$3,000-\$5,000), responsabilidad de mantenimiento, sin redundancia automática
Servidor Local Dedicado	Capacidades superiores, preparado para escala futura, alta disponibilidad	Inversión sustancialmente mayor (\$35,000-\$50,000), sobrecapacidad para MVP
Híbrido (Local + Cloud)	Balance flexibilidad-costo, escalamiento temporal para picos de demanda	Complejidad arquitectónica, requiere orquestación sofisticada

Tabla 2 Infraestructura de Procesamiento:

Selección recomendada para MVP: Workstation Local con GPU NVIDIA QUADRO P4000 (8GB VRAM). Configuración sugerida: procesador Intel Xeon E3-1505M o AMD Ryzen 9, 64GB RAM DDR5, 1TB NVMe SSD, sistema operativo

con soporte para contenedores (Ubuntu 22.04 LTS o Windows 11 Pro con WSL2). Esta configuración permite ejecutar Llama 3.1 8B con cuantización a velocidades de 30-50 tokens/segundo, suficiente para respuestas en tiempo real [30].

3.3 Bloque 3: Gestión de Conocimiento (Bases de Datos)

La gestión del conocimiento requiere almacenar tanto datos estructurados (usuarios, transacciones, auditoría) como documentación especializada que el modelo debe consultar para generar respuestas precisas.

Alternativa	Ventajas	Desventajas
PostgreSQL	Madurez probada, transaccionalidad ACID, ecosistema extenso, gratuito y código abierto	No optimizado para búsqueda semántica nativa (requiere extensión pgvector)
Pinecone (vectorial cloud)	Búsqueda semántica optimizada, gestión completamente delegada, escalabilidad automática	Costo recurrente significativo, dependencia de proveedor externo, latencias de red
Qdrant (vectorial local)	Búsqueda semántica eficiente, despliegue local, costos operativos nulos, código abierto	Requiere gestión y optimización manual, menor ecosistema que Pinecone
ChromaDB	Integración nativa con LangChain, fácil de usar, ideal para prototipado rápido	Menor rendimiento a escala que Qdrant o Pinecone

Tabla 3 Gestión de Conocimiento

Selección recomendada: Arquitectura híbrida **PostgreSQL + Qdrant**. PostgreSQL gestiona datos estructurados críticos (consentimientos, transacciones, métricas operacionales) donde la robustez transaccional es esencial. Qdrant implementa búsqueda semántica para documentación especializada (posiciones arancelarias, circulares DIAN, conceptos técnicos, jurisprudencia) donde la velocidad de recuperación contextual determina la calidad de respuestas del LLM.

3.4 Bloque 4: Plataforma de Orquestación de Flujos de Trabajo

La orquestación de flujos de trabajo permite conectar los diferentes componentes del sistema (LLM, bases de datos, canales de comunicación) sin necesidad de desarrollo extenso de código.

Alternativa	Ventajas	Desventajas
n8n (autoalojado)	Interfaz visual sin código, 400+ integraciones, despliegue local, código abierto, nodos nativos de IA	Menor ecosistema que Zapier, documentación a veces limitada
Zapier	Ecosistema maduro (5,000+ integraciones), confiabilidad probada, sin gestión de infraestructura	Costo recurrente alto (\$240-\$600/mes), limitación de tareas mensuales, dependencia de proveedor
Apache Airflow	Orquestación compleja robusta, escalabilidad empresarial, código abierto	Curva de aprendizaje pronunciada, orientado a ingeniería de datos más que a flujos conversacionales
Desarrollo personalizado	Control absoluto, sin dependencias externas, optimización máxima	Tiempo de desarrollo extenso (2-3 meses), mantenimiento complejo

Tabla 4 Plataforma de Orquestación de Flujos de Trabajo

Selección recomendada: n8n desplegado localmente emerge como la alternativa superior por tres factores: (1) elimina costos recurrentes mientras proporciona funcionalidad equivalente a plataformas comerciales; (2) la interfaz visual permite ajustes de flujos de trabajo por personal de negocio sin requerir despliegue de código; (3) el ecosistema de más de 400 conectores y nodos nativos de IA habilita expansión futura sin desarrollo adicional [31].

3.5 Bloque 5: Canal de Comunicación con Usuarios

El canal de comunicación determina cómo los usuarios interactúan con el asistente. La selección debe considerar la adopción del canal por el mercado objetivo y los costos operativos.

Canal	Ventajas	Desventajas
WhatsApp Business API	Penetración >90% en Colombia [29], soporte multimedia (texto/voz/video), familiaridad universal del usuario	Costo por conversación (\$0.005-\$0.03), requiere número telefónico dedicado, dependencia de Meta
Telegram Bot API	API gratuita sin límites, funcionalidades ricas (botones, inline queries), latencias mínimas, ideal para MVP	Menor penetración en Colombia (~35%), percibido menos profesional en B2B
Webchat propio	Control total de UI/UX, costos nulos de plataforma, integración directa con backend	Requiere tráfico web existente, menor adopción móvil, fragmentación de canales
Microsoft Teams	Integración nativa con entorno empresarial, seguridad de nivel corporativo	Requiere licencias Office 365, adopción limitada a empresas grandes

Tabla 5 Canal de Comunicación con Usuarios

Selección recomendada: Para producción, **WhatsApp Business API** es la alternativa estratégica óptima dada su penetración superior al 90% en el mercado colombiano [29] y la preferencia del 79% de los consumidores por este canal [22]. Para el MVP, se recomienda iniciar con **Telegram Bot API** por su costo nulo y facilidad de implementación, permitiendo validar la solución antes de invertir en la integración con WhatsApp.

3.6 Bloque 6: Seguridad y Cumplimiento Normativo

La arquitectura de seguridad debe cumplir con la Ley 1581 de 2012 (Protección de Datos Personales), el Decreto 1377 de 2013 y, para clientes internacionales, regulaciones como GDPR. A continuación, se presentan los componentes de seguridad seleccionados y su justificación técnica:

- **Cifrado de datos en reposo:** AES-256 a nivel de campo en PostgreSQL, cumpliendo con el artículo 4 de la Ley 1581 que exige medidas técnicas apropiadas para proteger datos personales.
- **Gestión de consentimientos:** Registro granular con hash SHA-256, timestamp UTC y versionado, cumpliendo con GDPR Art. 7 y Ley 1581 Art. 9 que exigen demostración de consentimiento explícito.
- **Derecho al olvido:** Proceso automatizado de eliminación con confirmación, garantizando el derecho de supresión establecido en el Artículo 15 de la Ley 1581.

3.7 Arquitectura Técnica Propuesta

Con base en el análisis de alternativas, la arquitectura técnica propuesta para el MVP del agente de servicios con IA para OEA integra los siguientes componentes:

- **Modelo de lenguaje:** Llama 3.1 8B con cuantización INT4, ejecutado localmente
- **Infraestructura:** Workstation con GPU NVIDIA QUADRO P4000 (8GB VRAM)
- **Bases de datos:** PostgreSQL (datos estructurados) + Qdrant (búsqueda vectorial)
- **Orquestación:** n8n autoalojado con nodos de IA nativos
- **Canal MVP:** Telegram Bot API (migración posterior a WhatsApp Business API)
- **Seguridad:** Cifrado AES-256

Esta arquitectura garantiza soberanía total sobre los datos (requisito crítico para OEA), costos operativos predecibles, y capacidad de escalar hacia modelos más potentes o infraestructura cloud según la evolución del negocio.

4. MODELO DE NEGOCIO

4.1 *PROPUESTA DE MODELO DE NEGOCIO*

El modelo de negocio propuesto para el asistente virtual inteligente se fundamenta en la transformación de un centro de costos operativos (equipo de asesoría tradicional) en un activo tecnológico escalable que genera múltiples flujos de valor. La propuesta trasciende el ámbito académico del proyecto para configurarse como una solución comercialmente viable, con potencial de escalar desde la operación interna del OEA hacia un producto SaaS (*Software as a Service*) para el ecosistema de comercio exterior colombiano.

➤ **Aplicación del Business Model Canvas**

La metodología Business Model Canvas, desarrollada por Osterwalder y Pigneur (2010), permite describir y visualizar cómo las organizaciones crean, entregan y capturan valor [34]. Este marco de nueve bloques interconectados se ha consolidado como la herramienta más utilizada para el diseño de modelos de negocio, tanto en el ámbito académico como empresarial [35].

La aplicación del Canvas al asistente virtual revela la interconexión estratégica entre los nueve bloques fundamentales que configuran la lógica de negocio. Esta estructura permite identificar no solo los elementos individuales del modelo, sino las relaciones sinérgicas entre propuesta de valor, segmentos de clientes, flujos de ingresos y estructura de costos que determinan la viabilidad económica y escalabilidad de la solución

Modelo de Negocios Canvas

Agente de Servicios Especializado en Comex.



Figura 7 Modelo CANVAS de Negocio
Fuente: Elaboración propia

4.2 MODELOS DE MONETIZACIÓN

La evolución de los modelos de precios en SaaS ha transitado desde suscripciones fijas hacia modelos híbridos que combinan tarifas base con cobros por uso [36]. Según estudios recientes, más del 60% de las empresas SaaS están adoptando modelos basados en uso o híbridos [37]. Para el asistente virtual, se proponen tres alternativas de monetización:

4.2.1 Alternativa 1: Modelo de Suscripción por Niveles

Este modelo ofrece predictibilidad de ingresos y facilita la planificación presupuestaria tanto para el proveedor como para el cliente. Se estructura en tres niveles:

- **Plan Básico:** Acceso al asistente con límite de consultas mensuales, horario extendido (no 24/7), soporte por email.
- **Plan Profesional:** Consultas ilimitadas, disponibilidad 24/7, integración con CRM, reportes mensuales, soporte prioritario.
- **Plan Enterprise:** Todo lo anterior más personalización de modelo, API dedicada, SLA garantizado, gerente de cuenta dedicado.

4.2.2 Alternativa 2: Modelo Basado en Uso

El cobro se basa en métricas de consumo real: número de consultas procesadas, minutos de audio transcritos, o documentos analizados. Este modelo es percibido como más justo por los clientes ya que pagan proporcionalmente a lo que consumen [38]. Es particularmente atractivo para empresas con volúmenes de consultas variables o estacionales.

4.2.3 Alternativa 3: Modelo Híbrido

Combina una tarifa base de suscripción que incluye un volumen determinado de consultas, con cobros adicionales por uso excedente. Según Andreessen Horowitz, los modelos híbridos funcionan especialmente bien para productos SaaS con componentes de IA, donde el costo marginal de procesamiento puede variar significativamente [39]. Este modelo balancea la predictibilidad de ingresos con la flexibilidad para capturar valor de clientes de alto volumen.

4.3 VALIDACIÓN DEL MODELO DE NEGOCIO

Para validar el modelo de negocio en el contexto del MVP, se deben especificar los ítems a evaluar una vez implementada la solución y realizadas las pruebas piloto. Esta validación permitirá obtener datos concretos para la mejora continua y potencial comercialización.

4.3.1 Marco de Validación

- **Validación de la Propuesta de Valor**
 - ¿El asistente resuelve efectivamente el problema de tiempos de respuesta?
 - ¿Los usuarios perciben valor diferencial frente a la atención tradicional?
 - ¿Las ventajas competitivas (24/7, multicanal, precisión) se mantienen en operación real?
- **Validación del Segmento de Clientes**
 - ¿Los segmentos definidos (grandes, medianos, PYMES) tienen necesidades diferenciadas?
 - ¿Los canales seleccionados (WhatsApp, Telegram) son los preferidos por cada segmento?
 - ¿Se requieren ajustes en la segmentación basados en patrones de uso reales?
- **Validación de Fuentes de Ingresos**
 - ¿El modelo de precios propuesto es aceptable para los clientes objetivo?
 - ¿Existe disposición a pagar por el servicio? ¿A qué precio?
 - ¿El modelo híbrido (suscripción + uso) es preferido sobre alternativas?
- **Validación de Estructura de Costos**
 - ¿Los costos operativos reales coinciden con las proyecciones?
 - ¿La infraestructura local es más costo-efectiva que alternativas cloud a largo plazo?
 - ¿Se requieren recursos adicionales no contemplados inicialmente?
- **Validación del MVP**
 - ¿Las funcionalidades mínimas son suficientes para demostrar valor?
 - ¿El tiempo de implementación (3-4 meses) es realista?

- ¿El MVP genera retroalimentación suficiente para iterar?

4.3.2 Métricas de Éxito

Para medir el desempeño del modelo de negocio, se definen indicadores clave en cuatro categorías:

Categoría	Métrica	Objetivo
Operacional	Tasa de resolución automatizada	>80% de consultas resueltas sin intervención humana
Operacional	Tiempo promedio de respuesta	<8 minutos para consultas de complejidad media
Calidad	Precisión de respuestas	>90% de respuestas técnicamente correctas
Satisfacción	Net Promoter Score (NPS)	>50 (categoría "excelente")
Adopción	Tasa de adopción	>70% de usuarios activos en primer mes
Financiero	Reducción de costos operativos	>40% vs. modelo tradicional en 12 meses

Tabla 6 Métricas de Éxito

5. PROPUESTA DE LA SOLUCIÓN TECNOLÓGICA

La presente propuesta de solución tecnológica define la arquitectura y componentes para el desarrollo e implementación de un Producto Mínimo Viable (MVP) de un agente de servicios inteligente basado en Inteligencia Artificial, diseñado para optimizar los procesos de asesoría y atención al cliente en el contexto de un Operador Económico Autorizado en Colombia.

La solución responde a las deficiencias identificadas en el sistema tradicional de asesoría: demoras en respuesta operativa, disponibilidad limitada, sobrecostos en recursos humanos y oportunidades de mejora en la satisfacción del cliente. El MVP se fundamenta en una estrategia de implementación pragmática que prioriza la viabilidad económica y el control operativo mediante el uso de infraestructura local propia y modelos de inteligencia artificial de código abierto.

Esta decisión arquitectónica responde a tres factores críticos:

- **Eliminación de dependencias:** Se evitan proveedores cloud externos y costos operativos recurrentes elevados, manteniendo control total sobre la infraestructura.
- **Soberanía de datos:** Se garantiza la privacidad y protección completa de datos sensibles de clientes en cumplimiento estricto con la Ley 1581 de 2012 de Protección de Datos Personales.
- **Personalización profunda:** Se habilita la capacidad de ajustar el modelo de IA con datos propios del dominio de comercio exterior colombiano mediante técnicas de ajuste fino.

La solución integra capacidades de procesamiento multimodal (texto y voz) a través de WhatsApp y Telegram como canales de interacción primarios, aprovechando su penetración superior al 90% en el mercado colombiano [29].

5.1 Componente de Orquestación: n8n

Un componente arquitectónico fundamental que distingue esta propuesta es la integración de n8n como plataforma de automatización central [31]. n8n actúa como el orquestador principal que coordina todos los flujos de trabajo entre los diferentes componentes del sistema, desde la recepción inicial de mensajes hasta la generación y distribución de reportes de cumplimiento e historial de interacción. Esta capa de automatización proporciona tres ventajas estratégicas significativas:

- **Agilidad en modificaciones:** Los flujos de trabajo se configuran visualmente mediante interfaz sin código, permitiendo ajustes operativos sin requerir reescritura de código o redespiegue de servicios.
- **Visibilidad y trazabilidad:** Cada ejecución de flujo de trabajo se registra con registros detallados, facilitando la depuración, auditorías y optimización de procesos.
- **Extensibilidad simple:** La adición de nuevas integraciones (correo electrónico, Slack, CRM, etc.) se realiza mediante conectores preconstruidos sin desarrollo personalizado, acelerando la evolución del sistema posterior al MVP.

5.2 Arquitectura de la Solución

La arquitectura propuesta se estructura en seis capas funcionales principales que operan de manera coordinada para proporcionar un servicio de asesoría inteligente disponible las 24 horas del día, los 7 días de la semana.

5.2.1 Capa de Interacción Multicanal

- **Tecnología:** WhatsApp Business API / Telegram Bot API
Esta capa gestiona la comunicación directa con los usuarios finales a través de los canales de mensajería más populares. Sus características principales incluyen:
 - Gestión unificada de mensajes de texto y notas de voz
 - Procesamiento de mensajes de voz para consultas complejas
 - Interfaz familiar para usuarios del sector logístico
 - Capacidad de procesamiento de múltiples mensajes concurrentes

5.2.2 Capa de Orquestación y Automatización

- **Tecnología:** n8n (autoalojado)
Esta capa actúa como el cerebro operativo del sistema, coordinando todos los flujos de trabajo:
 - Coordinación de flujos de trabajo entre todas las capas del sistema
 - Gestión de estados conversacionales y contexto de usuarios
 - Enrutamiento inteligente basado en tipo de consulta y carga del sistema

- Integración nativa con más de 400 servicios externos para futura expansión [31]
- Monitoreo en tiempo real de flujos y detección de anomalías

5.2.3 Capa de Procesamiento de Lenguaje Natural

- **Tecnología:** Llama 3.1 (8B/70B parámetros) + Whisper
Esta capa constituye el núcleo de inteligencia artificial del sistema, encargada de comprender las consultas y generar respuestas precisas:
 - Comprensión contextual de consultas técnicas sobre comercio exterior
 - Generación de respuestas precisas basadas en normativa aduanera colombiana
 - Procesamiento multilingüe (español, inglés, portugués) para clientes internacionales [30]
 - Capacidad de ajuste fino con terminología específica del sector
 - Transcripción de audio mediante Whisper con precisión superior al 92%

5.2.4 Capa de Gestión de Conocimiento

- **Tecnología:** PostgreSQL + Qdrant
Esta capa implementa una arquitectura híbrida que combina bases de datos relacionales y vectoriales para optimizar diferentes tipos de consultas:
- **PostgreSQL (datos estructurados):**
 - Registro de consentimientos con marca de tiempo y hash de validación
 - Historial de consultas y respuestas
 - Métricas de uso y patrones de consulta
- **Qdrant (búsqueda semántica):**
 - Indexación de posiciones arancelarias y documentación normativa
 - Almacenamiento de vectores de documentos (embeddings)
 - Búsqueda por similitud semántica con alta precisión
 - Actualización incremental de regulaciones sin interrupciones de servicio

5.2.5 Capa de Infraestructura Local

- **Tecnología:** Estación de trabajo con GPU dedicada
La especificación de hardware recomendada para el MVP considera los requisitos de memoria y procesamiento del modelo Llama 3.1 [30]:

Componente	Especificación Recomendada
Procesador	Intel XeonE3 o AMD Ryzen 9 (mínimo 4 núcleos)
GPU	NVIDIA QUADRO P4000 (8GB VRAM) para Llama 3.1 8B con cuantización
Memoria RAM	64GB DDR5 para carga del modelo y procesamiento concurrente
Almacenamiento	1TB NVMe SSD (lectura >3,000 MB/s) para modelos y bases de datos
Sistema Operativo	Ubuntu 22.04 LTS o Windows 11 Pro con WSL2
Redundancia (MVP)	No incluida en fase MVP; planificada para producción

Tabla 7 Capa de Infraestructura Local

5.2.6 Capa de Cumplimiento Normativo

Esta capa garantiza el cumplimiento de la Ley 1581 de 2012 y el Decreto 1377 de 2013 mediante los siguientes componentes:

- Módulo de gestión de consentimiento con registro explícito e inequívoco
- Registro auditable de todas las interacciones con sello de tiempo UTC
- Generación automática de reportes con métricas de transaccionalidad
- Política de retención de datos configurable según normativa vigente

5.3 Flujo Operativo del Sistema

➤ Proceso de Atención de Consultas

El flujo de atención de consultas se estructura en cinco etapas secuenciales que garantizan una experiencia de usuario consistente y el cumplimiento normativo:

Etapas 1 - Recepción y Validación:

- Usuario envía consulta vía WhatsApp o Telegram (texto, voz)
- n8n valida formato y extrae metadatos del mensaje
- Sistema verifica existencia de consentimiento previo

Etapas 2 - Gestión de Consentimiento:

- Si no existe consentimiento: solicitud automática con términos claros
- Captura de respuesta explícita (Sí/No)
- Almacenamiento en PostgreSQL con hash SHA-256 y marca de tiempo

Etapas 3 - Procesamiento de Consulta:

- Conversión voz a texto si aplica (Whisper)
- Análisis de intención y extracción de entidades
- Búsqueda semántica en Qdrant de información relevante
- Generación de contexto aumentado (RAG) con documentos pertinentes

Etapas 4 - Generación de Respuesta:

- Llama 3.1 procesa consulta con contexto recuperado
- Aplicación de plantilla de respuesta según tipo de cliente
- Validación de coherencia y completitud

Etapas 5 - Entrega y Registro:

- Envío de respuesta por el canal original
- Almacenamiento de interacción completa en PostgreSQL
- Actualización de métricas de servicio
- Activación de reportes automáticos si corresponde

5.4 Funcionalidades del MVP

El alcance funcional del MVP contempla las capacidades esenciales requeridas para validar la propuesta de valor con usuarios reales, manteniendo un alcance acotado que facilita implementación ágil en 3-4 meses.

➤ **Procesamiento Multimodal**

- **Texto:** Consultas escritas sobre clasificación arancelaria, requisitos documentales, procedimientos aduaneros
- **Voz:** Notas de audio convertidas a texto con precisión superior al 92%

➤ **Gestión de Datos Personales**

- Captura estructurada: nombre completo, correo electrónico, celular, tipo de cliente
- Consentimiento granular por tipo de dato
- Trazabilidad completa de uso de datos

➤ **Inteligencia de Negocio**

- Clasificación automática de consultas por categoría
- Alertas de consultas recurrentes no resueltas
- Sugerencias de mejora en base de conocimiento

➤ **Generación de Reportes**

- **Reporte diario:** Volumen de consultas, tipos, tiempos de respuesta
- **Reporte semanal:** Análisis de satisfacción y temas emergentes
- **Reporte mensual:** Indicadores clave de servicio y recomendaciones
- Envío automático por correo electrónico a partes interesadas definidas

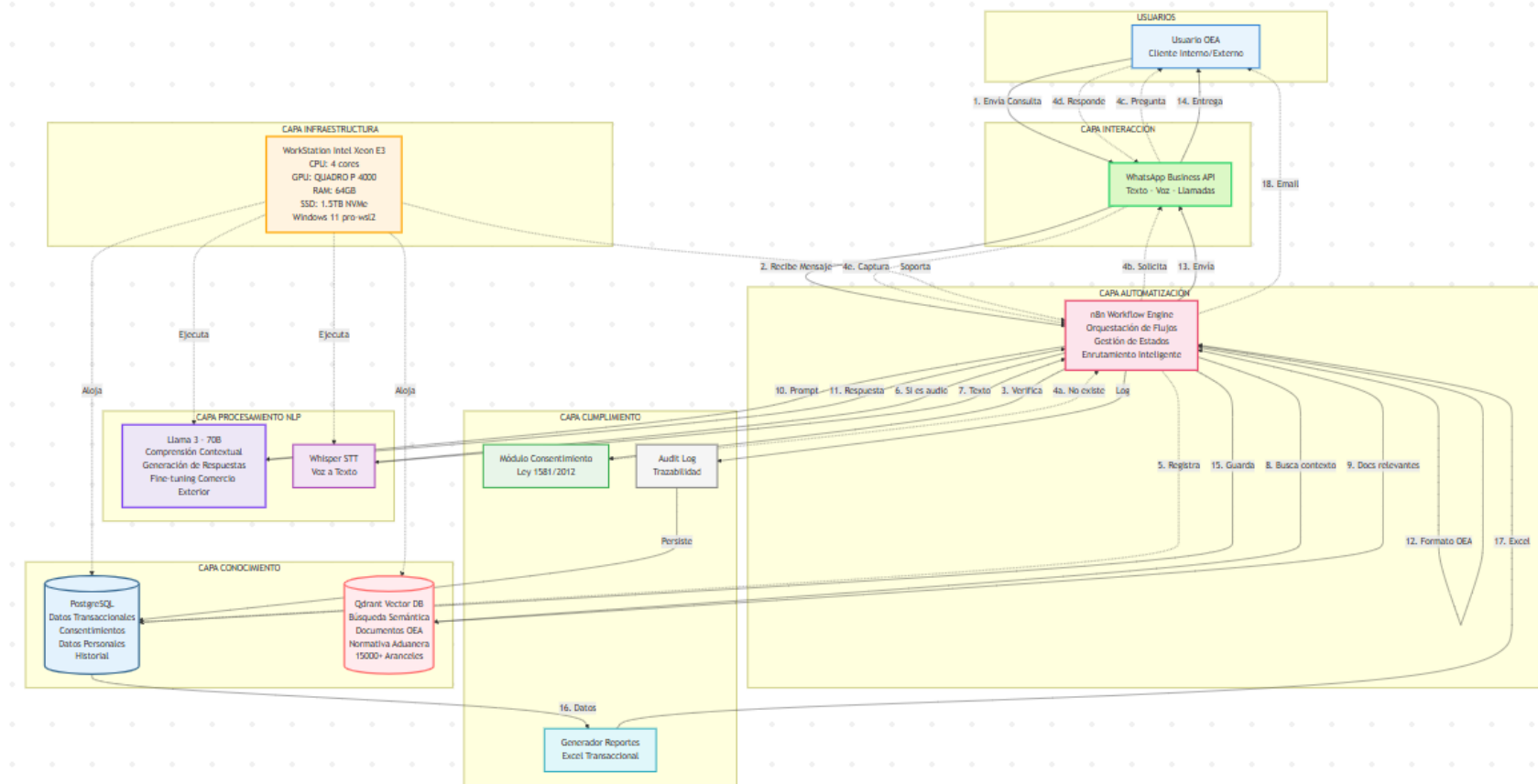


Figura 8 Diagrama de Arquitectura Solución Funcional
Fuente: Elaboración propia

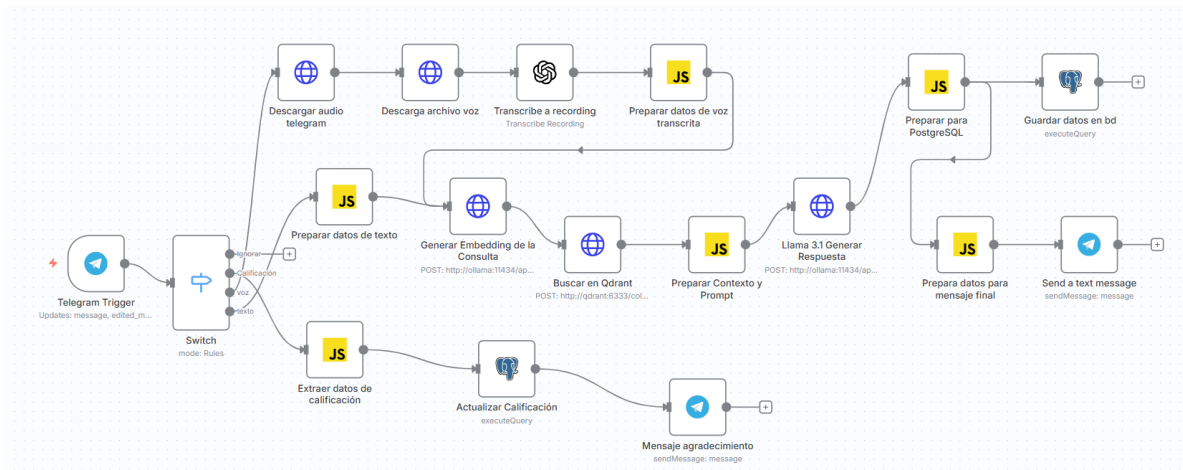


Figura 9 Workflow n8n MVP Solución Funcional
Fuente: Elaboración propia

```

n8n=# CREATE DATABASE chat_history;
CREATE DATABASE
n8n=# \c chat_history
You are now connected to database "chat_history" as user "root".
chat_history=# CREATE TABLE conversaciones (
chat_history(#      id SERIAL PRIMARY KEY,
chat_history(#      chat_id BIGINT NOT NULL,
chat_history(#      usuario VARCHAR(255) NOT NULL,
chat_history(#      telefono VARCHAR(50) NOT NULL,
chat_history(#      consulta TEXT NOT NULL,
chat_history(#      respuesta TEXT NOT NULL,
chat_history(#      es_voz BOOLEAN DEFAULT false,
chat_history(#      duracion_segundos INTEGER DEFAULT 0,
chat_history(#      calificacion INTEGER,
chat_history(#      session_id VARCHAR(100),
chat_history(#      timestamp TIMESTAMP DEFAULT CURRENT_TIMESTAMP
chat_history(# );
CREATE TABLE
chat_history=# CREATE INDEX idx_chat_id ON conversaciones(chat_id);
CREATE INDEX
chat_history=# CREATE INDEX idx_session_id ON conversaciones(session_id);
CREATE INDEX
chat_history=# CREATE INDEX idx_timestamp ON conversaciones(timestamp);
CREATE INDEX
chat_history=# \dt
          List of relations
Schema |      Name      | Type  | Owner
-----|-----|-----|-----
public | conversaciones | table | root
(1 row)

chat_history=# \d conversaciones
          Table "public.conversaciones"
   Column   |      Type      | Collation | Nullable | Default
-----|-----|-----|-----|-----
id          | integer        |           | not null | nextval('conversaciones_id_seq'::regclass)
chat_id     | bigint         |           | not null |
usuario     | character varying(255) |         | not null |
telefono    | character varying(50) |         | not null |
consulta    | text           |         | not null |
respuesta   | text           |         | not null |
es_voz      | boolean        |         |         | false
duracion_segundos | integer        |         |         | 0
calificacion | integer        |         |         |
session_id  | character varying(100) |         |         |
timestamp   | timestamp without time zone |         |         | CURRENT_TIMESTAMP
Indexes:
    "conversaciones_pkey" PRIMARY KEY, btree (id)
    "idx_chat_id" btree (chat_id)
    "idx_session_id" btree (session_id)
    "idx_timestamp" btree ("timestamp")

```

Figura 10 Creación y parametrización Base de Datos Postgres
Fuente: Elaboración propia

5.5 Modelado General del Workflow

El workflow implementado en n8n orquesta todo el flujo de atención del asistente virtual, desde la recepción del mensaje del usuario en Telegram hasta el envío de la respuesta generada por inteligencia artificial. El diseño contempla tres rutas de procesamiento según el tipo de entrada: mensajes de voz, mensajes de texto y calificaciones del servicio.

La arquitectura sigue el patrón RAG (Generación Aumentada por Recuperación), donde las consultas del usuario se convierten en vectores para buscar información relevante en la base de conocimiento antes de generar la respuesta con el modelo de lenguaje Llama 3.1.

1) Telegram Trigger

- **Tipo:** Trigger (Disparador)
- **Función:** Es el punto de entrada del workflow. Escucha activamente los mensajes entrantes en el bot de Telegram configurado. Detecta tanto mensajes nuevos como mensajes editados (Updates: message, edited_m...). Cada vez que un usuario envía un mensaje al bot, este nodo captura la información y la pasa al siguiente nodo para su procesamiento.

2) Switch

- **Tipo:** Enrutador condicional (mode: Rules)
- **Función:** Clasifica el mensaje entrante y lo dirige a la ruta de procesamiento correspondiente según su tipo. Define cuatro salidas posibles: "Ignorar" para mensajes que no requieren procesamiento, "Calificación" para cuando el usuario envía una valoración del servicio, "voz" para notas de audio, y "texto" para mensajes escritos. Esta bifurcación permite que cada tipo de entrada reciba el tratamiento especializado que necesita.

3) Ruta de procesamiento de voz

➤ Descargar audio telegram

- **Tipo:** HTTP Request
- **Función:** Obtiene la información del archivo de audio desde la API de Telegram. Cuando un usuario envía una nota de voz, Telegram almacena el archivo en sus servidores y proporciona un identificador. Este nodo consulta la API para obtener la ruta de descarga del archivo de audio.

4) Descarga archivo voz

- **Tipo:** HTTP Request
- **Función:** Descarga el archivo de audio real desde los servidores de Telegram usando la ruta obtenida en el paso anterior. El archivo descargado (generalmente en formato OGG) queda disponible en memoria para ser procesado por el servicio de transcripción.

5) Transcribe a recording

- **Tipo:** Transcribe Recording (integración con servicio de transcripción)
- **Función:** Convierte el audio a texto utilizando un modelo de reconocimiento de voz (como Whisper). Recibe el archivo de audio descargado y devuelve la transcripción en texto plano. Esta conversión es fundamental para poder procesar consultas de voz con el mismo pipeline que se usa para texto.

6) Preparar datos de voz transcrita

- **Tipo:** Code (JavaScript)
- **Función:** Procesa y estructura la transcripción obtenida. Normaliza el texto, elimina caracteres especiales o ruido de la transcripción, y formatea los datos en la estructura que esperan los nodos posteriores. Unifica el formato de salida para que sea compatible con la ruta de texto.

RUTA DE PROCESAMIENTO DE TEXTO

7) Preparar datos de texto

- **Tipo:** Code (JavaScript)
- **Función:** Extrae y prepara el contenido del mensaje de texto enviado por el usuario. Limpia el texto, extrae metadatos relevantes (ID del usuario, nombre, timestamp) y estructura la información para los pasos siguientes del procesamiento RAG.

8) PIPELINE RAG [Común para voz y texto] - (*Generación Aumentada por Recuperación*)

9) Generar Embedding de la Consulta

- **Tipo:** HTTP Request (POST a <http://ollama:11434/api...>)
- **Función:** Convierte el texto de la consulta del usuario en un vector numérico (embedding) usando el servicio de Ollama. Este vector representa el significado semántico de la consulta y permite buscar documentos

similares en la base de conocimiento. Es el primer paso crítico del patrón RAG.

10) Buscar en Qdrant

- **Tipo:** HTTP Request (POST a <http://qdrant:6333/col...>)
- **Función:** Realiza una búsqueda por similitud semántica en la base de datos vectorial Qdrant. Envía el embedding generado y recupera los fragmentos de documentos más relevantes para la consulta del usuario. Estos documentos contienen información sobre normativa aduanera, procedimientos OEA y respuestas a preguntas frecuentes.

11) Preparar Contexto y Prompt

- **Tipo:** Code (JavaScript)
- **Función:** Combina los documentos recuperados de Qdrant con la consulta original del usuario para construir el prompt completo que se enviará al modelo de lenguaje. Incluye instrucciones del sistema, el contexto relevante recuperado y la pregunta del usuario, estructurados de manera que el modelo genere respuestas precisas y fundamentadas.

12) Llama 3.1 Generar Respuesta

- **Tipo:** HTTP Request (POST a <http://ollama:11434/api...>)
- **Función:** Envía el prompt construido al modelo Llama 3.1 ejecutándose localmente a través de Ollama. El modelo procesa el contexto proporcionado y genera una respuesta en lenguaje natural que responde a la consulta del usuario basándose en la información recuperada de la base de conocimiento.

PERSISTENCIA Y ENVÍO DE RESPUESTA

13) Preparar para PostgreSQL

- **Tipo:** Code (JavaScript)
- **Función:** Estructura todos los datos de la interacción para su almacenamiento en la base de datos relacional. Prepara campos como: ID del usuario, consulta original, tipo de entrada (voz/texto), respuesta generada, documentos consultados, marca de tiempo y métricas de procesamiento. Esta información es esencial para auditoría, análisis y mejora continua del sistema.

14) Guardar datos en bd

- **Tipo:** PostgreSQL (executeQuery)
- **Función:** Ejecuta la inserción de los datos de la interacción en PostgreSQL. Almacena el registro completo de la conversación para cumplir con los requisitos de trazabilidad establecidos por la normativa de protección de datos (Ley 1581/2012) y para generar los reportes de servicio.

15) Prepara datos para mensaje final

- **Tipo:** Code (JavaScript)
- **Función:** Formatea la respuesta generada por Llama 3.1 para su envío a través de Telegram. Aplica formato de texto (Markdown si corresponde), ajusta la longitud si excede límites de la API, y estructura el mensaje según las plantillas definidas para cada tipo de consulta.

16) Send a text message

- **Tipo:** Telegram (sendMessage: message)
- **Función:** Envía la respuesta final al usuario a través de Telegram. Utiliza el ID del chat original para responder en la misma conversación donde se recibió la consulta. Este nodo cierra el ciclo de atención para consultas de texto y voz.

RUTA DE CALIFICACIÓN DEL SERVICIO

17) Extraer datos de calificación

- **Tipo:** Code (JavaScript)
- **Función:** Procesa el mensaje de calificación enviado por el usuario (por ejemplo, una puntuación numérica o estrellas). Extrae el valor de la calificación, identifica la interacción que está siendo evaluada y prepara los datos para actualizar el registro correspondiente.

18) Actualizar Calificación

- **Tipo:** PostgreSQL (executeQuery)
- **Función:** Ejecuta una consulta UPDATE en PostgreSQL para registrar la calificación del usuario asociada a la interacción correspondiente. Esta información alimenta las métricas de satisfacción (NPS, CSAT) y los reportes de calidad del servicio.

19)Mensaje agradecimiento

- **Tipo:** Telegram (sendMessage: message)
- **Función:** Envía un mensaje de agradecimiento al usuario por proporcionar su calificación. Este cierre cortés mejora la experiencia del usuario y confirma que su retroalimentación fue recibida correctamente.

20)Resumen de Tecnologías Integradas

Tecnología	Rol en el Workflow
Telegram Bot API	Canal de comunicación con usuarios (entrada y salida de mensajes)
Whisper / Transcribe	Conversión de notas de voz a texto (speech-to-text)
Ollama	Servidor local para ejecutar modelos LLM y generar embeddings
Llama 3.1	Modelo de lenguaje para generación de respuestas en lenguaje natural
Qdrant	Base de datos vectorial para búsqueda semántica (RAG)
PostgreSQL	Almacenamiento de interacciones, usuarios y calificaciones
JavaScript (Code nodes)	Transformación de datos y lógica de negocio personalizada

Figura 11 Resumen de Tecnologías Integradas

Este workflow representa la implementación real del MVP descrito en la Sección 5 de la monografía, demostrando cómo n8n actúa como orquestador central que coordina todas las tecnologías de la arquitectura propuesta.

6. ANÁLISIS DEL PROCESO DE TRANSFORMACIÓN DIGITAL

La implementación del asistente virtual inteligente en un Operador Económico Autorizado representa un caso paradigmático de transformación digital que trasciende la mera adopción tecnológica. Este proceso reconfigura fundamentalmente los procesos operativos, la propuesta de valor y el modelo de relacionamiento con clientes.

La transformación digital, en su mejor comprensión, implica adoptar procesos de negocio y prácticas que ayuden a la organización a competir efectivamente en un mundo cada vez más digital [43]. No se trata simplemente de implementar nuevas tecnologías, sino de adaptar toda la organización a las tendencias digitales del entorno competitivo.

6.1 Marco de Análisis: Modelo de Madurez Digital

El análisis del proceso se estructura mediante el marco de Madurez Digital desarrollado por MIT Sloan Management Review en colaboración con Capgemini y Deloitte [44]. Este modelo evalúa la transformación en dos dimensiones principales:

- **Intensidad Digital:** El nivel de inversión en iniciativas habilitadas por tecnología destinadas a cambiar cómo opera la empresa.
- **Intensidad de Gestión de la Transformación:** El nivel de inversión en las capacidades de liderazgo necesarias para crear transformación digital dentro de la organización.

Según la investigación de MIT Sloan, las empresas digitalmente maduras superan a sus competidores menos maduros en métricas financieras clave: aquellas maduras en intensidad digital generan entre 6% y 9% más ingresos por empleado, mientras que las maduras en gestión de la transformación son entre 9% y 26% más rentables que el promedio de su industria [44].

6.2 Dimensiones de Transformación Evaluadas

El proyecto de asistente virtual genera transformación en cinco dimensiones clave que abarcan tanto aspectos tecnológicos como organizacionales:

➤ **Estrategia Digital**

La IA conversacional evoluciona de ser una herramienta de soporte a convertirse en una competencia central diferenciadora. Las organizaciones digitalmente maduras tienen horizontes de planificación estratégica consistentemente más largos, con casi el 30% mirando cinco años o más hacia adelante, comparado con solo el 13% de las organizaciones menos maduras [45]. El asistente virtual posiciona la tecnología como generador primario de valor, no como elemento periférico.

➤ **Cultura Organizacional**

La implementación requiere un cambio cultural hacia la colaboración interfuncional. Más del 70% de los encuestados de empresas digitalmente maduras indican que sus organizaciones están cada vez más organizadas alrededor de equipos interfuncionales, comparado con solo el 28% de empresas en etapas tempranas de desarrollo digital [45]. El asistente virtual demanda integración entre áreas de tecnología, operaciones, servicio al cliente y cumplimiento normativo.

➤ **Capacidades Tecnológicas**

El proyecto desarrolla capacidades tecnológicas en procesamiento de lenguaje natural, gestión de conocimiento mediante bases de datos vectoriales, orquestación de flujos de trabajo automatizados y cumplimiento normativo digitalizado. Estas capacidades constituyen activos reutilizables que pueden expandirse a otros procesos de la organización.

➤ **Experiencia del Cliente**

La transformación de la experiencia del cliente es uno de los nueve elementos de la transformación digital identificados por MIT Sloan [46]. El asistente virtual transforma radicalmente esta experiencia: de tiempos de respuesta de 48-72 horas a menos de 8 minutos, de disponibilidad limitada (8AM-6PM) a servicio continuo 24/7, y de canal único (teléfono/email) a multicanal (WhatsApp, Telegram, voz).

➤ Modelos Operativos

Los procesos operativos se reconfiguran desde un modelo basado en recursos humanos hacia un modelo híbrido donde la IA maneja consultas rutinarias (70-85% del volumen) mientras los asesores humanos se concentran en casos complejos de alto valor. Esta transformación no elimina el factor humano, sino que lo reposiciona hacia actividades que requieren juicio experto y construcción de relaciones.

6.3 Evaluación del Nivel de Madurez Digital

El modelo de madurez digital clasifica a las organizaciones en cuatro cuadrantes según sus niveles de intensidad digital e intensidad de gestión de la transformación [44]:

Cuadrante	Características
Principiantes	Baja intensidad digital y baja gestión de transformación. Hacen poco con tecnologías avanzadas.
Fashionistas	Alta intensidad digital pero baja gestión. Implementan tecnología sin estrategia clara de transformación.
Conservadores	Alta gestión pero baja intensidad digital. Tienen visión pero avanzan lentamente en implementación.
Digirati	Alta intensidad en ambas dimensiones. Combinan actividad digital con liderazgo fuerte. Superan a competidores.

Tabla 8 Evaluación del Nivel de Madurez Digital

El proyecto del asistente virtual posiciona al OEA en una trayectoria hacia el cuadrante Digirati, combinando implementación tecnológica avanzada (modelo LLM, arquitectura RAG, automatización con n8n) con elementos de gestión de la transformación (estrategia clara, métricas definidas, enfoque en experiencia del cliente).

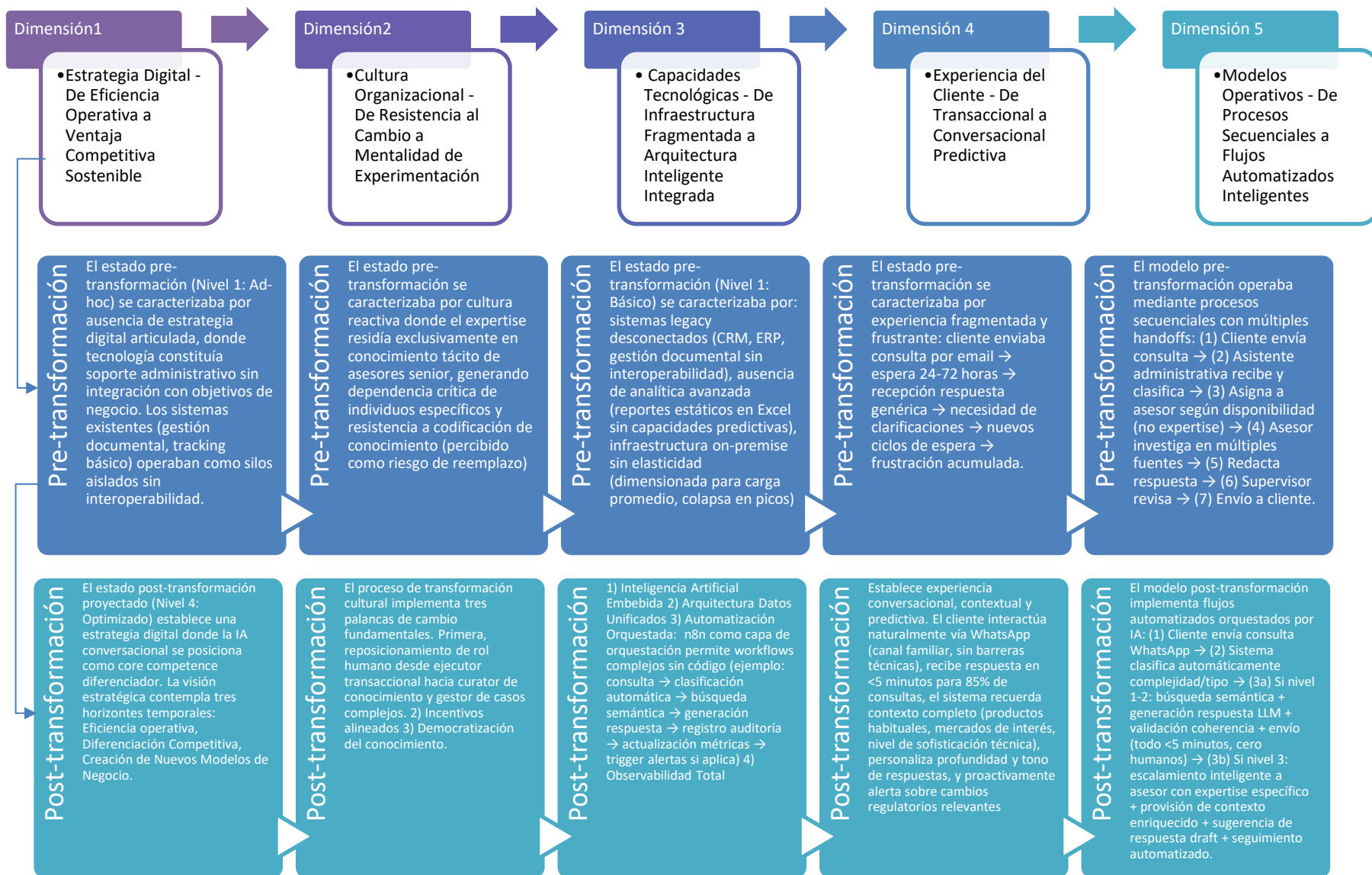


Figura 12 Modelo de Análisis Madurez Digital de MIT Sloan
 Fuente: Elaboración propia

7. ASPECTOS LEGALES Y CONTRATACIÓN

La implementación del asistente virtual inteligente opera en la intersección de múltiples marcos regulatorios nacionales e internacionales que determinan la viabilidad legal del proyecto. El análisis se estructura en cinco dimensiones: cumplimiento normativo de protección de datos, regulación aduanera, responsabilidad civil y profesional, propiedad intelectual y licenciamiento, y contratos marco para prestación del servicio.

7.1 Cumplimiento Normativo de Protección de Datos Personales

➤ Marco Regulatorio Nacional

La Ley Estatutaria 1581 de 2012 establece el régimen general de protección de datos personales en Colombia, concretando el derecho fundamental de todas las personas frente al correcto uso de su información [47]. Esta ley, junto con su Decreto reglamentario 1377 de 2013, establece el marco vinculante para el tratamiento de datos personales.

El artículo 4 de la Ley 1581 define "Tratamiento" como cualquier operación sobre datos personales, incluyendo recolección, almacenamiento, uso, circulación o supresión. El asistente virtual ejecuta todas estas operaciones, configurándose el OEA como Responsable del Tratamiento según el artículo 3 literal k.

El Decreto 1377 de 2013 (artículo 13) exige que la autorización sea "previa, expresa e informada". El sistema cumple mediante:

- **Prevía:** No se procesa ningún dato hasta recibir consentimiento explícito.
- **Expresa:** Se requiere acción afirmativa inequívoca ("Sí" textual, no inferida de continuar interacción).
- **Informada:** Se comunican claramente finalidades, datos recopilados, destinatarios y derechos del titular.

➤ Circular Externa 002 de 2024 - SIC

La Superintendencia de Industria y Comercio, mediante la Circular Externa 002 del 21 de agosto de 2024, estableció lineamientos específicos para el tratamiento de datos personales en sistemas de inteligencia artificial [48]. Esta circular aclara que

las disposiciones de la Ley 1581 de 2012 aplican sin distinción del medio utilizado, ya sea tecnológico o manual.

La Circular establece que el tratamiento de datos personales en sistemas de IA debe basarse en cuatro criterios:

- **Idoneidad:** El tratamiento debe ser adecuado y cumplir objetivos legítimos.
- **Necesidad:** La recolección debe limitarse a lo estrictamente necesario.
- **Razonabilidad:** Debe ser coherente con valores constitucionales.
- **Proporcionalidad:** Debe mantener equilibrio justo entre beneficios y afectaciones a la privacidad.

Adicionalmente, la Circular exige realizar y documentar un estudio de impacto de privacidad antes del diseño y desarrollo de sistemas de IA que traten datos personales con alto riesgo de afectación a los titulares [48].

➤ **Marco Regulatorio Internacional – GDPR**

Aunque el OEA opera principalmente en Colombia, clientes internacionales (especialmente europeos) pueden estar protegidos por el Reglamento General de Protección de Datos (GDPR) de la Unión Europea. El GDPR aplica si el OEA ofrece servicios a interesados en la UE (artículo 3.2.a), lo cual ocurre al asesorar importadores o exportadores europeos. El sistema se diseña para cumplir con ambos marcos normativos.

7.2 Regulación Aduanera y Validez de Asesorías Automatizadas

El marco regulatorio de la DIAN establece los lineamientos para la prestación de servicios de asesoría aduanera. Las asesorías generadas por sistemas de IA son válidas bajo supervisión humana, lo cual el sistema garantiza mediante:

- Escalamiento a asesores humanos para consultas complejas o de alto riesgo.
- Registro completo de trazabilidad de cada respuesta generada.
- Revisión periódica de calidad de respuestas por personal calificado.
- Advertencias claras sobre el carácter informativo (no vinculante) de las respuestas automatizadas.

7.2Responsabilidad Civil y Profesional

El Código Civil Colombiano (artículo 2341) establece responsabilidad extracontractual para quien cause daño a otro por delito o culpa. En el contexto del proyecto, se identifican dos escenarios de responsabilidad potencial:

➤ Escenario 1: Error en Asesoría Automatizada

Mitigación implementada:

- Declaración expresa de que las respuestas son informativas y no constituyen asesoría legal vinculante.
- Recomendación de verificación con fuentes oficiales para decisiones críticas.
- Registro de trazabilidad que permite auditar el origen de cada respuesta.

➤ Escenario 2: Brecha de Seguridad o Filtración de Datos

Mitigación implementada:

- Cifrado AES-256 de datos en reposo.
- Infraestructura local que elimina exposición a terceros.
- Monitoreo de anomalías y alertas en tiempo real.

7.3 Propiedad Intelectual y Licenciamiento

El Código Civil (artículo 671) define propiedad como derecho de usar, gozar y disponer. La Decisión Andina 351 (artículo 23) establece que el software se protege como obra literaria bajo derecho de autor. La titularidad del asistente virtual involucra múltiples componentes con regímenes distintos:

Componente	Licencia	Titularidad
Modelo LLM Base (Llama 3.1)	Llama Community License (Meta)	Meta Platforms, Inc. Uso permitido bajo licencia.
Modelo con ajuste fino	Derivada de Llama License	OEA (mejoras propias sobre modelo base).
Base de Conocimiento	Propiedad exclusiva	OEA (documentación, Q&A, procedimientos).
Flujos de trabajo n8n	Fair-code (n8n License)	OEA (configuraciones personalizadas).

Tabla 9 Propiedad Intelectual y Licenciamiento

➤ **Contratos Marco para Prestación del Servicio**

Para operación comercial escalable, se diseña una arquitectura contractual con tres documentos interrelacionados:

- **Términos y Condiciones de Servicio:** Documento público que regula el uso del servicio por usuarios finales, incluyendo limitaciones de responsabilidad, política de privacidad y condiciones de uso aceptable.
- **Contrato Marco Empresarial:** Acuerdo para clientes corporativos que incluye niveles de servicio (SLA), condiciones de facturación, obligaciones de confidencialidad y procedimientos de escalamiento.
- **Acuerdo de Licenciamiento de Marca Blanca:** Para OEAs que deseen implementar la solución bajo su propia marca, regulando derechos de uso, personalización permitida y obligaciones de mantenimiento.

7.3 Marco Regulatorio de Contratación Electrónica

La Ley 527 de 1999 (Comercio Electrónico) y el Decreto 2364 de 2012 regulan la validez de contratos electrónicos en Colombia. Estas normas reconocen la equivalencia funcional de documentos electrónicos con documentos físicos, habilitando la celebración válida de contratos de servicio a través de medios digitales.

CONCLUSIONES

A lo largo de este trabajo hemos tenido la oportunidad de explorar un territorio que, hace apenas unos años, parecía reservado exclusivamente para grandes corporaciones tecnológicas: la implementación de inteligencia artificial conversacional en un contexto empresarial colombiano. Lo que comenzó como una inquietud académica sobre cómo mejorar la atención al cliente en operaciones de comercio exterior, se transformó en un ejercicio de diseño integral que nos llevó a cuestionar supuestos, investigar alternativas y proponer soluciones que consideramos viables y pertinentes para nuestra realidad.

Una de las principales lecciones que nos deja este proyecto es que la brecha tecnológica entre las grandes corporaciones y las empresas medianas se está reduciendo de manera significativa. Cuando analizamos las alternativas técnicas disponibles, encontramos que modelos de lenguaje como Llama 3.1 ofrecen capacidades que hace tres años solo estaban disponibles a través de costosas suscripciones a servicios en la nube. La posibilidad de ejecutar estos modelos en infraestructura local, con una inversión inicial que puede recuperarse en menos de un año, cambia completamente la ecuación económica para empresas como los Operadores Económicos Autorizados.

Esto no significa que implementar IA sea sencillo o que cualquier empresa pueda hacerlo sin preparación. Lo que significa es que las barreras ya no son principalmente económicas o tecnológicas, sino de conocimiento y capacidad organizacional. Y precisamente ahí es donde la gestión de servicios TIC cobra relevancia: no se trata solo de instalar software, sino de orquestar un cambio que involucra procesos, personas y cultura.

Antes de este trabajo, el concepto de soberanía de datos nos parecía más una preocupación de académicos y reguladores que una necesidad práctica del día a día empresarial. Sin embargo, al profundizar en los requisitos de la Ley 1581 de 2012 y especialmente en los lineamientos de la Circular 002 de 2024 de la Superintendencia de Industria y Comercio, entendimos que mantener el control sobre los datos de los clientes no es solo una obligación legal, sino una ventaja competitiva real.

En un sector como el de comercio exterior, donde se maneja información sensible sobre operaciones comerciales, clientes y proveedores internacionales, poder garantizar que esos datos nunca salen de la infraestructura controlada por la

empresa genera confianza. Y la confianza, como aprendimos al analizar el modelo de negocio, es quizás el activo más valioso en relaciones comerciales de largo plazo.

El marco de madurez digital de MIT Sloan nos ayudó a entender algo fundamental: muchas empresas fracasan en sus iniciativas tecnológicas no porque la tecnología falle, sino porque implementan herramientas sofisticadas sin transformar la manera en que la organización trabaja. Hemos visto casos en nuestro entorno laboral donde se adquieren sistemas costosos que terminan subutilizados porque nadie cambió los procesos ni preparó a las personas para usarlos.

Por eso insistimos a lo largo del documento en que el asistente virtual no puede verse como un proyecto de tecnología aislado, requiere equipos interfuncionales y, que los asesores humanos redefinan su rol hacia actividades de mayor valor, también requiere métricas claras y seguimiento constante. En otras palabras, requiere gestión. Y esa es precisamente nuestra área de especialización.

Durante el desarrollo de este trabajo nos dimos cuenta de que, uno de los mayores desafíos no fue entender la tecnología, sino comunicarla de manera que tuviera sentido para diferentes audiencias. Tuvimos que explicar conceptos técnicos como RAG (generación aumentada por recuperación) y bases de datos vectoriales a personas con formación en comercio exterior, y al mismo tiempo traducir requisitos de negocio y normativas legales a especificaciones técnicas.

Este rol de traductor o puente entre el mundo técnico y el mundo del negocio es, creemos, el valor diferencial que puede aportar un especialista en gestión de servicios TIC. No somos los expertos más profundos en machine learning ni los conocedores más detallados de la normativa aduanera, pero sí podemos hacer que estos dos mundos conversen y se entiendan.

Sería poco asertivo presentar este trabajo como si hubiéramos tenido todo claro desde el principio. La realidad es que tuvimos que replantear varios supuestos a medida que avanzábamos. Por ejemplo, inicialmente pensamos que podríamos usar hardware más económico, pero al investigar a fondo los requisitos de memoria de los modelos de lenguaje actuales, entendimos que subestimar la infraestructura compromete todo el proyecto. También tuvimos que reconocer que algunas referencias que encontramos inicialmente no eran verificables, lo que nos obligó a ser más rigurosos en la búsqueda de fuentes confiables.

Este proceso de corrección y mejora continua refleja, de alguna manera, lo que sería la implementación real de un proyecto como este: no se puede pretender que todo salga perfecto desde el primer día. Lo importante es tener la capacidad de identificar problemas, aprender de ellos y ajustar el rumbo.

Queremos ser honestos también en nuestra valoración final: creemos que el proyecto es técnicamente viable, económicamente justificable y legalmente posible. Los números del modelo de negocio muestran un retorno razonable, la arquitectura técnica utiliza componentes probados y disponibles, y el marco normativo colombiano, aunque exigente, no impide este tipo de implementaciones, sino que más bien las orienta.

Sin embargo, también reconocemos que pasar del diseño a la implementación real implica desafíos que van más allá de lo que pudimos abordar en este ejercicio académico. Encontrar el talento técnico adecuado, gestionar la resistencia al cambio dentro de la organización, mantener y actualizar el sistema una vez implementado, son todos retos que requerirían atención cuidadosa en un proyecto real.

El campo de la inteligencia artificial está evolucionando a una velocidad que hace difícil cualquier predicción a largo plazo. Cuando comenzamos este trabajo, algunos de los modelos que hoy consideramos como opción ni siquiera existían. Esto significa que cualquier implementación basada en estas tecnologías debe diseñarse con flexibilidad para adaptarse a nuevos desarrollos.

Pero también significa que las oportunidades seguirán creciendo. Si hoy es posible implementar un asistente virtual inteligente con recursos accesibles para una empresa mediana colombiana, es razonable pensar que en los próximos años las capacidades aumentarán mientras los costos continúan disminuyendo. Las empresas que comiencen ahora a desarrollar las competencias necesarias para aprovechar estas tecnologías estarán mejor posicionadas que las que esperen a que todo sea más fácil.

REFLEXIÓN FINAL

Este proyecto nos permitió integrar conocimientos de diferentes áreas de nuestra especialización: gestión de servicios, arquitectura tecnológica, modelos de negocio, cumplimiento normativo, seguridad y transformación digital. Más que un ejercicio de diseño técnico fue un ejercicio de pensamiento sistémico donde cada decisión en un área tenía implicaciones en las demás.

Esperamos que este trabajo pueda servir como referencia para profesionales y empresas que estén considerando iniciativas similares. No pretendemos haber encontrado la única manera de hacer las cosas, pero sí creemos haber documentado un camino razonable, con sus fortalezas y limitaciones claramente expuestas.

La inteligencia artificial conversacional llegó para quedarse. La pregunta ya no es si las empresas colombianas la adoptarán, sino cuándo y cómo lo harán. Ojalá este trabajo contribuya, aunque sea de manera modesta, a que esa adopción se haga de manera informada, responsable y alineada con las necesidades reales del contexto local.

REFERENCIAS

- [1]** ANALDEX - Asociación Nacional de Comercio Exterior. Informes y estudios sobre el sector logístico colombiano. Disponible en: <https://analdex.org/>
- [2]** Zendesk. "Tiempo de respuesta al cliente: qué es + 5 tips para mejorarlo". Estadísticas sobre expectativas de tiempo de respuesta. Disponible en: <https://www.zendesk.com.mx/blog/tiempo-de-respuesta/>
- [3]** ANALDEX con datos DANE. "Informe mensual de exportaciones colombianas diciembre de 2024". Las exportaciones alcanzaron USD FOB \$49.552 millones en 2024. Disponible en: <https://analdex.org/>
- [4]** Organización Mundial de Aduanas (OMA). "SAFE Framework of Standards to Secure and Facilitate Global Trade". Marco adoptado en 2005 que incluye el programa AEO/OEA. Disponible en: <https://www.wcoomd.org/>
- [5]** International Chamber of Commerce (ICC). "ICC releases updated recommendations on Authorised Economic Operator programmes", mayo 2024. Más de 80 programas OEA a nivel global. Disponible en: <https://iccwbo.org/>
- [6]** DIAN - Dirección de Impuestos y Aduanas Nacionales. "Empresas OEA Autorizadas". Lista oficial actualizada de operadores certificados. Disponible en: https://www.dian.gov.co/aduanas/oea/inicio/Paginas/Empresas_oea_autorizadas.aspx
- [7]** DIAN. "Acuerdos de Reconocimiento Mutuo del programa OEA". Colombia tiene ARM con Alianza del Pacífico, Comunidad Andina, Mercosur, EE.UU., Costa Rica, Guatemala y República Dominicana. Disponible en: <https://www.dian.gov.co/aduanas/oea/>
- [8]** ITIL (Information Technology Infrastructure Library). Marco de mejores prácticas para gestión de servicios de TI, incluyendo gestión de niveles de servicio y mesa de ayuda.
- [9]** Freshworks. "Unveiling 20 Key Chatbot Statistics for 2024". Los chatbots pueden reducir en 70% las consultas por llamada, chat y correo. Disponible en: <https://www.freshworks.com/chatbots/statistics/>
- [10]** AIPRM. "50+ AI in Customer Service Statistics 2024". H&M reportó reducción del 70% en tiempos de respuesta con chatbot de IA generativa. Disponible en: <https://www.aiprm.com/ai-in-customer-service-statistics/>
- [12]** MinTIC. "Estrategia Nacional Digital de Colombia 2023-2026". Incluye ejes sobre IA y tecnologías emergentes para generación de valor económico y social. Disponible en: https://www.mintic.gov.co/portal/715/articles-334120_recurso_1.pdf
- [13]** MinTIC. "Durante 2024, la estrategia Potencia Digital invirtió cerca de \$42.500 millones para acompañar a mil empresas", diciembre 2024. Disponible en: <https://www.mintic.gov.co/>

- [14]** Gartner. "Gartner Predicts Chatbots Will Become a Primary Customer Service Channel Within Five Years", julio 2022. Para 2027, chatbots serán canal principal para ~25% de organizaciones. Disponible en: <https://www.gartner.com/en/newsroom/press-releases/2022-07-27-gartner-predicts-chatbots-will-become-a-primary-customer-service-channel-within-five-years>
- [15]** Gartner. "Gartner Predicts Agentic AI Will Autonomously Resolve 80% of Common Customer Service Issues Without Human Intervention by 2029", marzo 2025. Reducción 30% en costos operativos. Disponible en: <https://www.gartner.com/en/newsroom/press-releases/2025-03-05>
- [16]** OMC. "World Trade Report 2024: Trade and inclusiveness - How to make trade work for all". Reducción de costos comerciales mediante digitalización. Disponible en: https://www.wto.org/english/res_e/publications_e/wtr24_e.htm
- [17]** OMC. "Members hold focused discussions on harnessing digitalization to facilitate trade", abril 2024. Tema 2024: Uso de Digitalización para Facilitar el Comercio. Disponible en: https://www.wto.org/english/news_e/news24_e/fac_17apr24_e.htm
- [18]** MinTIC. "Centros de Transformación Digital Empresarial". Estrategia de MinTIC e iNNpulsa Colombia para acompañar empresarios en transformación digital. Disponible en: <https://www.mintic.gov.co/portal/715/w3-article-198857.html>
- [19]** ProColombia. Datos sobre estructura de empresas exportadoras colombianas. PYMES representan ~75% de empresas exportadoras pero ~17% del valor exportado. Disponible en: <https://www.procolombia.co/>
- [20]** Decreto 1074 de 2015. Clasificación empresarial en Colombia según ingresos anuales por actividades ordinarias.
- [21]** GM Insights - Los 7 principales transitarios (DHL, Kuehne+Nagel, DSV, DB Schenker, Bolloré, Nippon Express, CEVA) tienen una cuota de mercado superior al 30%. URL: <https://www.gminsights.com/industry-analysis/freight-forwarding-market>
- [22]** Treble.ai / El Tiempo (2024) - 79% de consumidores colombianos prefieren WhatsApp para interactuar con marcas; 70% de empresas lo reportan como canal con mejor desempeño. URL: <https://www.eltiempo.com/tecnosfera/novedades-tecnologia/los-consumidores-colombianos-prefieren-whatsapp-a-la-hora-de-comprar-productos-3326914>
- [23]** Kayako / Help Scout - 41% de consumidores prefieren chat en vivo como canal de soporte (vs 32% teléfono, 23% email). Comm100 - 73% de clientes se sienten más cómodos comunicándose vía chat en vivo que por email o teléfono. URL: <https://www.helpscout.com/blog/live-chat-statistics/>

- [24]** Freshworks. "Unveiling 20 Key Chatbot Statistics for 2024". Tiempo respuesta chatbot: 9.3 segundos vs 39 segundos humanos. Disponible en: <https://www.freshworks.com/chatbots/statistics/>
- [25]** ANALDEX/DANE (2025). Informe Mensual de Exportaciones Colombianas - Diciembre 2024. Exportaciones 2024: USD 49.552 millones. Disponible en: <https://analdex.org/2025/02/11/informe-mensual-de-exportaciones-colombianas-diciembre-de-2024/>
- [26]** ANALDEX/DANE (2025). Informe de Importaciones 2024. Importaciones: USD 64.104,7 millones CIF. Disponible en: <https://analdex.org/2025/02/14/en-2024-las-importaciones-crecieron/>
- [27]** GM Insights (2025). Freight Forwarding Market Size, Industry Analysis 2025-2034. Top 7 forwarders con cuota de mercado >30%. Disponible en: <https://www.gminsights.com/industry-analysis/freight-forwarding-market>
- [28]** DHL Global Forwarding (2024). myDHLi: A New Generative AI Assistant to Boost Your Logistics Efficiency. Virtual assistant 24/7. Disponible en: <https://www.dhl.com/us-en/home/global-forwarding/latest-news-and-webinars/mydhli-generative-ai-logistics.html>
- [29]** B2Chat / ACIS (2023). 90% de las conversaciones digitales de las empresas colombianas con sus clientes se hacen por WhatsApp. Estudio de 600 empresas. Disponible en: <https://acis.org.co/portal/content/90-de-las-conversaciones-digitales-de-las-empresas-colombianas-con-sus-clientes-se-hacen-por>
- [30]** Hugging Face (2024). Llama 3.1 - 405B, 70B & 8B with multilinguality and long context. Especificaciones técnicas, requisitos de hardware y capacidades multilingües. Disponible en: <https://huggingface.co/blog/llama31>
- [31]** n8n.io (2025). AI Workflow Automation Platform. 400+ integraciones, nodos nativos de IA, despliegue autoalojado. Disponible en: <https://n8n.io/> y <https://github.com/n8n-io/n8n>
- [32]** OpenAI (2024). API Pricing. Precios actualizados GPT-4o: \$3-5/1M tokens entrada. Disponible en: <https://openai.com/api/pricing/>
- [33]** Anthropic (2025). Claude API Pricing. Claude Sonnet 4: \$3/1M entrada, \$15/1M salida. Disponible en: <https://docs.anthropic.com/en/docs/about-claude/pricing>
- [34]** Osterwalder, A. & Pigneur, Y. (2010). Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers. Hoboken, NJ: John Wiley & Sons.
- [35]** Wikipedia (2025). Business Model Canvas. Los nueve bloques del modelo fueron propuestos inicialmente en 2005 por Alexander Osterwalder, basándose en su investigación doctoral. Disponible en: https://en.wikipedia.org/wiki/Business_model_canvas

- [36]** BCG on Tech (2024). The Evolution of Pricing Models for SaaS Companies. El interés por modelos basados en uso crece en todos los tipos de software. Disponible en: <https://medium.com/bcgontech/the-evolution-of-pricing-models-for-saas-companies>
- [37]** Zylo (2025). Pricing Models Explained: Usage-Based vs Subscription Pricing. Más del 60% de empresas SaaS están adoptando modelos basados en uso o híbridos. Disponible en: <https://zylo.com/blog/usage-based-pricing-vs-subscription/>
- [38]** Chargebee (2024). SaaS Pricing Models Guide. El modelo basado en uso es percibido como el más justo por los clientes. Disponible en: <https://www.chargebee.com/resources/guides/saas-pricing-models-guide/>
- [39]** Andreessen Horowitz (2024). Usage-Based Pricing Is Popular, But Is It Right For You? Los modelos híbridos funcionan especialmente bien para productos SaaS con componentes de IA. Disponible en: <https://a16z.com/usage-based-pricing-rule-of-thumb/>
- [40]** OpenAI (2024). Whisper: Robust Speech Recognition via Large-Scale Weak Supervision. Modelo de transcripción de audio con soporte multilingüe y precisión superior al 92%. Disponible en: <https://openai.com/research/whisper>
- [41]** Qdrant (2024). Vector Database for AI Applications. Base de datos vectorial de código abierto optimizada para búsqueda semántica. Disponible en: <https://qdrant.tech/>
- [42]** PostgreSQL Global Development Group (2024). PostgreSQL: The World's Most Advanced Open Source Relational Database. Sistema de gestión de bases de datos relacional con soporte ACID. Disponible en: <https://www.postgresql.org/>
- [43]** Kane, G.C. (2017). Digital Maturity, Not Digital Transformation. MIT Sloan Management Review. La mejor comprensión de la transformación digital es adoptar procesos y prácticas para competir efectivamente en un mundo digital. Disponible en: <https://sloanreview.mit.edu/article/digital-maturity-not-digital-transformation/>
- [44]** Westerman, G., Bonnet, D. & McAfee, A. (2012). The Advantages of Digital Maturity. MIT Sloan Management Review. Las empresas digitalmente maduras superan a competidores en 6-26% en métricas financieras. Disponible en: <https://sloanreview.mit.edu/article/the-advantages-of-digital-maturity/>
- [45]** MIT Sloan Management Review & Deloitte (2017). Achieving Digital Maturity. 70% de empresas maduras organizadas en equipos interfuncionales vs 28% de empresas en etapas tempranas. Disponible en: <https://sloanreview.mit.edu/projects/achieving-digital-maturity/>
- [46]** Westerman, G., Bonnet, D. & McAfee, A. (2014). The Nine Elements of Digital Transformation. MIT Sloan Management Review. Tres áreas clave: experiencia del cliente, procesos operativos y modelos de negocio. Disponible en: <https://sloanreview.mit.edu/article/the-nine-elements-of-digital-transformation/>

[47] Congreso de la República de Colombia (2012). Ley Estatutaria 1581 de 2012 - Régimen General de Protección de Datos Personales. Disponible en: <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=49981>

[48] Superintendencia de Industria y Comercio (2024). Circular Externa 002 de 2024 - Lineamientos para el tratamiento de datos personales en sistemas de inteligencia artificial. Criterios: idoneidad, necesidad, razonabilidad y proporcionalidad. Referenciado en: <https://www.grantthornton.com.co/> y <https://www.cuatrecasas.com/>

[49] Congreso de la República de Colombia (1999). Ley 527 de 1999 - Comercio Electrónico. Regula la validez de contratos electrónicos y equivalencia funcional.

LISTA DE TABLAS

<i>Tabla 1 Modelos de Lenguaje de Gran Escala</i>	39
<i>Tabla 2 Infraestructura de Procesamiento:</i>	40
Tabla 3 Gestión de Conocimiento	41
<i>Tabla 4 Plataforma de Orquestación de Flujos de Trabajo</i>	42
Tabla 5 Canal de Comunicación con Usuarios	43
<i>Tabla 6 Métricas de Éxito</i>	49
<i>Tabla 7 Capa de Infraestructura Local</i>	53
<i>Tabla 8 Evaluación del Nivel de Madurez Digital</i>	65
<i>Tabla 9 Propiedad Intelectual y Licenciamiento</i>	69

LISTA DE FIGURAS

<i>Figura 1</i> <i>Árbol de Problemas</i>	16
<i>Figura 2</i> <i>Esquema Flujo de análisis</i>	19
<i>Figura 3</i> <i>Árbol de Objetivos</i>	29
<i>Figura 4</i> <i>Situación Actual - Proceso Manual de Asesoría</i>	31
<i>Figura 5</i> <i>Situación Esperada - Ecosistema Inteligente de Asesoría</i>	33
<i>Figura 6</i> <i>Mapa de Valor</i>	38
<i>Figura 7</i> <i>Modelo CANVAS de Negocio</i>	46
<i>Figura 8</i> <i>Diagrama de Arquitectura Solución Funcional</i>	56
<i>Figura 9</i> <i>Workflow n8n MVP Solución Funcional</i>	57
<i>Figura 10</i> <i>Creación y parametrización Base de Datos Postgres</i>	57
<i>Figura 11</i> <i>Resumen de Tecnologías Integradas</i>	62
<i>Figura 12</i> <i>Modelo de Análisis Madurez Digital de MIT Sloan</i>	66