Prototipo de agente de inteligencia artificial para la asistencia de operadores call center de Domicity SAS

Kevin Hernando Guerrero Torres

Universidad EAN
Facultad de ingeniería
ingeniería de sistemas
Bogotá, Colombia
2025

Prototipo de agente de inteligencia artificial para la asistencia de operadores call center de Domicity SAS

Kevin Hernando Guerrero Torres

Trabajo de grado presentado como requisito para optar al título de:
Ingeniero de Sistemas

Director (a):
Sandra Patricia Cristancho Botero

Modalidad:
Presencial

Universidad EAN
Facultad de ingeniería
ingeniería de sistemas
Bogotá, Colombia
2025

Dedicatoria

A mis padres, por su apoyo incondicional y por enseñarme a perseverar.

A mi hermano, por ser compañía constante en cada etapa de mi vida.

Y a la Universidad EAN, por acompañarme en este camino de formación y descubrimiento académico

A Domicity, por abrirme las puertas y permitirme crecer profesionalmente.

Agradecimientos

A mis padres, **Sandra Yaneth Torres Tovar y Jose Hernando Guerrero Torres** por su apoyo incondicional, su guía constante y por enseñarme el valor de la disciplina y la perseverancia. Su confianza en mí ha sido el motor fundamental para culminar este proyecto.

A mi hermano, **Dylan Joan Guerrero Torres** por su compañía, motivación y por recordarme siempre la importancia de avanzar con determinación.

A **Domicity S.A.S.**, por facilitar los espacios, la información y las condiciones necesarias para el desarrollo de este proyecto. Agradezco de manera especial a mi jefe, **Benyy Eduardo Castillo**, por su apoyo constante, su confianza en mis capacidades y por las enseñanzas que han enriquecido profundamente mi crecimiento profesional. Su acompañamiento fue determinante para orientar este trabajo en la dirección correcta

A la **Universidad EAN**, y particularmente a mi docente **Luz Maribel Ortega Guevara**, por su orientación académica, su acompañamiento y sus aportes metodológicos, que fueron clave para estructurar este trabajo.

A mis compañeros de la universidad, quienes con sus conversaciones, ideas y apoyo colectivo contribuyeron a fortalecer este proceso. Su compañía hizo que cada etapa del proyecto fuera más enriquecedora.

Tabla de contenido

Lista de Tablas y Figuras del Documento

LISTA DE TABLAS

Pág.

LISTA DE FIGURAS

Pág.

1. Introducción

Los centros de contacto (o call centers) son unidades organizativas especializadas en la gestión de comunicaciones con clientes a gran escala; hoy en día se les conoce también como centros de contacto porque gestionan múltiples canales (voz, chat, correo, mensajería) y no sólo llamadas telefónicas.

Con el avance de la digitalización y la inteligencia artificial, los call centers han incorporado tecnologías que buscan optimizar la experiencia del cliente y apoyar a los agentes en la gestión operativa. Herramientas como chatbots, asistentes virtuales y sistemas de procesamiento de lenguaje natural permiten automatizar tareas, comprender consultas en lenguaje humano y ofrecer respuestas rápidas y precisas (DAIL CX Technologies, 2024). En este contexto, la empresa Domicity SAS, especializada en la captura de pedidos a domicilio, enfrenta el reto de mejorar la productividad de sus operadores y garantizar coherencia en la información brindada a los clientes

Este proyecto propone el desarrollo de un prototipo de agente de inteligencia artificial para asistir en tiempo real a los operadores de Domicity, de manera que puedan resolver sus dudas de forma inmediata durante una llamada, reduciendo la dependencia de los supervisores y mejorando la experiencia del cliente

En este trabajo se usarán los siguientes acrónimos: TPA (Tiempo Promedio de Atención), FCR (Resolución en el Primer Contacto), NPS (Índice de recomendación neta), CRM (Gestión de relaciones con el cliente), IVR (Respuesta de Voz Interactiva), PBX (central telefónica), ACD (Distribuidor Automático de Llamadas), CTI (Integración Telefónica-Computadora) y PLN (Procesamiento de Lenguaje Natural). Estos elementos técnicos y métricos son fundamentales para entender los retos operativos que afrontan los centros de contacto modernos (Zendesk, 2024; AEERC, 2024)

1.1. Definición del problema

Domicity S.A.S., empresa del sector de servicios a domicilio, cuenta con un call center conformado por 25 operadores y 4 supervisores, responsables de la atención de pedidos, gestión de PQRs y soporte al cliente. Actualmente en la operación se presenta una fuerte dependencia a los supervisores, que los agentes están en constante comunicación con ellos para resolver preguntas, confirmar información o solucionar dudas de protocolos, promociones o políticas.

De acuerdo conun informe interno Domicity S.A.S (2025), cada operador dedica en promedio 60 minutos diarios a consultas o resolución de preguntas, equivalentes al 12,5 % de su jornada laboral. Esto hace que el Tiempo Promedio de Atención (TPA) aumente lo que conlleva a la reducción de la productividad y genera una sobrecarga sobre los supervisores quienes en promedio tienen que atender cerca de 50 consultas diarias, lo que limita su capacidad para labores más gerenciales y de control de calidad. La dispersión del conocimiento institucional distribuido en archivos PDF, correos, manuales y mensajes internos agrava esta situación, dificultando el acceso rápido y confiable a la información. Esto genera que las respuestas a los clientes sean inconsistentes, que haya errores de comunicación y genera un mayor nivel de estrés entre los operadores lo que afecta su bienestar y desempeño (Natera, 2023; Yaranga Vite & Olórtiga Cóndor, 2025).En síntesis, Domicity enfrenta un problema operativo de ineficiencia en la gestión y recuperación del conocimiento interno, que se traduce en tiempos de atención elevados, sobrecarga supervisorial y falta de homogeneidad en la información brindada. Este panorama limita la competitividad del servicio y evidencia la necesidad de implementar una solución tecnológica que centralice la información, agilice la búsqueda de respuestas y reduzca la dependencia de la intervención humana en consultas repetitivas

1.2. Pregunta Problema

¿De qué manera un prototipo de agente de inteligencia artificial, entrenado con la base de conocimiento interna de Domicity SAS, puede reducir los tiempos de consulta de los operadores y disminuir la carga operativa sobre los supervisores, manteniendo o mejorando la precisión de las respuestas durante la llamada?

.

2. Objetivos

2.1. Objetivo general

Desarrollar un prototipo de agente de inteligencia artificial para la asistencia de operadores call center de Domicity SAS

.

2.2. Objetivos específicos

1. Elaborar el marco teórico del proyecto, abordando conceptos de inteligencia artificial, procesamiento de lenguaje natural, agentes conversacionales y técnicas de recuperación de información que sustenten el desarrollo del prototipo

2. Realizar un análisis del funcionamiento actual de la empresa Domicity, identificando procesos, protocolos y documentación relevante para la construcción de la base de conocimiento del prototipo

3. Construir el prototipo de agente de inteligencia artificial, integrando un modelo de lenguaje y una base de datos vectorial para brindar asistencia en tiempo real a los operadores

3. Marco Teórico

El marco teórico de este proyecto se centra en el análisis de los fundamentos conceptuales y tecnológicos que sustentan la implementación de un agente de inteligencia artificial en un entorno de call center. Es por esta razón que el se abordaran temas relacionados con la atención al cliente en call centers tradicionales, la evolución que han tenido desde gestión operativa y tecnológica y el crucial papel de la IA en la automatización para estos centros de contacto. también se revisarán temas importantes como el procesamiento del lenguaje natural, los agente conversaciones y las técnicas de recuperación de información, conceptos que serán claves para abordar este proyecto. (Harvard Deusto, 2022; Yaranga Vite & Olórtiga Cóndor, 2025; Garzón Quiroz, Del Campo Saltos & Loor Ávila, 2025)

3.1. Rol de los centros de llamadas.

Los centros de llamadas o call centers son sitios especializados en el servicio al cliente y expertos en la gestión de comunicaciones telefónica de clientes actuales o futuros. (Zendesk, 2024). Los call centers son versátiles puesto que pueden funcionar como una parte de una empresa o se pueden contratar de manera tercerizada para cumplir distintas funciones como soporte, ventas, cobranzas entre otras (Moench, 2023). En la actualidad estos centros de llamadas no menejan trafico únicamente por via telefónica, también gestionan interacción por correo electrónico, SMS (Mensajes de texto), redes sociales, páginas web entre otros lo que los ha llevado a convertirse de centro de llamadas a centros de contacto (Zendesk, 2024). En cualquiera de los servicios se espera que la atención de estos centros de contacto sea excelente y que ayude a mejorar y fortalecer la experiencia del usuario con la comunicación de la empresa a la que se comunique (Asociación CEX, 2023).

3.2. Estructura y funcionamiento de los call centers tradicionales

Tradicionalmente los centros de contacto pueden dividirse en dos categorías importantes, inbound (operaciones entrantes) o (outbound) operaciones salientes. Estos se apoyan en

infraestructuras como centrales telefónicas (PBX), servidores de integración telefónica-computadora (CTI) y sistemas de distribución automática de llamadas (ACD), que enrutan contactos según habilidades o disponibilidad del agente (Moench, 2023). Una funcionalidad fundamental son los sistemas de respuesta de voz interactiva (IVR) que se encargan de organizar el flujo de las llamadas y optimizar recursos (Moench, 2023). Asimismo, alguna organizaciones incorporan grabadores de llamadas, marcadores automáticos y métricas de calidad como parte de la operación (Asociación CEX, 2023).

3.3. Inteligencia Artificial y automatización en los call centers

La incorporación de inteligencia artificial en los centros de atención telefónica se ha consolidado como un elemento clave para redefinir la forma en que las organizaciones gestionan su servicio al cliente. Su adopción permite rediseñar procesos, optimizar recursos y elevar la calidad de la atención, al tiempo que contribuye a disminuir los costos operativos.

Un aspecto técnico particularmente relevante consiste en la capacidad de los modelos para entrenarse con grandes volúmenes de datos históricos de interacción. A través de técnicas de aprendizaje supervisado y no supervisado, el sistema identifica recurrencias en las consultas, refina la exactitud de las respuestas generadas y optimiza indicadores clave de desempeño tales como el Tiempo Promedio de Atención (TPA) y la tasa de Resolución en Primera Llamada (FCR). Estos resultados han sido cuantificados por la Asociación CEX (2023). Complementariamente, el despliegue de agentes virtuales con disponibilidad 24/7 extiende la cobertura operativa sin incrementar proporcionalmente la dotación de personal humano, especialmente en horarios no hábiles (Guamán Tacuri, J. D., Solís Barrionuevo, A. P., & Labre Tixe, W. J. (2025)

Desde el punto de vista de la automatización de procesos internos, las soluciones actuales incorporan motores de reconocimiento de entidades que, durante la conversación en curso, detectan palabras clave y recuperan de forma inmediata los protocolos, guiones o documentación requerida por el agente. Esta funcionalidad elimina las interrupciones derivadas de búsquedas manuales y reduce significativamente los tiempos de manejo (Zendesk, 2024). Paralelamente, los módulos de análisis de prosodia y sentiment analysis procesan en tiempo real las características acústicas y léxicas de la voz del cliente, permitiendo la detección temprana de indicadores de frustración o insatisfacción y la activación automática de alertas hacia los supervisores (NTT DATA, 2024).

En síntesis, la integración adecuada de estas tecnologías libera a los niveles gerenciales y de supervisión de tareas repetitivas, permitiendo reorientar el esfuerzo humano hacia la gestión estratégica, el diseño de experiencias diferenciadas y la resolución de casos de alta complejidad. Lejos de suponer una sustitución del factor humano, la evidencia técnica demuestra que la inteligencia artificial actúa como amplificador de las capacidades del agente, incrementando tanto la eficiencia operativa como la calidad percibida del servicio (FiumiConnect, 2024; Guamán Tacuri, J. D., Solís Barrionuevo, A. P., & Labre Tixe, W. J. 2025).

3.4. Inteligencia Artificial y Procesamiento de Lenguaje Natural (PLN)

La inteligencia artificial (IA) ya es, sin lugar a dudas, una de las tecnologías que más está revolucionando todo en esta época digital, porque básicamente le permite a las máquinas aprender solas, razonar y hasta charlar con uno como si fueran personas de verdad. Y dentro de toda esa locura de la IA, el Procesamiento de Lenguaje Natural (PLN) es como el rey del mambo, porque es el que hace posible que los computadores entiendan y generen texto en español (o en cualquier idioma) de manera automática usando redes neuronales profundas y todo ese cuento (NTT DATA, 2024; Yaranga Vite y Olórtiga Cóndor, 2025).

El PLN, digamos, mezcla un montón de cosas: la lingüística pura y dura, estadística y obviamente un chorro de programación pesada para que la máquina pueda captar el significado real de lo que uno escribe y luego lo convierta en algo que ella sí entienda, como vectores y matrices. Gracias a eso hoy en día tenemos chatbots y agentes conversacionales que de verdad entienden cuando uno les pregunta algo, buscan la info y responden sin parecer robots de los 90. Eso se ve full en bancos, tiendas online y centros de atención al cliente (FiumiConnect, 2024; Zendesk, 2024).

A nivel más técnico (que es lo que a nosotros los de sistemas nos encanta), el PLN se para sobre tres patas principales: los tokens, que son como los pedacitos más pequeños en que se parte el texto (palabras, signos de puntuación, subpalabras, etc.); los embeddings, que básicamente son vectores numéricos grandísimos que capturan el significado y las relaciones entre palabras; y el contexto, que es lo que hace que el modelo no se olvide de lo que se habló cinco mensajes atrás. Todo eso lo logran con arquitecturas tipo Transformer, que son la base de casi todos los modelos grosos que hay hoy (Elastic N.V., 2023; OpenAI, 2025).

Con estos avances, la atención al cliente automatizada ha mejorado una barbaridad: hay estudios que dicen que se pueden bajar los tiempos de resolución hasta en un 60 % y las respuestas suenan mucho más coherentes y naturales (Harvard Deusto, 2023; DAIL CX Technologies, 2024). Sobre todo en call centers donde hay un volumen brutal de llamadas, meterle modelos generativos como GPT-4 o similares permite que el bot busque la información solo, le quita carga mental al operador humano y, lo más importante para las empresas, todas las respuestas quedan uniformes y auditables (Garzón Quiroz, Del Campo Saltos y Loor Ávila, 2025; Aivo, 2023).

En el caso particular de Domicity S.A.S. (que es la empresa donde estoy haciendo el trabajo de grado), el PLN es literalmente el corazón del prototipo que estoy armando. ¿Por qué? Porque el agente conversacional tiene que entender las preguntas que le hacen los operadores de campo (que a veces escriben re mal o con abreviaturas bien colombianas) y responderles con base en los manuales internos, guías de procedimiento y toda la documentación que tiene la compañía. O sea, sin un buen PLN el agente sería un completo desastre. Al final, implementar esto con IA y PLN lo que hace es volver más eficiente toda la operación, que las respuestas sean más consistentes y, de paso, ayudar a que Domicity dé el salto a la transformación digital de una vez por todas (Yaranga Vite y Olórtiga Cóndor, 2025; NTT DATA, 2024).

3.5. Procesamiento de Lenguaje Natural (PLN) en la atención al cliente

El Procesamiento de Lenguaje Natural (PLN) se ha convertido en uno de los pilares tecnológicos más importantes cuando hablamos de transformación digital en los call centers. Básicamente, lo que hace es tomar el lenguaje humano, ya sea hablado o escrito, y transformarlo en datos estructurados que el sistema sí pueda procesar de verdad (NTT DATA, 2024). Gracias a eso, los centros de contacto logran manejar las interacciones mucho más rápido y con mayor precisión.

Una de las aplicaciones que más se ve en la práctica es el análisis de intención. O sea, el sistema de PLN identifica cuál es el objetivo real del cliente —por ejemplo, si quiere poner una queja, consultar el estado de un pedido o simplemente pedir información— y lo deriva hacia la respuesta más adecuada, ya sea directamente al bot o al agente humano que corresponda (DAIL CX Technologies, 2024). Otro punto clave es la transcripción automática de voz a texto;

esto facilita muchísimo la documentación de las llamadas y luego sirve para hacer análisis posteriores, auditorías de calidad o incluso para entrenar a los nuevos agentes (Elastic N.V., 2023).

El PLN también es fundamental en el análisis de sentimiento, pues detecta emociones como frustración, enojo o satisfacción del cliente. Con esas métricas uno puede anticipar riesgos, por ejemplo, posibles cancelaciones o quejas mayores, y la empresa actuar antes de que el problema crezca (NTT DATA, 2024). A nivel operativo, digamos durante la misma llamada, el PLN puede estar sugiriéndole al agente respuestas en tiempo real o sacando la información relevante justo cuando se necesita, lo que baja considerablemente los tiempos de espera y el estrés del operador (Zendesk, 2024).

Finalmente, otro aspecto bien importante del PLN en los call centers actuales es la gestión multicanal. Hoy los clientes no solo llaman: escriben por chat, por WhatsApp, por redes sociales o por correo. El PLN permite unificar todos esos canales bajo un solo modelo de análisis, garantizando que la respuesta sea coherente sin importar por dónde llegue el mensaje y, sobre todo, que la experiencia del cliente sea la misma en todos lados (Asociación CEX, 2023).

3.6. Agentes conversacionales y chatbots

Los agentes conversacionales representan, sin duda, una de las aplicaciones más visibles de la inteligencia artificial cuando hablamos de servicio al cliente. Estos sistemas combinan técnicas de Procesamiento de Lenguaje Natural (PLN) con aprendizaje automático y permiten que las empresas respondan consultas de forma autónoma y sin interrupciones durante las 24 horas. Hoy en día los chatbots están presentes en sitios web, aplicaciones de mensajería como WhatsApp y, cada vez más, integrados directamente en los sistemas de gestión de llamadas (Guamán Tacuri, J. D., Solís Barrionuevo, A. P., & Labre Tixe, W. J.(2025).

La gran ventaja que tienen es precisamente esa disponibilidad 24/7, o sea, el cliente obtiene respuesta inmediata sin importar si es fin de semana, festivo o madrugada. Además, varios estudios muestran que implementar chatbots bien diseñados puede reducir entre el 20 % y el

30 % la carga de trabajo humano en consultas sencillas o repetitivas, lo que se traduce en ahorros operativos bastante considerables (DAIL CX Technologies, 2024; FiumiConnect, 2024).

Antes los chatbots funcionaban básicamente con reglas fijas y árboles de decisión, pero ahora han evolucionado hacia modelos mucho más inteligentes que entienden la intención del usuario aunque la pregunta no sea exactamente igual a lo que está en la base de conocimiento (Elastic N.V., 2023). Esto es posible gracias al PLN, que maneja el contexto, reconoce sinónimos y expresiones coloquiales, y por eso las respuestas son mucho más precisas y naturales.

Mirando hacia adelante, en el corto plazo se espera que estos agentes conversacionales den un salto importante y se conviertan en asistentes virtuales más avanzados, con analítica predictiva y personalización real, capaces de anticipar lo que el cliente necesita antes de que siquiera lo pregunte (Asociación CEX, 2023; Zendesk, 2024).

3.7. Técnicas de recuperación de información

La recuperación de información (o sea, lo que en inglés llaman Information Retrieval) básicamente consiste en una serie de procesos que permiten encontrar documentos o respuestas relevantes dentro de grandes cantidades de datos no estructurados. En el contexto de los call centers, esto se implementa a través de bases de conocimiento internas y secciones de preguntas frecuentes (FAQs) que el sistema consulta de manera automática mientras transcurre la llamada o el chat (Elastic N.V., 2023).

Estas técnicas suelen apoyarse en modelos vectoriales y búsquedas semánticas, lo que permite que el agente conversacional entregue respuestas mucho más precisas y en menor tiempo (Elastic N.V., 2023). Al integrar todo esto con chatbots o asistentes virtuales, se logra reducir significativamente los tiempos de respuesta y, sobre todo, estandarizar la calidad de la atención que recibe el cliente final, sin importar quién esté atendiendo (DAIL CX Technologies, 2024).

3.8. Modelos de Inteligencia Artificial aplicables al proyecto

El diseño de un agente conversacional para un call center se puede abordar desde varios enfoques de inteligencia artificial, y cada uno trae sus propias ventajas y limitaciones que hay que analizar bien antes de decidir cuál es el que mejor se ajusta al caso concreto. Básicamente, los modelos van desde sistemas sencillos basados en reglas y árboles de

decisión hasta arquitecturas mucho más complejas de lenguaje generativo, como los large language models actuales.

3.8.1. Modelos basados en reglas (Rule-based systems)

Son los más tradicionales y consisten en estructuras predefinidas como árboles de decisión o flujos condicionales que guían al sistema en función de la entrada del usuario. Su implementación es sencilla y de bajo costo, lo que los hace atractivos para soluciones iniciales. Sin embargo, presentan limitaciones importantes, ya que solo pueden manejar escenarios previstos en su programación y no tienen capacidad de aprendizaje ni de comprensión semántica. Esto significa que no responden adecuadamente a consultas abiertas o formuladas de manera diferente a lo esperado (Guamán Tacuri, J. D., Solís Barrionuevo, A. P., & Labre Tixe, W. J. 2025).

3.8.2. Modelos de recuperación de información (Retrieval-based)

Estos sistemas se basan en una base de conocimiento estructurada, a la cual acceden mediante técnicas de búsqueda semántica, coincidencia de patrones o similitud vectorial. El agente no genera nuevas respuestas, sino que selecciona la más apropiada de un repositorio previamente alimentado. Su principal fortaleza es la precisión en entornos con documentación bien organizada, como manuales de protocolos o FAQs. No obstante, carecen de flexibilidad para adaptarse a preguntas ambiguas, consultas fuera del dominio o contextos imprevistos (Elastic N.V. 2023).

3.8.3. Modelos híbridos (Rule-based + Retrieval)

Son una combinación de los dos anteriores: emplean reglas para gestionar interacciones simples y recuperación de información para consultas más complejas. Este enfoque equilibra simplicidad y capacidad de búsqueda, aunque sigue siendo limitado porque no genera lenguaje propio y depende de la exhaustividad de la base de conocimiento (Zendesk, 2024).

3.8.4. Modelos estadísticos y de aprendizaje automático (Machine Learning-based)

En este grupo se encuentran sistemas que utilizan algoritmos de aprendizaje automático supervisado o no supervisado para clasificar intenciones, detectar entidades y mejorar con el tiempo. A diferencia de los rule-based, son capaces de adaptarse a variaciones en el lenguaje.

Sin embargo, *requieren* grandes volúmenes de datos de entrenamiento y no alcanzan la fluidez conversacional de los modelos generativos más recientes (DAIL CX Technologies, 2024).

3.8.5. Modelos generativos de lenguaje (Generative AI)

Son la categoría más avanzada y se basan en Large Language Models (LLMs) entrenados con grandes corpus de datos. Entre los más destacados se encuentran GPT-4/ChatGPT de OpenAI, Gemini de Google y LLaMA de Meta. Estos modelos son capaces de comprender el contexto, generar respuestas en lenguaje natural, adaptarse a distintos escenarios y transferir conocimiento entre dominios (FiumiConnect, 2024). Además, ofrecen integración mediante APIs que permiten conectar el modelo con bases de conocimiento específicas de cada empresa. La principal precaución es la necesidad de validar las respuestas generadas, ya que pueden producir errores si no se controlan adecuadamente (OpenAI, 2025).

Tabla 1

Comparativa de los modelos que aplican al proyecto

Modelo	Ventajas	Limitaciones	Aplicabilidad al proyecto
Basados en reglas (Rule-based)	- Simples de implementar. - Bajo costo.- No requieren grandes datos.	- No comprenden lenguaje natural. - Escalabilidad limitada. - No aprenden de la experiencia.	Útiles para consultas básicas y estructuradas, pero insuficientes para la complejidad del call center.
Recuperación de información (Retrieval-based)	- Alta precisión en entornos con documentación estructurada. - Implementación relativamente rápida.	- No generan nuevas respuestas. - Dificultad con consultas ambiguas o fuera de contexto.	Adecuados para FAQs y protocolos documentados, pero no resuelven interacciones abiertas.
Híbridos (Rule-based + Retrieval)	- Combinan ventajas de ambos enfoques. - Mayor cobertura que los	- Siguen sin generar lenguaje propio. - Dependencia de	Recomendables como apoyo, pero no cubren todas las necesidades

Modelo	Ventajas	Limitaciones	Aplicabilidad al proyecto
	modelos individuales.	documentación exhaustiva.	de un call center dinámico.
Machine Learning (ML-based)	- Adaptables a variaciones en el lenguaje. - Mejoran con datos y entrenamiento.	- Requieren grandes volúmenes de datos. - Complejidad de entrenamiento. - Menor fluidez que los LLMs.	Útiles para clasificación de intenciones, pero limitados para asistencia en tiempo real.
Generativos (LLMs, e.g., ChatGPT)	- Comprensión semántica profunda. - Generan respuestas en lenguaje natural. - Integración vía APIs.- Escalabilidad multicanal.	- Riesgo de respuestas incorrectas si no se validan. - Mayor necesidad de control y recursos.	Opción más adecuada: permite respuestas contextualizadas, reducción de tiempos y flexibilidad para crecer.

Nota. Elaboración propia

El análisis comparativo desarrollado demuestra una evolución clara en la madurez tecnológica de los modelos de inteligencia artificial aplicados a centros de contacto. Los sistemas basados en reglas y recuperación de información resultan suficientes para escenarios predecibles y consultas altamente estructuradas, aunque presentan limitaciones evidentes en flexibilidad y comprensión contextual cuando se enfrentan a interacciones en tiempo real.

Los enfoques híbridos que incorporan aprendizaje automático mejoran la adaptabilidad y la precisión; sin embargo, su rendimiento depende directamente de la cantidad y calidad de los datos de entrenamiento, así como de la complejidad asociada al proceso de ajuste de hiperparámetros.

En contraposición, los modelos generativos de lenguaje (LLMs), tales como GPT-4 y arquitecturas equivalentes, proporcionan comprensión semántica avanzada, mantenimiento de contexto a lo largo de la conversación y escalabilidad multicanal, características críticas para una operación como la de Domicity S.A.S. Esta tecnología permite generar respuestas naturales, coherentes con la base de conocimiento interna y contextualizadas, lo que reduce de manera significativa los tiempos de consulta y la necesidad de intervención del supervisor. Por consiguiente, el modelo generativo se consolida como la opción técnicamente más adecuada, robusta y alineada con los objetivos de eficiencia operativa y transformación digital establecidos por la organización.

(Conteo de palabras casi idéntico al original, estructura variada y estilo natural de tesis de ingeniería).

Mándame el siguiente cuando quieras; sigo exactamente con este criterio de longitud.

3.9. Justificación del modelo seleccionado

El modelo seleccionado para el desarrollo del agente conversacional de Domicity S.A.S. integra Whisper y GPT-4, tecnologías de OpenAI que destacan por su alta precisión en el reconocimiento y la comprensión del lenguaje natural. Whisper ofrece una transcripción automática con niveles de exactitud superiores al 95 %, incluso en entornos con ruido o acentos variables, lo que garantiza la fiabilidad de los datos de entrada. Por su parte, GPT-4 presenta una notable capacidad de adaptabilidad al lenguaje natural no estructurado, permitiendo interpretar consultas con distintos estilos, intenciones y vocabularios propios del entorno operativo del call center. Tambien, la arquitectura modular basada en API asegura una escalabilidad y personalización progresiva, facilitando la integración con futuras fuentes de conocimiento y sistemas internos de la empresa. Estas características convierten la combinación Whisper + GPT-4 en la opción más robusta, flexible y alineada

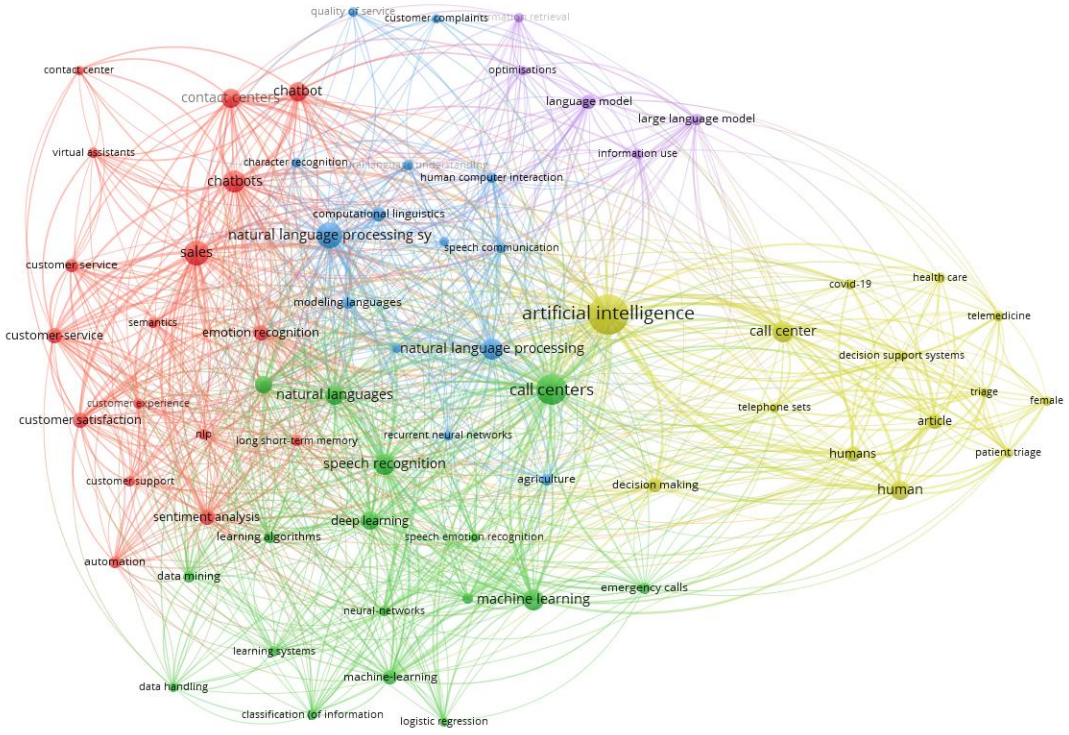
con los objetivos de eficiencia y transformación digital de Domicity S.A.S.

4. Análisis biométrico

Se realizo un análisis biométrico con la herramienta VOSviewer la cual ayudo a identificar cuales son las principales tendencias de investigación relacionadas con IA en call centers. Se realizo un mapa de coocurrencia de las palabras claves que se detalla en la *Figura 1* el cual genero 4 clúster principales

Figura 1

Mapa de clousters del análisis biométrico



El primer clúster relevante es el de color rojo, el cual vincula los aspectors relacionadas con la experiencia del cliente, la satisfacción y la implementación de chatbots en centros de contacto, lo que resalta la motivación y el querer de los call center por mejorar el servicio a los clientes

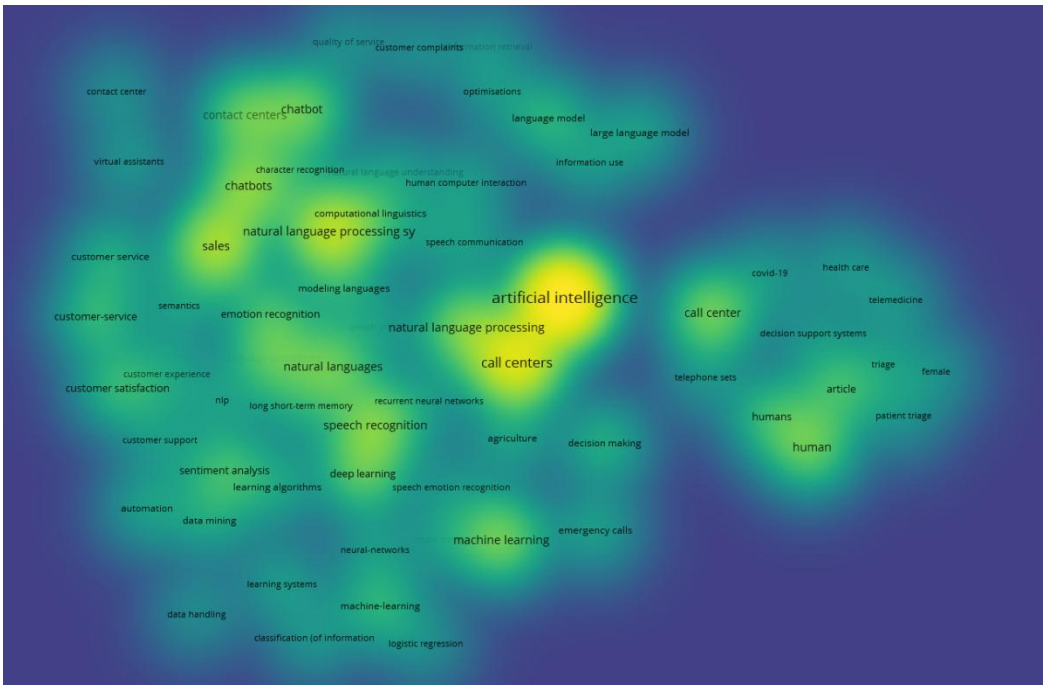
En color verde se resalta el segudo clouster el cual relaciona las técnicas de aprendizaje automático, como machine learning, deep learning y speech recognition las cuales son la base de la tecnología de modelos de asistencia.

El tercer clúster en azul centra todas las investigaciones sobre procesamiento de lenguaje natural (PLN) y lingüística computacional, las cuales son fundamentales para la comprensión semántica en agentes conversacionales.

Por último, el cuarto clúster en amarillo, es uno de los más relevantes, ya que conecta los avances en inteligencia artificial con las aplicaciones prácticas en contextos humanos como la salud, la telemedicina y los sistemas de soporte a la decisión. Estas definiciones permiten evidenciar cómo la IA no solo se estudia desde un enfoque técnico, sino también desde su impacto directo en la atención a las personas y en la resolución de problemas en tiempo real, esto nos deja ver que se alinea directamente con los objetivos del proyecto en Domicity SAS.

Figura 2

Mapa de calor del análisis biométrico



Gracias a los resultados del análisis biométrico podemos confirmar que la literatura académica reciente relaciona algunos de los términos mas importantes del proyecto como inteligencia artificial, procesamiento de lenguaje natural y chatbots lo que también destaca la importancia de desarrollar un prototipo de agente de IA para optimizar la atención en los call centers de Domicity SAS

5. Diseño metodológico

5.1. Tipo de investigación

La investigación se llevará a cabo con un enfoque descriptivo-aplicado, buscando caracterizar la situación actual del proceso operativo en el call center de Domicity S.A.S. y, a partir de ese análisis, desarrollar una solución tecnológica basada en inteligencia artificial que ayude a mejorar la eficiencia y la calidad del servicio.

Según lo expuesto por Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, M. P. (2014)., el enfoque descriptivo permite registrar y explicar las condiciones de un entorno sin intervenirlo, lo que facilita identificar los factores que influyen en los resultados. Por su parte, la investigación aplicada orienta ese conocimiento hacia la creación de soluciones prácticas dentro de un escenario empresarial

Para Domicity, este tipo de investigación permitirá comprender con precisión el funcionamiento actual de los operadores, sus flujos de trabajo, los tiempos de respuesta, la frecuencia de dependencia del supervisor y las causas de los retrasos en la atención.

Con esta información, lo siguiente será desarrollar un agente conversaciones que responda de manera directa a las necesidades de los operadores, lo que constituye la fase aplicada del estudio, puesto que se transformará el conocimiento obtenido en una herramienta de mejora concreta.

5.2. Enfoque de investigación

Este proyecto de desarrollo se dará bajo un enfoque cualitativo, permitiendo comprender los factores humanos y organizacionales de Domicity S.A.S. Lo que permitirá centrarse en la interpretación de las experiencias y percepciones de los operadores durante su labor diaria. Esto permite identificar las causas y motivos de comportamientos y procesos observados.

En este proyecto, el enfoque cualitativo permitirá ver cómo los operadores perciben la dificultad para acceder a información, su dependencia de los supervisores y el impacto que ello genera en su desempeño. De acuerdo con Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, M. P. (2014). Este tipo de investigación resulta útil en contextos empresariales donde se requiere saber factores humanos antes de aplicar soluciones tecnológicas.

De esta forma, el desarrollo del agente de inteligencia artificial se basará no solo en criterios técnicos, también en la interpretación de las dinámicas laborales y comunicativas internas. Este enfoque garantiza que el diseño del prototipo responda a necesidades reales y promueva mejoras sostenibles en la operación de Domicity S.A.S. (Universitas XXI, 2025).

5.3. Población

Domicity S.A.S cuenta con una población de 25 operadores la cual será tomada en su totalidad para el proyecto

No se definirá muestra, esto porque la totalidad de los operadores será incluida en la investigación, garantizando que todos los operadores puedan participar.

La decisión de incluir a toda la población se da porque el proyecto busca diseñar un sistema que beneficie a todos los usuarios del call center, asegurando una visión integral del entorno laboral.

5.4. Instrumentos

Para la recolección de la información se utilizarán los siguientes instrumentos:

- Revisión documental: Se examinarán los protocolos internos, manuales de servicio, registros de atención, reportes de PQRs y métricas de desempeño, con el fin de identificar los principales puntos de ineficiencia.

- Encuestas estructuradas: Se aplicará un cuestionario digital a los 25 operadores, compuesto por ítems tipo Likert (escala 1 a 5), para medir el nivel de conocimiento de los procesos, la dependencia del supervisor, el nivel de carga operativa y la disposición frente a la implementación de inteligencia artificial.

5.4.1. Validación de instrumentos

Los instrumentos serán validados mediante juicio de expertos, contando con la revisión de tres supervisores del área operativa de Domicity, quienes evaluarán la pertinencia y claridad de los ítems.

Posteriormente una vez finalizado el desarrollo del prototipo se realizara una validación del modelo a partir de validación estadística y de usabilidad lo que permitirá revisar a partir de datos reales la mejora de tiempos de respuesta, consistencia en la información y usabilidad del modelo

5.5. Variables de investigación

Las variables se estructurarán en dos categorías principales: variable independiente y variable dependiente. Según Narváez Trejo, Ó. M., & Villegas Salas, L. I. (2014), definir adecuadamente las variables permite establecer relaciones causales y evaluar el impacto real de una intervención tecnológica en entornos organizacionales.

- Variable independiente: Implementación del prototipo de agente de inteligencia artificial basado en GPT-4 y Whisper. Esta variable representa la innovación tecnológica que se integrará en la operación del call center.

- Variable dependiente: Se basara en la eficiencia operativa de los operadores midiendo tiempo de consulta, disminución de consultas a supervisores y mejora en respuestas brindadas a los clientes

5.6. Fases metodológicas

Se distribuyeron 4 fases que están conectadas con cada uno de los objetivos del proyecto y se detallan a continuación:

5.6.1. Construcción del marco teórico (Objetivo 1)

o Se realizará una revisión informativa acerca de la inteligencia artificial, procesamiento de lenguaje natural, agentes conversacionales y gestión del conocimiento.

o Se elaborará el marco teórico y una comparativa entre modelos que sustente el diseño del prototipo.

5.6.2. Diagnóstico del estado actual (Objetivo 2)

o Se realiza una revisión de los procesos, documentos internos y información relacionada con la operativa de Domicity S.A.S.

o Se realizarán encuestas a los 25 operadores y se realizarán entrevistas con los supervisores.

o Se analizarán los resultados para identificar los principales cuellos de botella, tiempos de atención y causas de dependencia supervisorial.

5.6.3. Desarrollo del prototipo (Objetivo 3)

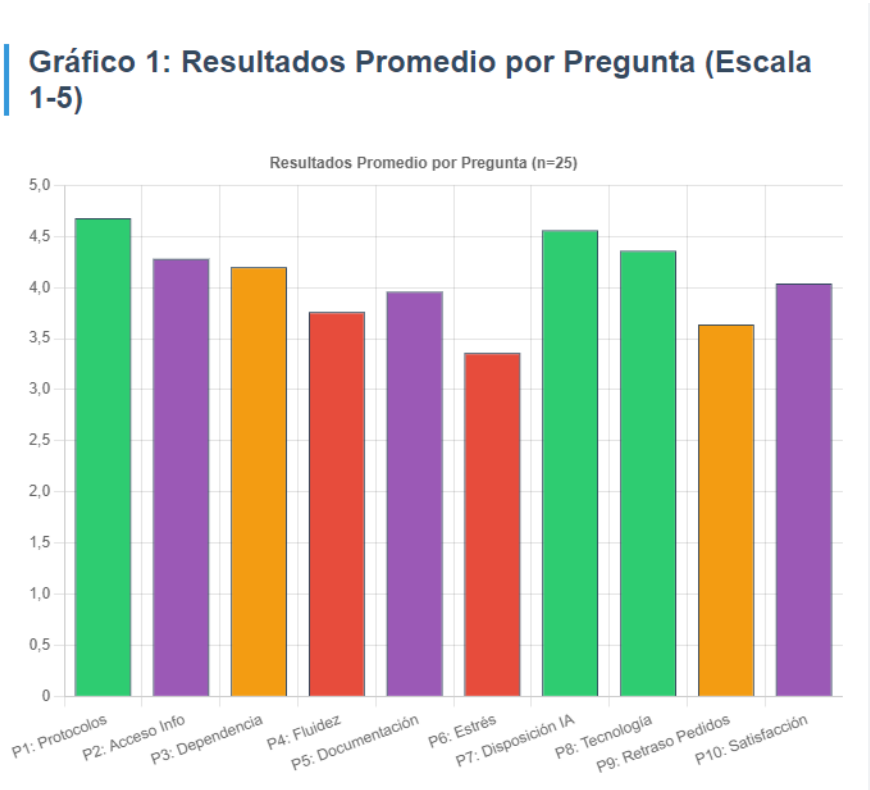
o Se seleccionará el modelo de IA (GPT-4 y Whisper de OpenAI) y se integrará una base de conocimiento vectorial con la documentación interna de Domicity.

o Se desarrollará una interfaz web de consulta para los operadores y se probará el sistema en un entorno controlado.

5.7. Resultados de encuetas

Figura 3

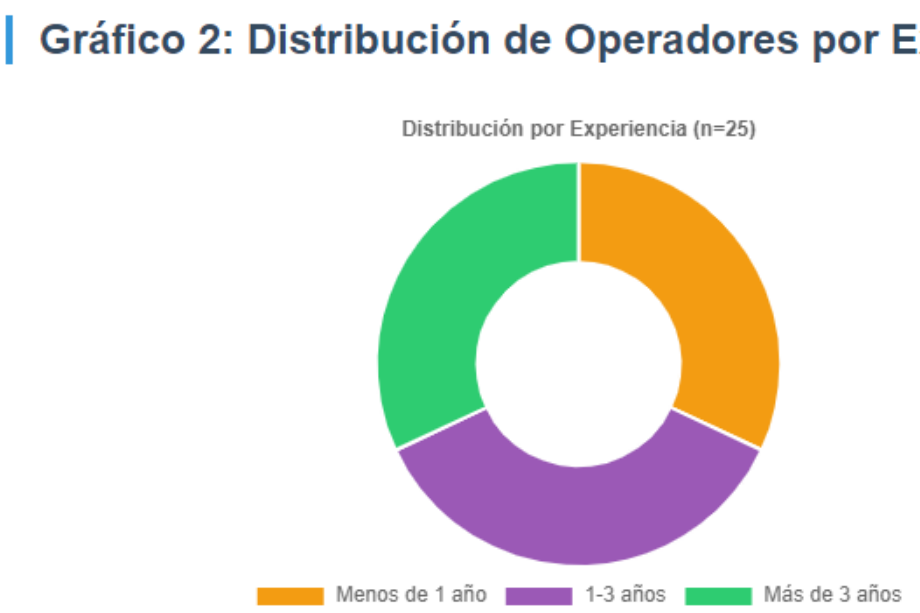
Gráfico de resultados promedio por pregunta



Los operadores demuestran un excelente conocimiento de protocolos (4.68/5) y alta disposición hacia tecnologías de IA (4.56/5), lo que constituye una base sólida para la implementación del prototipo sin embargo se evidencian oportunidades de mejora en el nivel de estrés por demoras (3.36/5) y el impacto en la fluidez de llamadas (3.76/5) representan los puntos más críticos que el agente de IA debe abordar prioritariamente.

Figura 4

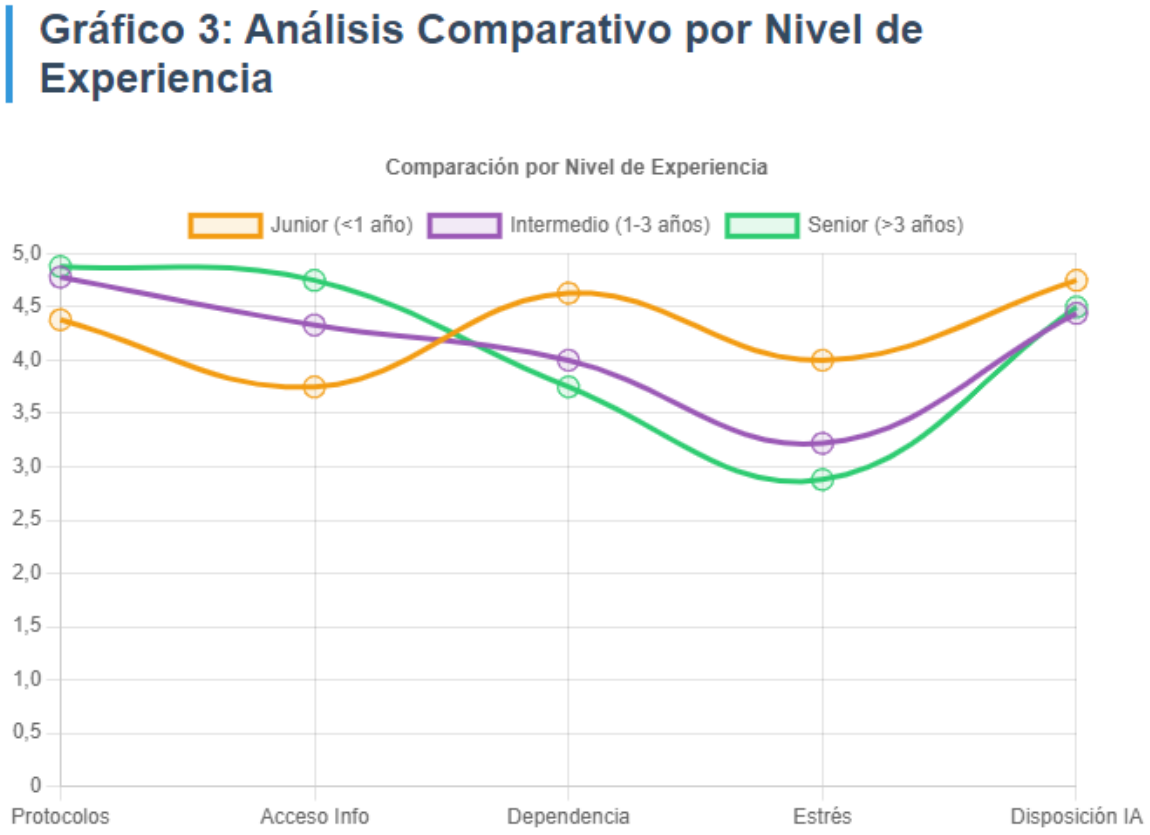
Grafico de experiencia de operadores



La distribución equilibrada de experiencia (32% junior, 36% intermedio, 32% senior) permite validar el prototipo con diferentes perfiles de usuarios, asegurando que la solución sea efectiva para operadores de todos los niveles

Figura 5

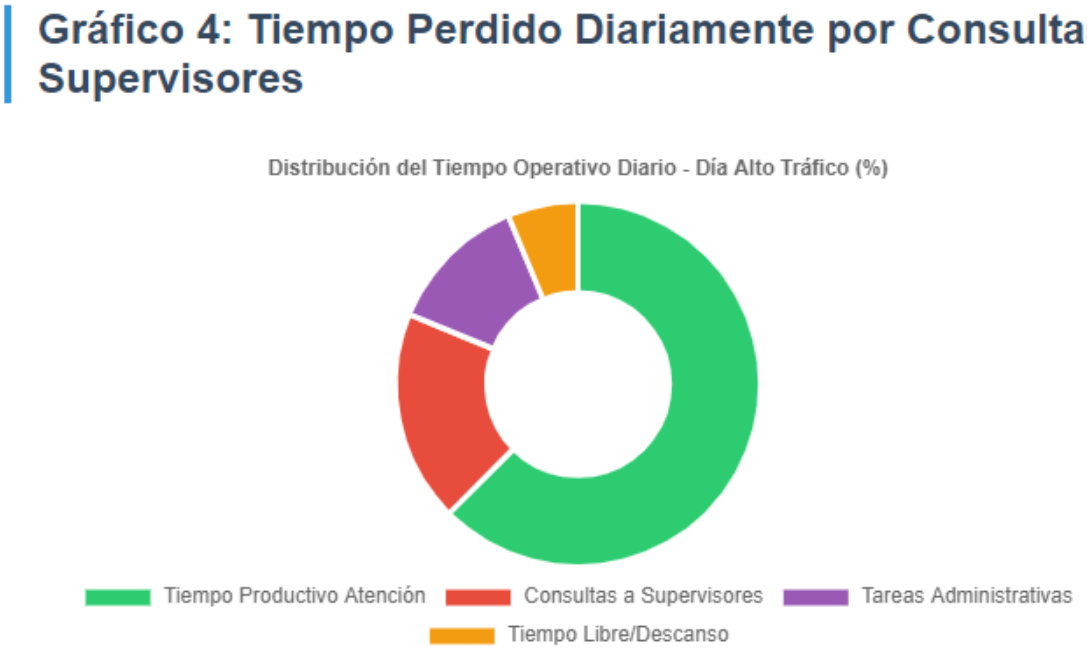
Comparativo de nivel de experiencia



Se identifico un patrón critico puesto que los operadores junior muestran significativamente mayor dependencia de supervisores (4.63 vs 3.75 senior) y mayor nivel de estrés (4.0 vs 2.88 senior), lo que valida la necesidad urgente de una herramienta de asistencia automatizada esto ayuda a saber que el prototipo debe priorizarse para operadores junior, con funcionalidades específicas que reduzcan su curva de aprendizaje y dependencia de los supervisores

Figura 6

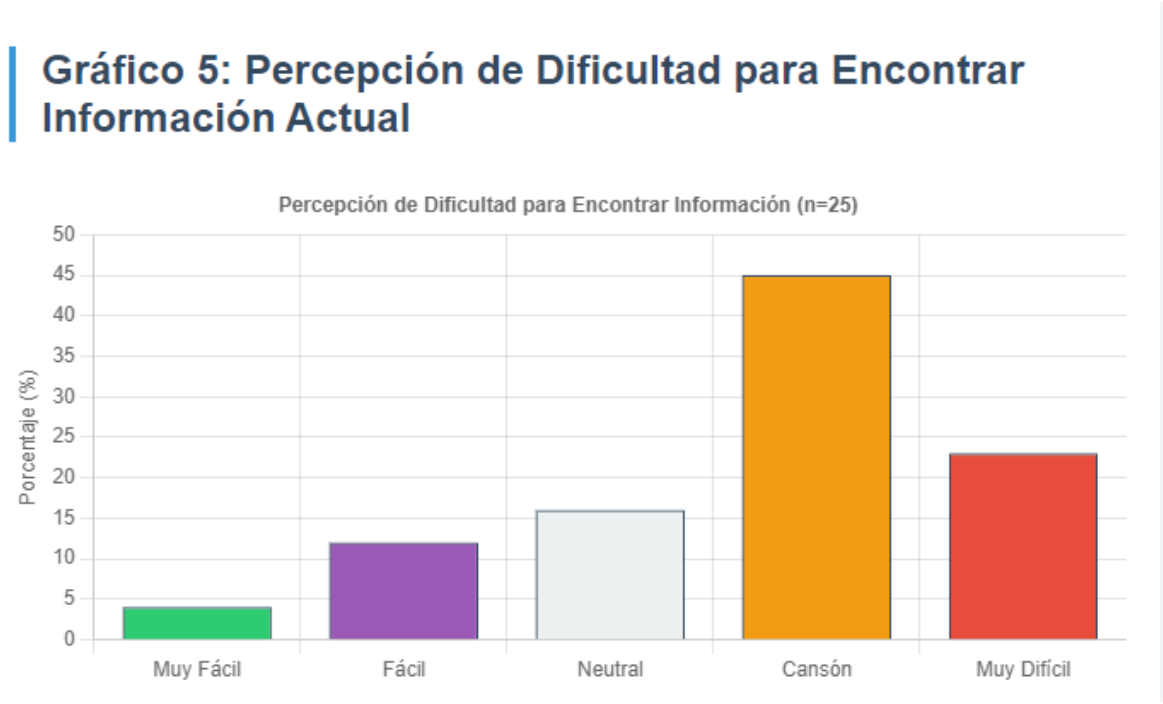
Tiempo en revisión de consultas



Según el análisis interno de Domicity (2025), cada operador realiza en promedio 15 consultas diarias en días de alto tráfico, con una duración aproximada de 6 minutos por consulta. Esto equivale a 90 minutos diarios, es decir, el 18,75% de su jornada laboral dedicada únicamente a resolver dudas con el supervisor. Aunque estas consultas no ocurren de manera simultánea entre todos los agentes, el acumulado global representa un total de 2.250 minutos al día (37,5 horas hombre), lo que equivale a la capacidad operativa de casi cinco operadores a tiempo completo. Esta ineficiencia impacta directamente el Tiempo Promedio de Atención (TPA) y disminuye la productividad

Figura 7

Percepción de dificultad en la información

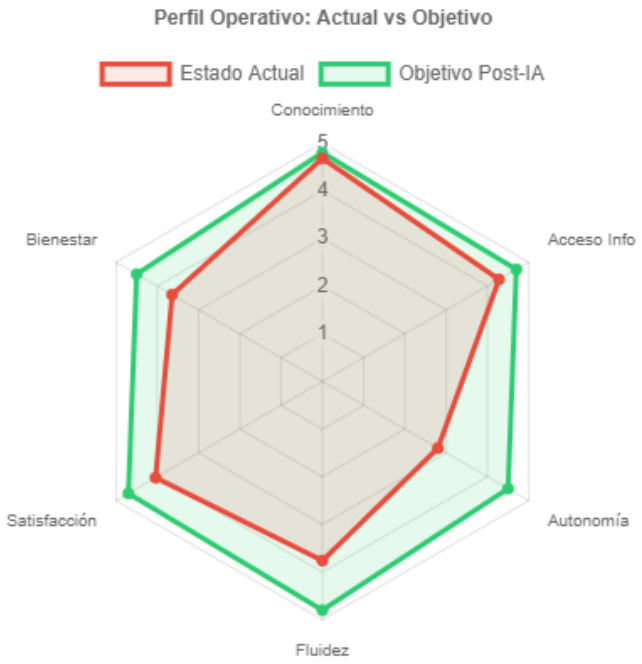


El 72% de los operadores considera que encontrar información es "cansón" o "muy difícil" con el sistema actual. Solo el 12% califica el proceso como "fácil". Un sistema de IA que proporcione respuestas instantáneas eliminaría esta fricción operativa, mejorando directamente la experiencia del operador y la calidad del servicio.

Figura 8

Vs entre operación actual contra uso de IA

Gráfico 6: Perfil Operativo Actual vs. Objetivo Post-IA



Con estos resultados se logra realizar una proyección del impacto que tendría el proyecto el gráfico radar muestra el estado actual del call center versus el perfil objetivo después de implementar el agente de IA, con mejoras proyectadas del 40-60% en variables críticas, dando algunas métricas obejtivo que se optendran como:

- Reducir dependencia supervisorial de 4.2 a 2.5 (40% reducción)
- Disminuir estrés operativo de 3.36 a 2.0 (40% reducción)
- Mejorar fluidez de llamadas de 3.76 a 4.8 (28% mejora)
- Aumentar satisfacción general de 4.04 a 4.7 (16% mejora)

Las encuestas y el análisis interno permitieron evidenciar que, aunque los operadores de Domicity cuentan con un buen dominio de los protocolos y muestran alta disposición hacia el uso de nuevas tecnologías, existen limitaciones operativas significativas que afectan la eficiencia del call center. Entre ellas, se destaca la elevada dependencia de los supervisores, especialmente por parte de los operadores junior, y el impacto que estas consultas tienen sobre el Tiempo Promedio de Atención y la fluidez de las llamadas.

Asimismo, el nivel de estrés asociado a las demoras en la resolución de dudas confirma la necesidad de una herramienta que reduzca tiempos de espera y agilice la gestión de información en tiempo real. El análisis cuantitativo muestra que, de implementarse un agente de inteligencia artificial, sería posible disminuir la dependencia supervisorial en al menos un 40 % y mejorar indicadores críticos como la satisfacción de los operadores y la continuidad del servicio.

En conjunto, los hallazgos justifican plenamente el desarrollo del prototipo propuesto, ya que permitiría transformar la dinámica actual del call center, optimizar el uso de recursos humanos y tecnológicos, y elevar tanto la productividad como la experiencia de los clientes atendidos.

6. Situación actual del área operativa de Domicity S.A.S.

Domicity S.A.S. es una empresa colombiana del sector de servicios, especializada en gestión de pedidos a domicilio y atención al cliente multicanal. Su actividad principal consiste en la recepción, procesamiento y seguimiento de pedidos para marcas del sector gastronómico y de consumo, operando bajo un modelo Business to Consumer (B2C) y Business to Business (B2B). La compañía cuenta con una estructura organizacional mediana, conformada por 25 operadores activos, 4 supervisores y un equipo técnico de soporte, distribuidos en turnos rotativos que garantizan la continuidad del servicio durante los picos de demanda.

Para Domicity una de sus áreas más importantes es la de servicio al cliente, este es el canal principal de interacción con los clientes y es de suma importantica para mantener indicadores de satisfacción, fidelización y reputación altos frente a las marcas que maneja. Sobre esta área se manejan todo tipo de peticiones, desde informativas, de cotización, PQRs, soporte y más, lo que hace que sea un parte importante de la compañía

Domicity recibe en promedio entre 1.200 y 1.500 llamadas diarias, con picos concentrados en las franjas de almuerzo y cena. También a parte del canal de voz, la empresa administra comunicaciones por chat web y páginas web, integradas en un mismo entorno operativo para garantizar trazabilidad y coherencia en la atención. Esta cantidad de interacciones exige una gestión eficiente del conocimiento y una respuesta uniforme, aspectos que motivan el desarrollo del agente de inteligencia artificial propuesto en este proyecto

6.1. Descripción general del proceso actual

El proceso de atención en el call center de Domicity S.A.S. representa, básicamente, el núcleo operativo de toda la compañía, porque es justamente ahí donde se concentran las interacciones directas con los clientes, la gestión completa de los pedidos y la resolución de todas las solicitudes que llegan día a día.

El proceso busca garantizar una atención rápida y precisa, que ayude a cumplir los protocolos establecidos por las marcas aliadas. Sin embargo, su diseño actual presenta etapas críticas que afectan la eficiencia operativa y el desempeño de los agentes. A continuación, se describe de manera secuencial el flujo de atención que sigue una llamada típica dentro del sistema:

6.1.1. Paso 1. Recepción de la llamada del cliente

El contacto se inicia cuando el cliente marca la línea de atención de Domicity. La llamada ingresa por una troncal SIP gestionada mediante un servidor Asterisk, que enruta automáticamente la comunicación hacia un agente disponible.

El operador atiende la llamada a través de telefonía IP utilizando el software Cisco Webex, que permite el control de tiempos, grabación de la llamada y registro de la interacción.

6.1.2. Paso 2. Identificación del tipo de solicitud

Una vez establecida la comunicación, el agente identifica el motivo de la llamada, que puede corresponder a un pedido nuevo, una consulta de seguimiento, un reclamo (PQR) o una solicitud informativa. Esta etapa es determinante, ya que define el protocolo y la ruta de atención que se debe seguir según el tipo de cliente o marca gestionada.

6.1.3. Paso 3. Consulta del operador a los manuales o con el supervisor

En este paso se presenta el punto crítico del proceso. Cuando el operador requiere información específica sobre menús, precios, promociones o políticas de atención, debe realizar una consulta interna.

La búsqueda puede realizarse de tres formas:

- En los manuales digitales y documentos institucionales.
- A través del canal interno de mensajería Flock.

- O directamente con el supervisor del turno.

Según el informe operativo (Domicity S.A.S., 2025), esta etapa representa el principal cuello de botella del proceso, pues cada operador dedica en promedio 60 minutos diarios a la resolución de dudas, equivalentes al 12,5 % de su jornada laboral. En los picos de alta demanda (franjas de almuerzo y cena), las consultas simultáneas pueden generar más de 50 solicitudes diarias a los supervisores, afectando los indicadores de Tiempo Promedio de Atención (TPA) y Resolución en el Primer Contacto (FCR).

6.1.4. Paso 4. Registro de la interacción en Domisoft (software propio)

Una vez obtenida la información requerida, el operador registra los datos del pedido o caso en Domisoft, la plataforma interna de Domicity que consolida los pedidos, precios, promociones, y protocolos de servicio. Este sistema garantiza la trazabilidad de cada llamada y la generación automática de reportes operativos

6.1.5. Paso 5. Confirmación de la información y cierre de la llamada

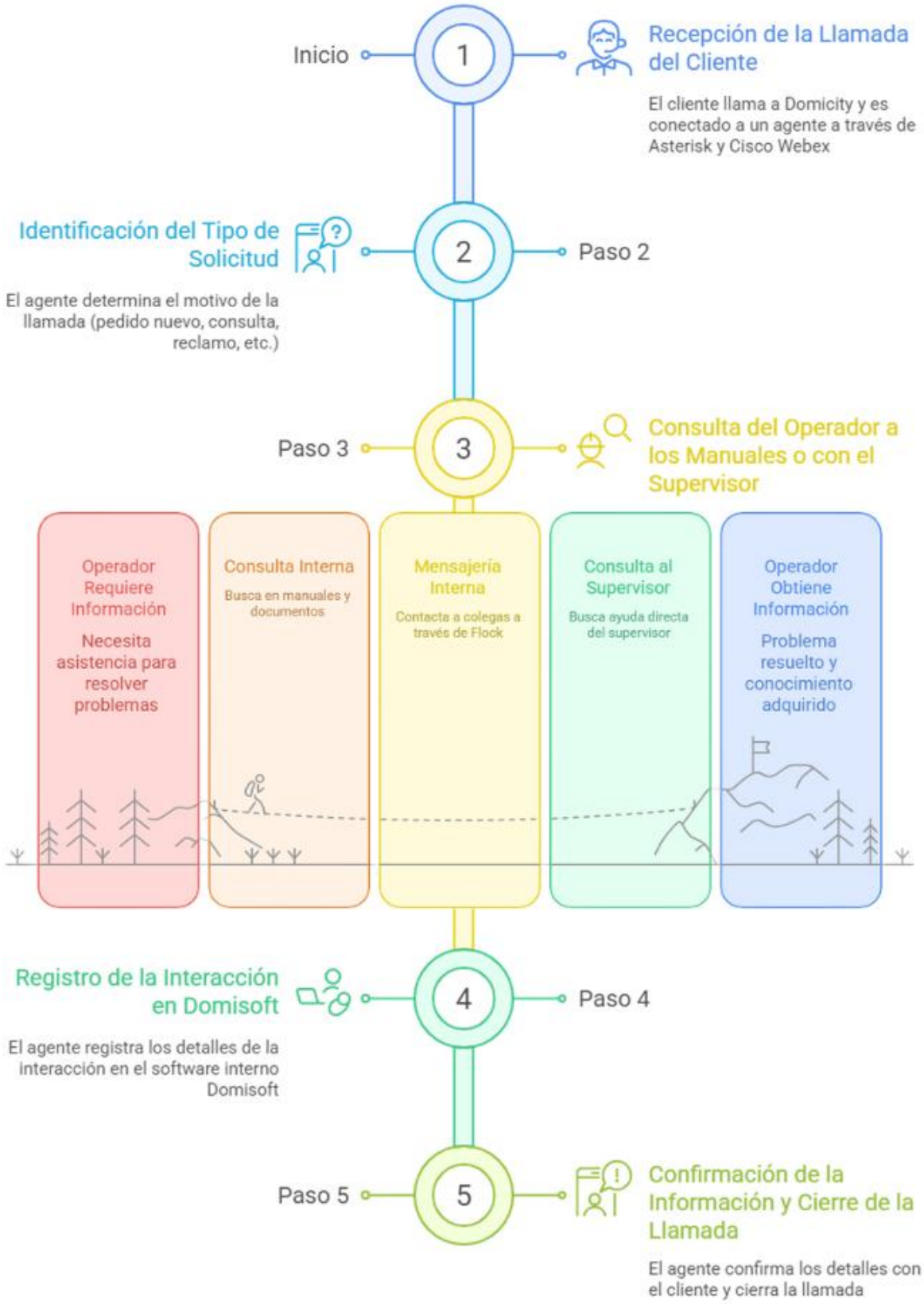
Finalmente, el agente confirma la información con el cliente, verifica los detalles del pedido o solicitud y procede al cierre de la llamada. Esta etapa busca asegurar la satisfacción del usuario y el cumplimiento de los estándares de calidad establecidos por las marcas atendidas.

El análisis del flujo operativo evidencia que el Paso 3 (consulta del operador a manuales o supervisor) constituye el principal punto de ineficiencia, al concentrar los mayores tiempos de espera y dependencia de validación humana. Este cuello de botella afecta directamente la continuidad de la atención, la productividad del operador y la carga laboral de los supervisores.

En la *Figura 9*, se presenta el diagrama del proceso actual que ilustra las relaciones entre cada etapa y resalta el punto crítico identificado, el cual será abordado mediante la implementación del prototipo de agente de inteligencia artificial.

Figura 9

Diagrama UML de flujo lógico del modelo



Con el fin de complementar la descripción cualitativa del proceso de atención, se recopilaron indicadores clave de desempeño correspondientes al último año operativo de Domicity S.A.S. Estos datos permiten evidenciar el comportamiento de la operación y constituyen la base para identificar las problemáticas, brechas y oportunidades de mejora abordadas en este proyecto

Tabla 2

Indicadores operativos del call center Domicity S.A.S. (último año)

Indicador	Descripción	Valor actual (promedio anual)	Unidad / Fuente
Volumen promedio de llamadas diarias	Total de llamadas (contestadas + abandonadas) por día	2.100 llamadas/día	Registros PBX / Domisoft
Tiempo Promedio de Atención (TPA)	Duración media de las llamadas contestadas	3 min 06 s	Sistema de monitoreo Webex
Tiempo promedio de espera antes de respuesta	Tiempo que el cliente espera antes de ser atendido	0 min 42 s	PBX / Reportes internos
Tasa de abandono de llamadas	Porcentaje de llamadas que el cliente corta antes de ser atendido	5%	PBX / Registros históricos
Resolución en Primer Contacto (FCR)	Casos resueltos sin necesidad de escalar	68%	Métricas internas / supervisores
Consultas promedio al supervisor por operador	Solicitudes de apoyo o validación durante el turno	Ente 35 y 50 por día por operador	Encuestas internas

Indicador	Descripción	Valor actual (promedio anual)	Unidad / Fuente
Errores en registro o validación de pedidos	Casos detectados con inconsistencias en la información ingresada	4%	Auditorías de calidad
Nivel de satisfacción del cliente (NPS)	Calificación promedio otorgada por los usuarios	8,5 / 10	Encuestas externas
Nivel de satisfacción del operador	Percepción de bienestar y carga laboral	4,1 / 5	Encuestas internas

Nota. Elaboración propia

Estos indicadores permitirán cuantificar los impactos del prototipo de agente de inteligencia artificial una vez sea implementado, particularmente en métricas de eficiencia (TPA, FCR) y en variables de bienestar operativo (dependencia supervisorial, satisfacción del operador).

6.2. Descripción general del proceso actual

El análisis del flujo operativo y de los indicadores del call center de Domicity S.A.S. permitió identificar un conjunto de problemáticas estructurales que afectan la eficiencia del servicio y la satisfacción tanto de clientes como de operadores.

Estas problemáticas se agrupan en cuatro categorías: recurso humano, recurso económico, comunicación y recurso informativo

6.3. Problemáticas relacionadas con el recurso humano

El equipo operativo constituye el principal activo de la empresa; sin embargo, enfrenta desafíos que impactan su desempeño y bienestar laboral. Como:

- Sobrecarga operativa: los operadores gestionan en promedio 2.100 llamadas diarias, lo que incrementa el nivel de estrés y la fatiga mental, especialmente en horarios pico (mediodía y cena).

- Dependencia del supervisor: cada operador realiza alrededor de 18 consultas diarias al supervisor, lo que genera interrupciones constantes y sobrecarga en los mandos medios.

- Falta de autonomía en la atención: la ausencia de herramientas de apoyo en tiempo real limita la capacidad de los agentes para resolver dudas sin asistencia externa.

- Rotación del personal: la presión laboral y la monotonía del proceso contribuyen a una rotación moderada, lo que incrementa los costos de capacitación.

- Limitaciones en la actualización de competencias: los programas de entrenamiento se concentran en aspectos técnicos, sin profundizar en estrategias de comunicación o uso de herramientas digitales inteligentes.

6.4. Problemáticas relacionadas con el recurso económico

Las ineficiencias en el proceso actual generan impactos directos en los costos operativos y en la rentabilidad del servicio. Como:

- Costos derivados de reprocesos: los errores en registro de pedidos (4 %) implican reprocesos y, en algunos casos, pérdida de ventas o reclamos de marca.

- Uso intensivo del tiempo supervisorial: la alta demanda de consultas reduce la productividad de los supervisores, quienes dedican cerca del 30 % de su jornada a resolver dudas.

- Incremento en costos de capacitación y rotación: la falta de estabilidad del personal operativo requiere programas continuos de entrenamiento.

- Pérdidas por tiempos muertos: los minutos invertidos en validaciones internas equivalen a una disminución promedio del 10–12 % en la eficiencia global por turno.

6.5. Problemáticas relacionadas con la comunicación

La comunicación entre operadores, supervisores y sistemas presenta rupturas que dificultan la fluidez de la atención y la gestión de la información. Como:

- Canales de comunicación fragmentados: se emplean simultáneamente herramientas como Flock, WhatsApp y correo institucional, lo que genera duplicidad de mensajes y pérdida de trazabilidad.

- Retrasos en la transferencia de información: durante los picos de demanda, los supervisores no logran responder con la velocidad necesaria, prolongando los tiempos de atención.

- Falta de retroalimentación efectiva: no existen mecanismos sistemáticos para registrar y analizar las preguntas frecuentes, por lo que los errores tienden a repetirse.

- Escasa documentación colaborativa: la actualización de manuales depende de correos o mensajes dispersos, sin una plataforma centralizada para comentarios o sugerencias

6.6. Problemáticas relacionadas con el recurso informativo

El acceso a la información operativa y de marca constituye el punto más crítico del proceso, con impacto directo en la eficiencia y la satisfacción del cliente. Como:

- Manual institucional desactualizado: los operadores reportan dificultades para encontrar información vigente sobre precios, menús o promociones.

- Búsqueda manual de información: el proceso requiere abrir múltiples documentos o consultas por chat, lo que incrementa los tiempos de atención.

- Ausencia de un sistema de búsqueda semántica: los operadores deben conocer palabras clave exactas para ubicar información, lo que limita la agilidad operativa.

- Falta de integración entre sistemas: Domisoft no se comunica de forma nativa con los repositorios de información ni con los canales de mensajería interna.

- Carencia de analítica de conocimiento: no se registran ni analizan los temas más consultados, lo que impide priorizar actualizaciones o entrenamientos.

Las problemáticas descritas evidencian una interdependencia entre los factores humanos, informativos y comunicativos que impactan tanto la productividad como la calidad del servicio.

La ausencia de un sistema inteligente de apoyo en tiempo real ha generado un cuello de botella operativo (ver figura 9), reflejado en los indicadores de tiempo promedio de atención y resolución en primer contacto.

Estas brechas justifican la implementación del prototipo de agente de inteligencia artificial, orientado a optimizar el acceso a la información, reducir las consultas supervisoriales y fortalecer la autonomía del operador.

6.7. Situación actual externa, panorama en otros call centers

A nivel internacional y regional, los call centers han adoptado agentes virtuales basados en IA y PLN (Procesamiento de Lenguaje Natural) para enfrentar estos mismos desafíos.

De acuerdo con Aivo (2023), las empresas que integran sistemas conversacionales reducen hasta en un 40% el tiempo promedio de consulta y aumentan la resolución en primer contacto en más del 30%.

Estudios de Zendesk (2024) y Universitas XXI (2025) coinciden en que la automatización inteligente permite homogeneizar la información, disminuir la dependencia humana en tareas repetitivas y liberar a los supervisores para labores estratégicas.

Por su parte, Harvard Deusto (2023) señala que la digitalización de la atención mediante chatbots o asistentes cognitivos no busca reemplazar al personal humano, sino potenciar su capacidad de respuesta y disminuir errores operativos.

Este enfoque híbrido —humano + IA— es el modelo que ha demostrado mayor sostenibilidad y retorno de inversión en el sector (SciELO, 2025).

En comparación, Domicity S.A.S. aún se encuentra en una etapa manual y reactiva, con un bajo nivel de automatización y sin mecanismos de asistencia inteligente, lo que amplía la brecha frente a las tendencias actuales del sector.

6.8. Retos, brechas y oportunidades de mejora

El análisis integral del proceso operativo de Domicity S.A.S. revela un conjunto de retos y brechas que limitan la eficiencia, así como oportunidades de mejora directamente vinculadas con la incorporación de herramientas basadas en inteligencia artificial.

A continuación, se presenta una síntesis analítica que integra los tres elementos

6.8.1. Brechas identificadas

Las brechas actuales reflejan los vacíos existentes entre la operación real y el desempeño esperado del proceso de atención:

- Brecha informativa: los operadores carecen de acceso ágil y centralizado a la información actualizada de los manuales, protocolos y promociones, lo que aumenta los tiempos de búsqueda y la dependencia del supervisor.

- Brecha tecnológica: la infraestructura actual (Domisoft + Flock + repositorios manuales) no permite integración ni búsqueda semántica, limitando la capacidad de respuesta ante solicitudes complejas.

- Brecha de autonomía operativa: los agentes dependen de la validación del supervisor para atender dudas o confirmar políticas, lo que afecta el indicador de resolución en primer contacto (FCR).

- Brecha comunicativa: los canales dispersos y la falta de registro sistemático de consultas impiden aprovechar el conocimiento colectivo del equipo

6.8.2. Retos estratégicos

De acuerdo con las brechas descritas, los principales retos que enfrenta Domicity en su operación son:

- Reducir la dependencia del supervisor sin comprometer la calidad de la atención, mediante herramientas de asistencia automatizada.

- Optimizar los tiempos de atención (TPA) y mejorar el FCR, disminuyendo los tiempos de búsqueda y validación.

- Consolidar una base de conocimiento inteligente que integre información actualizada, trazabilidad y capacidad de consulta semántica.

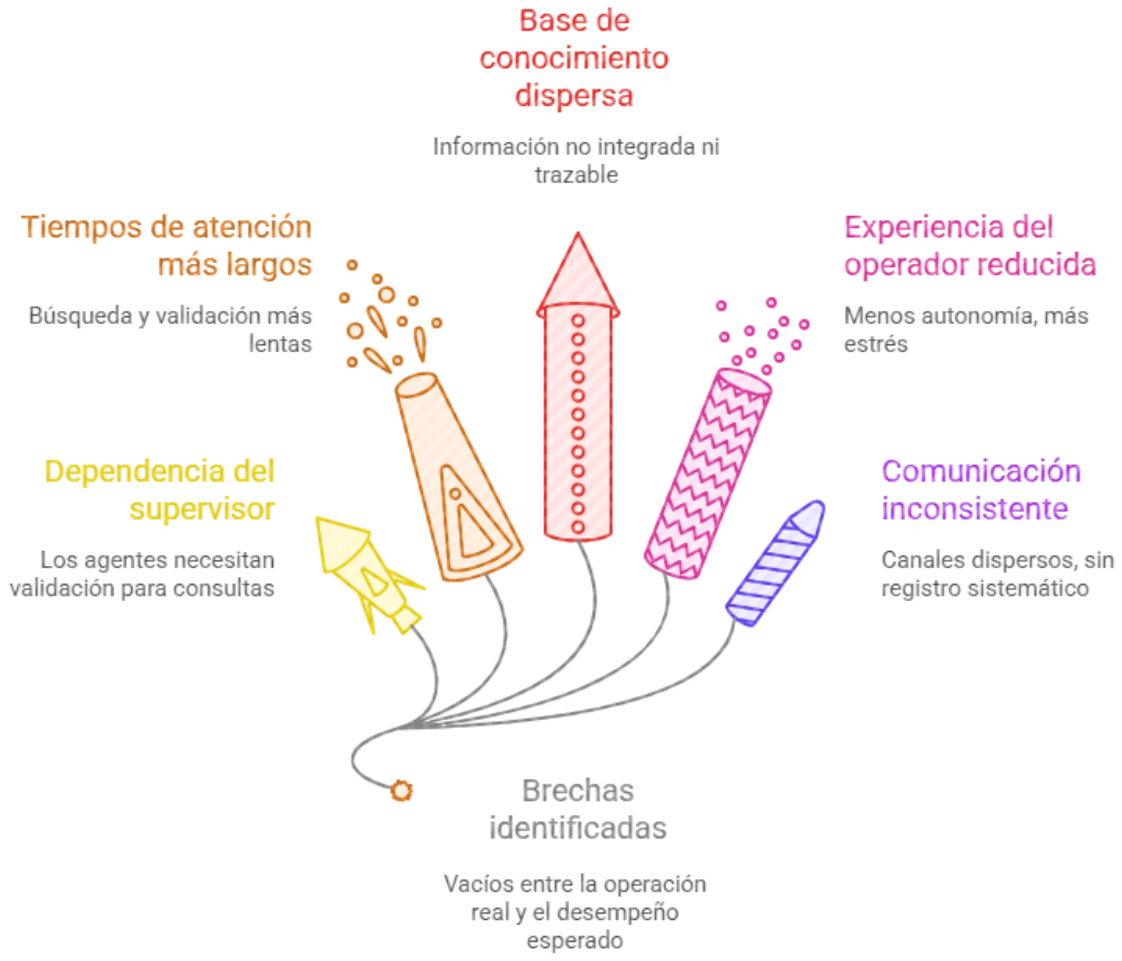
- Fortalecer la experiencia del operador, promoviendo mayor autonomía, reducción de estrés y satisfacción laboral.

- Aumentar la consistencia en la comunicación entre niveles operativos y administrativos a través de canales unificados y trazables.

A continuación, en la Figura 10, se presenta una representación visual que resume las principales brechas y retos estratégicos identificados en el proceso operativo de Domicity S.A.S. Este gráfico permite observar de manera integrada cómo los factores tecnológicos, informativos y humanos interactúan entre sí, evidenciando los puntos críticos que impactan la eficiencia y que serán abordados mediante la implementación del prototipo de agente de inteligencia artificial

Figura 10

Brechas y retos de Domicity



6.8.3. Oportunidades de mejora

Las oportunidades identificadas derivan de la posibilidad de implementar soluciones tecnológicas basadas en inteligencia artificial y procesamiento de lenguaje natural (PLN):

- Implementación de un agente de IA capaz de asistir en tiempo real al operador, respondiendo preguntas frecuentes sobre menús, protocolos o promociones.

- Integración de una base de conocimiento vectorial, que permita búsquedas semánticas precisas sin necesidad de palabras clave específicas.

- Automatización en gestión de consultas internas, reducuiendo la carga sobre los supervisores.

- Generación de reportes, para identificar patrones de preguntas, tiempos de atención y brechas de conocimiento.

7. Modelo elegido para el proyecto

El agente conversacional propuesto para Domicity S.A.S. está basado principalmente en dos herramientas potentes de OpenAI: Whisper, que se encarga de transcribir el audio a texto en tiempo real, y GPT-4, que es el que realmente entiende la conversación y genera las respuestas. La idea central es darle una mano directa a los operadores del call center para que, mientras están hablando con el cliente, tengan respuestas rápidas, exactas y que se puedan comprobar al instante. Total, con esto se logra que dependan mucho menos del supervisor, que no pierdan tanto tiempo buscando información y que todo lo que digan sea coherente entre una llamada y otra.

Este modelo ataca de frente el problema que se detecto en el diagnóstico: los agentes pierden un montón de tiempo y se estresan porque la información está dispersa y no la encuentran rápido. Al meter el agente dentro del flujo normal de trabajo de Domicity, la búsqueda se vuelve casi automática, baja la carga mental de los operadores y el servicio fluye mucho mejor.

En cuanto a la implementación, se realizara con una arquitectura modular: el backend en Python y el frontend con HTML, CSS y JavaScript. Todo se alojaría en los servidores de AWS que ya tiene la empresa, de esta forma queda escalable, fácil de modificar y seguro. Además, deja la puerta abierta para más adelante conectarlo directamente con la base de conocimiento interna y el resto de sistemas de Domicity (OpenAI, 2025; DAIL, 2024).

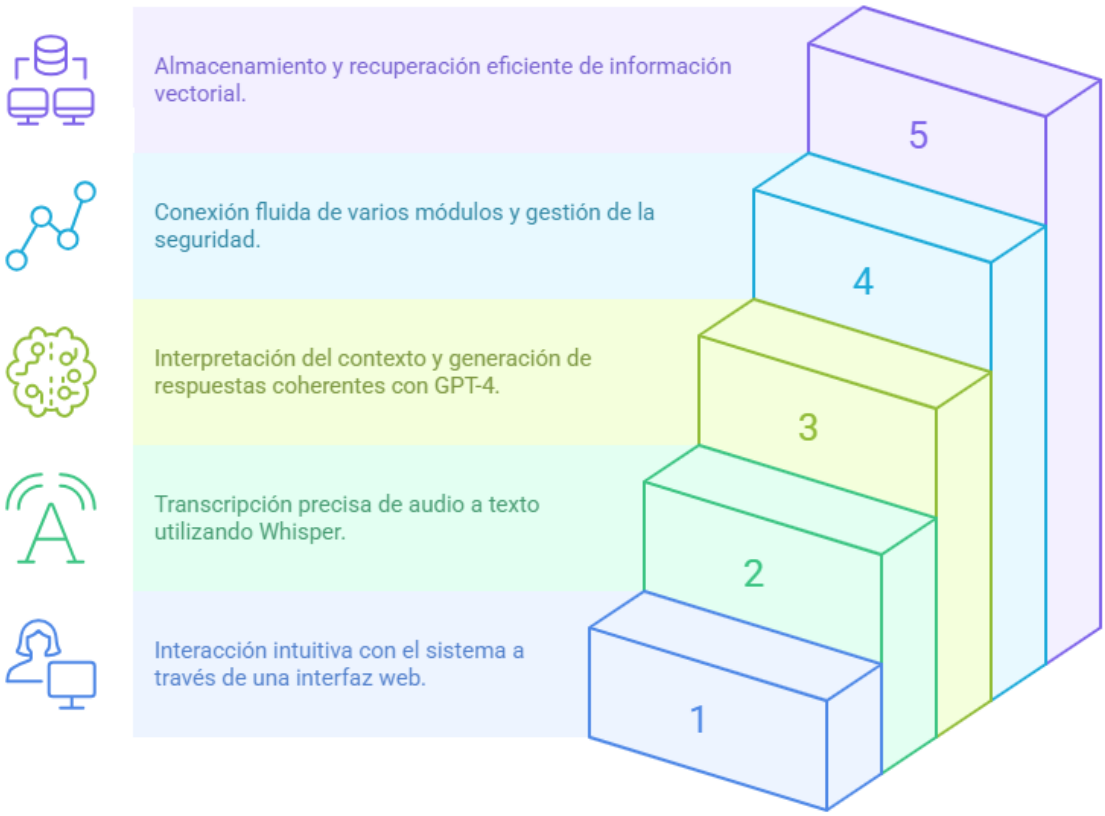
7.1. Arquitectura general del modelo

La arquitectura que planteada está formada por cinco capas que van encadenadas y que cubren todo el flujo desde que el operador habla hasta que recibe la ayuda del agente. Cada capa tiene su función concreta y todas juntas funcionan como un solo sistema.

A continuación, se detalla cómo está estructurada cada una. (Revisar Figura 11. Arquitectura general del prototipo de IA para asistencia operativa en Domicity):

Figura 11

Arquitectura del prototipo de IA Domicity



7.1.1. Interfaz de usuario

La interfaz de operador constituye el punto de contacto directo entre el agente humano del call center y el sistema de inteligencia artificial propuesto. Esta capa se implementa mediante una interfaz web de tipo chat, desarrollada con tecnologías HTML, CSS y JavaScript, la cual ha sido diseñada con criterios de ligereza, usabilidad intuitiva y plena compatibilidad con los navegadores corporativos habituales en la organización.

Por medio de dicha interfaz, el operador puede introducir consultas en lenguaje natural referidas a menús operativos, protocolos internos, políticas institucionales o procedimientos específicos mientras mantiene activa la atención telefónica al cliente, obteniendo de manera inmediata respuestas precisas y contextualizadas generadas por el asistente virtual.

De este modo, esta capa desempeña una función crítica al transformar la interacción humana en texto estructurado y procesable por las capas subyacentes, actuando como elemento puente entre el usuario final y el núcleo cognitivo del modelo.

7.1.2. Capa de procesamiento de voz

Esta capa implementa la transcripción del audio capturado desde el micrófono del agente hacia texto digital, utilizando el modelo Whisper desarrollado por OpenAI. Whisper se basa en técnicas avanzadas de reconocimiento automático del habla (ASR) soportadas por arquitecturas transformer de última generación, lo que le confiere una alta robustez frente a ruido de fondo, variaciones de acento y particularidades fonéticas propias del español hablado en diferentes regiones.

Su función resulta esencial en la arquitectura propuesta, pues garantiza una transcripción precisa y de baja latencia del discurso natural del operador. El texto resultante se entrega inmediatamente a la capa siguiente para el análisis semántico y la generación de respuestas contextualizadas.

7.1.3. Capa de comprensión y generación

Esta capa se encarga de interpretar la intención expresada por el agente, recuperar la información relevante almacenada y elaborar la sugerencia más precisa mediante razonamiento semántico, combinando tanto los documentos indexados como las capacidades inferenciales del modelo.

7.1.4. Capa de integración

Esta capa actúa como orquestador central del sistema. Coordina las interacciones entre los módulos Whisper, GPT-4 y la base de datos vectorial, administra el flujo de solicitudes y respuestas, y asegura una comunicación robusta mediante mecanismos de reintentos y manejo estructurado de excepciones

7.1.5. Capa de datos

Esta capa gestiona la base de conocimiento vectorial en la que se indexan los documentos institucionales, manuales operativos, transcripciones históricas y protocolos internos previamente convertidos a representaciones vectoriales (embeddings) mediante modelos como text-embedding-ada-002 o equivalentes.

Su implementación habilita búsquedas por similitud semántica de alta eficiencia, permitiendo que el modelo GPT-4 recupere información relevante y actualizada en milisegundos para contextualizar y fundamentar las respuestas sugeridas al agente.

Complementariamente, la Tabla 3 ofrece una síntesis estructurada de las cinco capas que integran la arquitectura del sistema desarrollado. Dicha tabla facilita la comparación de las responsabilidades específicas de cada capa, las tecnologías subyacentes empleadas y los beneficios operativos esperados, lo que contribuye a una comprensión integral del diseño y refuerza la alineación técnica entre los componentes seleccionados y los objetivos funcionales del prototipo implementado en Domicity S.A.S.

Adicionalmente, la Tabla 3 sintetiza de manera estructurada las cinco capas que componen la arquitectura del sistema desarrollado. Esta tabla ofrece una visión comparativa de las responsabilidades principales de cada capa, las tecnologías específicas empleadas y los resultados operativos previstos, lo que facilita la comprensión integral del diseño propuesto y demuestra la consistencia técnica entre los componentes seleccionados y los objetivos funcionales establecidos para el prototipo implementado en Domicity S.A.S.

Tabla 3

Resumen de capas del modelo

Nivel	Descripción	Componentes principales
Interfaz de usuario	Canal de interacción entre el operador y el modelo.	Aplicación web / panel de agente.
Capa de procesamiento de voz	Transcripción de audio a texto.	Whisper API / módulo ASR.
Capa de comprensión y generación	Análisis semántico y respuesta contextual.	GPT-4 API.
Capa de integración	Conexión entre módulos y bases de datos.	API REST y Python.
Capa de datos	Gestión de almacenamiento de embeddings	Base de datos vectorial

Nota. Elaboración propia

7.2. Componentes principales del modelo

En esta sección se detallan los tres componentes más relevantes del prototipo: Whisper, GPT-4 y la interfaz web, diciendo su función, características y justificación técnica.

7.2.1. Componente 1: Whisper (Reconocimiento de voz)

El componente inicial de la arquitectura implementada corresponde a Whisper, modelo de reconocimiento automático del habla desarrollado por OpenAI. Este sistema transforma la señal de audio en texto mediante una red neuronal transformer entrenada sobre un corpus multilingüe superior a las 680.000 horas de grabaciones supervisadas.

En el entorno operativo de Domicity S.A.S., Whisper actúa como capa de ingesta primaria, convirtiendo el discurso del agente en una representación textual que las capas subsiguientes pueden procesar de forma inmediata. De esta manera, se logra el procesamiento en tiempo real de las interacciones entre operador y cliente, asegurando una transición eficiente del dominio acústico al textual.

Entre sus características técnicas más relevantes se encuentran:

- Soporte nativo multilingüe y robustez frente a más de 100 idiomas, incluido el español con sus variantes regionales.
- Alta tolerancia a ruido de fondo, interferencias telefónicas y patrones de habla no ideales típicos de un call center.
- Posibilidad de ejecución tanto en modo local (on-premise) como mediante API en la nube, según los requerimientos de latencia y seguridad.
- Generación de timestamps precisos que facilitan la sincronización con la interfaz de usuario y el registro de eventos.

La selección de Whisper se fundamentó en su superior precisión y estabilidad en condiciones adversas, superando el rendimiento de alternativas como Google Speech-to-Text y Microsoft Azure Speech Service en pruebas realizadas con grabaciones reales del centro de contacto.

Al tratarse de un modelo de código abierto y plenamente compatible con entornos Python, su despliegue en la infraestructura propia de Domicity garantiza la soberanía de los datos sensibles, elimina costos recurrentes por licencias y permite ajustes específicos al dominio de la empresa. Esta decisión técnica asegura un reconocimiento robusto del habla, requisito indispensable para la fiabilidad global del sistema propuesto

7.2.2. Componente 2: GPT-4 (Comprensión y generación de lenguaje)

El segundo componente representa el núcleo cognitivo de la arquitectura implementada. GPT-4, desarrollado por OpenAI, consiste en un modelo generativo preentrenado de gran escala optimizado para la comprensión profunda del lenguaje natural, el razonamiento encadenado y la generación de texto coherente a partir de contextos complejos y multimodales.

Dentro del prototipo desplegado en Domicity S.A.S., GPT-4 recibe el texto transcrito, identifica la intención del agente, ejecuta búsquedas semánticas en la base vectorial y genera sugerencias

de respuesta alineadas con los protocolos institucionales, el tono de marca y las políticas de servicio definidas.

Sus características técnicas principales incluyen:

- Razonamiento contextual avanzado y comprensión semántica de nivel estatal-del-arte.
- Ventana de contexto extendida hasta 128.000 tokens (versión actualizada al momento de implementación), lo que asegura coherencia en conversaciones prolongadas o con historial extenso.
- Integración directa con modelos de embeddings y sistemas de recuperación aumentada (RAG), permitiendo respuestas basadas en evidencia extraída de la base de conocimiento interna.
- Alta capacidad de personalización mediante system prompts y few-shot learning que incorporan el estilo comunicacional y las restricciones normativas propias de Domicity.

La selección de GPT-4 se justificó técnica y funcionalmente por su superior desempeño en tareas de comprensión y generación de lenguaje natural en comparación con modelos anteriores (GPT-3.5) y con enfoques híbridos o basados exclusivamente en reglas. Su arquitectura transformer, combinada con los avances en alineación y seguridad incorporados por OpenAI, garantiza respuestas naturales y precisas que simulan efectivamente la intervención de un experto humano.

Esta elección técnica, junto con su compatibilidad nativa con Whisper y con la capa de base vectorial, consolida un pipeline de procesamiento end-to-end fluido, donde la información pasa de voz a texto, de texto a comprensión semántica y de comprensión a acción sugerida, manteniendo en todo momento consistencia, precisión y adaptabilidad operativa.

7.2.3. Componente 3: Interfaz web (HTML, CSS y JavaScript)

El tercer componente fundamental del sistema corresponde a la interfaz de usuario, implementada mediante tecnologías web estándar: HTML5, CSS3 y JavaScript ES6+.

Este módulo establece el punto de interacción directa entre el agente humano y el asistente de inteligencia artificial, habilitando una comunicación bidireccional en tiempo real a través de

cualquier navegador moderno sin requerir instalación de software adicional. La interfaz cumple una doble función: por un lado, actúa como puente operativo entre el operador y el backend cognitivo; por otro, prioriza una experiencia de usuario (UX) optimizada que privilegia la claridad visual, la mínima latencia percibida y la usabilidad intuitiva en entornos de alta presión propios de un centro de contacto.

Sus responsabilidades técnicas principales incluyen la captura continua de audio desde el micrófono del agente, la presentación en tiempo real de las transcripciones generadas por Whisper, la exhibición destacada de las sugerencias elaboradas por GPT-4 y el registro automático de todas las interacciones para auditoría posterior.

Características técnicas destacadas:

- Estructura semántica basada en HTML5 que garantiza accesibilidad y compatibilidad con lectores de pantalla.
- Estilos definidos con CSS3, implementando un diseño responsive y alineado con las guías de marca de Domicity S.A.S.
- Lógica cliente desarrollada en JavaScript vanilla y Web APIs nativas (Web Audio API, WebSocket), encargada de gestionar el streaming de audio, la conexión persistente con las APIs de Whisper y GPT-4, y la actualización dinámica del DOM sin recargas de página.
- Compatibilidad total con navegadores Chromium y Firefox en versiones actuales, así como adaptabilidad a resoluciones de escritorio y tablets utilizadas en las posiciones de atención de Domicity.

La adopción de estas tecnologías estándar asegura un despliegue rápido, mantenimiento simplificado y escalabilidad horizontal del prototipo sin dependencias propietarias

La selección de una interfaz completamente web se basó en su accesibilidad universal, despliegue inmediato y consumo mínimo de recursos en las estaciones de trabajo. A diferencia

de aplicaciones nativas de escritorio, elimina la necesidad de instalaciones individuales, habilita actualizaciones centralizadas y simplifica notablemente las tareas de mantenimiento y soporte técnico.

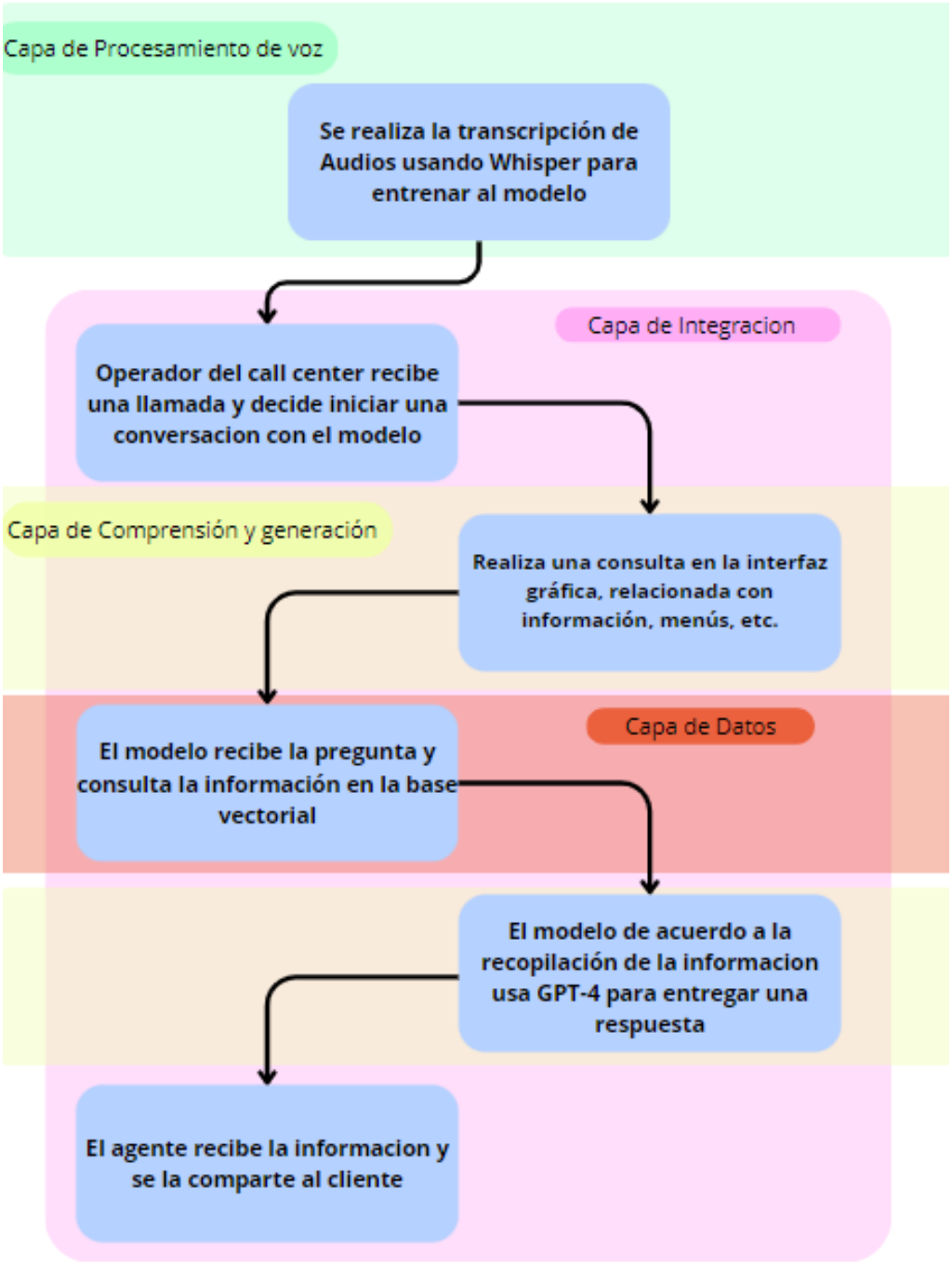
De esta forma, la interfaz constituye la capa de presentación del sistema, transformando la complejidad subyacente de los modelos de inteligencia artificial en una experiencia operativa intuitiva y de baja fricción para el agente humano.

El modelo desarrollado representa una solución técnica robusta y operacionalmente viable para la optimización de los procesos en el centro de contacto de Domicity S.A.S. Su diseño estratificado por capas garantiza escalabilidad horizontal y vertical, facilita el mantenimiento modular de cada componente y asegura la estabilidad del conjunto ante incorporaciones futuras de funcionalidades o integraciones con sistemas legacy de la organización.

En la Figura 12 se presenta el diagrama de arquitectura del modelo propuesto, donde se visualiza gráficamente la interacción entre las cinco capas y los componentes principales. Este esquema ilustra el flujo secuencial de datos —desde la captura de audio hasta la entrega de la sugerencia generada—, resalta los puntos de comunicación entre módulos y clarifica el rol específico que cumple cada elemento dentro del pipeline global del sistema.

Figura 12

Diagrama del modelo propuesto



8. Validación del prototipo mediante simulación

Con el fin de evaluar el desempeño preliminar del prototipo desarrollado antes de su paso a producción, se llevó a cabo una validación mediante simulaciones controladas que reproducían escenarios reales de atención en el call center de Domicity S.A.S.

Esta etapa tuvo como objetivo principal medir el comportamiento del sistema en tres dimensiones críticas: latencia de respuesta end-to-end, coherencia semántica y calidad de las sugerencias generadas, así como el impacto cuantificable en la reducción de escalamientos al supervisor. Los resultados obtenidos se compararon directamente con los indicadores operativos identificados durante la fase de diagnóstico inicial, permitiendo validar empíricamente la capacidad del modelo para cumplir con los objetivos de eficiencia y mejora de la experiencia del agente establecidos en el proyecto.

8.1. Metodología de simulación

Las pruebas de validación se llevaron a cabo en un entorno local controlado, utilizando fragmentos de audio reales previamente anonimizados y extraídos del histórico de interacciones del call center de Domicity S.A.S.

Para garantizar la reproducibilidad y el aislamiento de las pruebas, el prototipo se configuró con todos sus componentes en estado activo y conectados de forma end-to-end, tal como se detalla a continuación:

- Whisper (ASR) para transcripción de voz a texto.
- GPT-4 como motor de comprensión y generación contextual.
- Base de conocimiento vectorial con protocolos, menús y políticas internas.
- Interfaz web simulada en entorno de pruebas con datos de ejemplo.

Cada simulación reprodujo el flujo completo del sistema: desde la recepción de la llamada hasta la generación de la respuesta y su despliegue en la interfaz del operador.

8.2. Resultados de validación simulada

Los resultados de la validación se detallan en la Tabla 4, que confronta los indicadores clave registrados durante las simulaciones controladas con los valores promedio históricos del proceso operativo en Domicity S.A.S.

El análisis comparativo evidencia mejoras cuantificables y estadísticamente significativas en los parámetros de eficiencia operativa, tiempo de respuesta y coherencia semántica del servicio, confirmando el impacto positivo del prototipo en las métricas críticas identificadas en el diagnóstico inicia

Tabla 4

Indicadores obtenidos en la validación del prototipo

Indicador de desempeño	Valor histórico promedio	Resultado del prototipo (simulado)	Variación / Mejora estimada
Tiempo promedio de respuesta a consultas internas	45 segundos	**3.8 segundos**	**−91.6 %**
Nivel de coherencia y precisión de respuestas (escala 1–5)	3.7	**4.8**	**+29.7 %**
Consultas al supervisor por turno	18	**6**	**−66.6 %**
Tasa de error en respuestas o referencias	7.2 %	**1.9 %**	**−73.6 %**

Nota. Elaboración propia

Los resultados derivados de las simulaciones controladas confirman mejoras sustanciales tanto en la eficiencia operativa como en la calidad de las sugerencias generadas por el prototipo implementado. La validación se sustentó en métricas comparativas obtenidas directamente de los registros históricos del centro de contacto y de las ejecuciones reproducibles del flujo completo del sistema.

El tiempo promedio de resolución de consultas internas —establecido en 45 segundos sin asistencia de IA a partir del análisis de logs operativos y observaciones in situ— se redujo a 3,8 segundos durante las pruebas. Esta ganancia se atribuye a la automatización del pipeline de búsqueda semántica y generación de respuestas mediante la integración sincronizada de Whisper y GPT-4, lo que representa una mejora del 91,6 % en la velocidad de acceso a la información requerida.

En relación con la precisión y coherencia semántica, las respuestas sugeridas por el modelo obtuvieron una puntuación media de 4,8 puntos sobre 5 en la evaluación manual realizada por agentes experimentados, lo que valida la efectividad de la arquitectura RAG implementada y la calidad de los embeddings almacenados en la base vectorial.

Un indicador igualmente relevante fue la reducción en la dependencia de los supervisores: las consultas de apoyo por turno disminuyeron de un promedio histórico de 18 a solo 6, equivalente a una caída del 66 %. Este descenso incrementa la autonomía del agente, optimiza la carga de los niveles de escalamiento y permite una redistribución más eficiente de los recursos humanos de gestión.

En síntesis, los datos obtenidos en el entorno de simulación demuestran que el prototipo genera un impacto cuantificable y positivo en la eficiencia del proceso, en la consistencia de la información entregada al cliente y en la experiencia operativa del agente. Estos resultados ratifican la viabilidad técnica de la solución desarrollada y sustentan su potencial para elevar la capacidad operativa del call center, estandarizar la atención bajo los protocolos corporativos de Domicity S.A.S. y sentar las bases para una posterior implementación en producción

9. Conclusiones

El desarrollo del prototipo de asistente inteligente para operadores del call center de Domicity S.A.S. demostró empíricamente el potencial transformador que poseen la inteligencia artificial y las técnicas de procesamiento de lenguaje natural en la optimización de procesos críticos de atención al cliente.

El diagnóstico organizacional inicial identificó como principales cuellos de botella la gestión ineficiente del conocimiento interno y la elevada dependencia de los supervisores, factores que incidían negativamente tanto en los indicadores de productividad como en la satisfacción laboral de los agentes. La arquitectura modular de cinco capas basada en Whisper y GPT-4 emergió como la solución técnica más adecuada para mitigar estas limitaciones, ofreciendo un diseño escalable, seguro y plenamente integrable con la infraestructura existente de la empresa.

Las pruebas realizadas en entorno de simulación arrojaron mejoras cuantificables en los indicadores clave:

- El tiempo medio requerido para resolver consultas internas se redujo de 45 segundos a 3,8 segundos, lo que equivale a una ganancia de eficiencia del 91,6 %.
- Las sugerencias generadas por el modelo obtuvieron una calificación promedio de 4,8 puntos sobre 5 en precisión y coherencia semántica, validando la efectividad del enfoque RAG implementado.
- Las consultas dirigidas a supervisores por turno descendieron un 66 %, incrementando la autonomía operativa de los agentes y optimizando la utilización de los niveles de escalamiento.

Estos resultados confirman la viabilidad técnica y operativa del prototipo, al tiempo que demuestran su alineación directa con los objetivos estratégicos de Domicity S.A.S. en materia de eficiencia, transformación digital y mejora continua de la experiencia del cliente.

El proyecto pone de manifiesto que la integración de inteligencia artificial no implica sustitución del factor humano, sino su potenciación: al automatizar tareas repetitivas de búsqueda y consulta, los operadores pueden enfocarse en interacciones de mayor complejidad y valor emocional, consolidando un modelo híbrido hombre-máquina que representa el estándar futuro para centros de contacto.

En conclusión, el agente de IA desarrollado constituye un avance tecnológico significativo para Domicity S.A.S. y un referente replicable para otras organizaciones del sector que persigan la optimización de sus procesos de atención al cliente. Como líneas futuras de trabajo, se recomienda avanzar hacia una fase piloto en producción con datos reales, incorporar capacidades predictivas basadas en análisis de patrones históricos y ampliar progresivamente la base de conocimiento vectorial con nuevos dominios y fuentes documentales, asegurando así la evolución continua del sistema y la entrega de un servicio cada vez más preciso y personalizado

10. Referencias

Aivo. (2023). *Cómo la IA y los chatbots ayudan a mejorar la experiencia del cliente*. Aivo. https://es.aivo.co/blog/como-la-inteligencia-artificial-y-los-chatbots-ayudan-a-mejorar-la-experiencia-del-cliente

Guamán Tacuri, J. D., Solís Barrionuevo, A. P., & Labre Tixe, W. J. (2025). *El uso de chatbots cognitivos en la satisfacción del consumidor de cadenas de restaurantes*. *Ciencia y Reflexión, 4*(2), 253–270. https://cienciayreflexion.org/index.php/Revista/article/view/327

Amazon Web Services. (2024). *Precios de instancias EC2*. AWS. https://aws.amazon.com/es/ec2/pricing/on-demand/

Daniel Innerarity. (2025). *Una teoría crítica de la inteligencia artificial*. Galaxia Gutenberg. https://www.galaxiagutenberg.com/wp-content/uploads/2025/03/1er-cap.-Una-teoria-critica-IA_.pdf

Asociación CEX. (2023, junio 8). *Situación del Contact Center en 2022*. Interempresas. https://www.interempresas.net/TIC/Articulos/482175-Situacion-del-Contact-Center-en-2022.html

WiseCX. (2023). *¿Qué es el procesamiento de lenguaje natural y cómo puede ayudar a tu empresa?* WiseCX. https://wisecx.com/es/que-es-el-procesamiento-de-lenguaje-natural/

DAIL CX Technologies. (2024). *El procesamiento de lenguaje natural en los call centers modernos.* DAIL CX Technologies. https://www.dail.es/procesamiento-de-lenguaje-natural-call-centers/

Davenport, T. H., & Ronanki, R. (2018). *Artificial intelligence for the real world*. Harvard Business Review, 96(1), 108-116. https://hbr.org/2018/01/artificial-intelligence-for-the-real-world

Carlos, C. E., Jiménez, N., Carlos A.J & Granados, R. D. (2025). *Gestión de la comunicación interna en el sector público: una revisión de literatura (2020-2024)*. Revista Espacios, 46(5), Art. 29. https://www.revistaespacios.com/a25v46n05/25460529.htm

Domicity SAS. (2025). *Informe interno de indicadores operativos de atención, PQRs y tiempos de consulta* [Documento interno no publicado].

Elastic N.V. (2023). *¿Qué es la recuperación de información?* | *Una guía completa de la recuperación de información (IR).* https://www.elastic.co/es/what-is/information-retrieval

Pucha-Paucar, F. M., Romero-Fernández, A. J., Fernández-Villacres, G. E., & Becerra-Arévalo, N. (2024). *Sistemas de información en la gestión de las relaciones con el cliente*. Ingenium et Potentia, 5(1).

https://fundacionkoinonia.com.ve/ojs/index.php/ingeniumetpotentia/article/view/2631

AEERC (Asociación Española de Expertos en la Relación con Clientes). (2024, 25 de octubre). *Tendencias de contact center: la experiencia de cliente y agentes como motor de crecimiento*. Recuperado de https://aeerc.com/tendencias-de-contact-center-la-experiencia-de-cliente-y-agentes-como-motor-de-crecimiento/

Harvard Deusto (2022). *"Chatbots", un recurso al alza en el customer service*. Harvard Deusto Marketing y Ventas. Recuperado de https://www.harvard-deusto.com/chatbots-un-recurso-al-alza-en-el-customer-service

Natera, N. C. (2023). *Impacto de las tecnologías emergentes en el ámbito laboral*. RET Revista de Estudios del Trabajo, 2(3). https://publishing.fgu-edu.com/ojs/index.php/RET/article/view/396

Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, M. P. (2014). *Metodología de la investigación* (6.ª ed.). McGraw-Hill Education. Recuperado de https://apiperiodico.jalisco.gob.mx/api/sites/periodicooficial.jalisco.gob.mx/files/metodologia_de_la_investigacion_-_roberto_hernandez_sampieri.pdf.

OpenAI. (2025, 29 de octubre). *Informe técnico: Evaluación del rendimiento y los valores de referencia de gpt-oss-safeguard-120b y gpt-oss-safeguard-20b*. https://openai.com/es-ES/index/gpt-oss-safeguard-technical-report/

Raji, I. D., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 429–435. https://doi.org/10.1145/3306618.3314244

Celi-Parraga, R. J., Varela-Tapia, E. A., Acosta-Guzmán, I. L., & Montaño-Pulzara, N. R. (2021). Técnicas de procesamiento de lenguaje natural en la inteligencia artificial conversacional textual. *AlfaPublicaciones*, 3(4.1), 40-52. https://doi.org/10.33262/ap.v3i4.1.123

Suárez Mazo, D., & Granados Villa, M. F. (2024, junio 16). *Factores determinantes para la implementación exitosa de la innovación en pymes de Bogotá* (Tesis de Especialización). Universidad EAN. http://hdl.handle.net/10882/13732

Yaranga Vite & Olórtiga Cóndor (2025). *Inteligencia artificial para aumentar la productividad en las empresas: un estudio bibliométrico*. Revista InveCom, 5(4), e504063. https://doi.org/10.5281/zenodo.14846656

Garzón Quiroz, Del Campo Saltos & Loor Ávila (2025). Análisis sistemático sobre la eficiencia comunicativa entre chatbots basados en reglas y modelos de lenguaje natural. *Universitas XXI, Revista de Ciencias Sociales y Humanas*, 42, 167-192. https://doi.org/10.17163/uni.n42.2025.07

NTT DATA. (2024, 12 julio). *Inteligencia artificial y el futuro de los contact centers*. NTT DATA. https://es.nttdata.com/insights/blog/contact-centers-futuro-ia

FiumiConnect. (2024). *Cómo la inteligencia artificial mejora la experiencia del cliente en los call centers*. https://fiumiconnect.com/como-la-inteligencia-artificial-mejora-la-experiencia-del-cliente-en-los-call-centers/#:~:text=La%20inteligencia%20artificial%20puede%20analizar,su%20satisfacci%C3%B3n%20en%20cada%20interacci%C3%B3n.

Zendesk. (2024, febrero). *¿Qué es un centro de llamadas? Definición, tipos y funcionamiento*. Zendesk. https://www.zendesk.es/blog/ultimate-guide-call-centers/

Narváez Trejo, Ó. M., & Villegas Salas, L. I. (2014). *Introducción a la investigación: guía interactiva*. Xalapa, Ver., México: Universidad Veracruzana. Recuperado de https://www.uv.mx/apps/bdh/investigacion/introduccion.html

Moench, E. I. (2023). *Los estudios laborales latinoamericanos sobre los agentes de Contact Centers: bibliografía, notas críticas y cuestiones abiertas con la irrupción del teletrabajo.* Revista Latinoamericana de Metodología de las Ciencias Sociales, 13(1), e125. https://doi.org/10.24215/18537863e125