Digital Object Identifier: 10.24054/rcta.v2i46.3530

Sistema multiagente enfocado en procesos de inteligencia artificial distribuida para la captura y el ruteo de pedidos

Multi-agent system focused on distributed artificial intelligence processes for order capture and routing

MSc. Anny Astrid Espitia Cubillos [1], PhD. Robinson Jiménez Moreno [2], Ing. Mateo Andrés Pulido Aponte [1]

[1] ***Universidad Militar Nueva Granada,*** *Facultad de Ingeniería, Programa de Ingeniería Industrial, Bogotá, Colombia*
[2] ***Universidad Militar Nueva Granada,*** *Facultad de Ingeniería, Programa de Ingeniería Mecatrónica, Bogotá, Colombia.*

Correspondencia: *anny.espitia@unimilitar.edu.co*

Recibido: *20 enero 2025.* ***Revisado:*** *20 mayo 2025.* ***Aceptado:*** *21 junio 2025.* ***Publicado:*** *01 julio 2025.*

Cómo citar: *A. A. Espitia Cubillos, R. Jiménez Moreno, y M. A. Pulido Aponte, «Sistema multiagente enfocado en procesos de inteligencia artificial distribuida para la captura y el ruteo de pedidos», RCTA, vol. 2, n.º 46, pp. 13–21, jul. 2025. Recuperado de https://ojs.unipamplona.edu.co/index.php/rcta/article/view/3530*

Esta obra está bajo una licencia internacional
Creative Commons Atribución-NoComercial 4.0.

Resumen: Este documento presenta el diseño de un sistema multiagente enfocado en procesos de inteligencia artificial distribuida operado por texto que funciona siguiendo un modelo de dialogo, orientado a la gestión eficiente de órdenes de pedido de productos industriales, éste sistema integra cinco agentes de razonamiento y acción: el agente general potenciado con el modelo de lenguaje de inteligencia artificial Gemini de Google que lleva el flujo de la conversación y redirecciona las tareas a los otros agentes, un agente que identifica clientes, un agente que reconoce los productos, un agente que genera los pedidos y un agente que rutea. El sistema multiagente está diseñado para interactuar de manera virtual y conversacional con los usuarios, facilitando la creación y gestión de pedidos mediante un enfoque innovador que integra tecnologías avanzadas de procesamiento natural del lenguaje, bases de datos vectoriales y relacionales, y métodos de optimización. Finalmente, para su validación en un entorno real, se utilizan datos de una empresa de producción y distribución de productos de aseo, lo que permite desarrollar distintas pruebas para la identificación de clientes, productos, cantidades y establecimiento de rutas de entrega de pedidos. Se concluye que el flujo conversacional y técnicas utilizadas permite a los usuarios realizar consultas sobre clientes con una exactitud promedio del 77,38% y de productos con una precisión promedio del 88.57%, aún en escenarios con ambigüedades semánticas, logrando gestionar órdenes de manera intuitiva. Se logra además optimizar el ruteo de los pedidos considerando simultáneamente dos criterios que pueden ponderarse: importancia del cliente y distancia recorrida.

Palabras clave: aplicación, agente de razonamiento y acción, optimización, bases de datos, ruteo, inteligencia artificial, sistema multiagente.

Abstract: This paper presents the design of a multi-agent system focused on text-operated distributed artificial intelligence processes that works following a dialog model, oriented to the efficient management of industrial product orders. This system integrates five reasoning and action agents: the general agent powered by Google's Gemini artificial intelligence language model that carries the conversation flow and redirects the tasks to the other agents, an agent that

identifies customers, an agent that recognizes the products, an agent that generates the orders and an agent that routes. The multi-agent system is designed to interact virtually and conversationally with users, facilitating order creation and management through an innovative approach that integrates advanced natural language processing technologies, vector and relational databases, and optimization methods. Finally, for its validation in a real environment, data from a company that produces and distributes cleaning products are used, allowing the development of different tests for the identification of customers, products, quantities and the establishment of order delivery routes. It is concluded that the conversational flow and techniques used allow users to make queries about customers with an average accuracy of 77.38% and about products with an average accuracy of 88.57%, even in scenarios with semantic ambiguities, managing orders in an intuitive way. It is also possible to optimize order routing by simultaneously considering two criteria that can be weighted: customer importance and distance traveled.

Keywords: app, reasoning and action agent, optimization, databases, routing, artificial intelligence, multi-agent system.

1. INTRODUCCIÓN

En el sector industrial las empresas necesitan operar con eficiencia, para ello se requiere la mejora continua de los procesos incluyendo el uso de tecnologías que reduzcan la carga de trabajo de los empleados y les permitan concentrarse en las tareas que más agregan valor y mejorar la capacidad de respuesta al cliente, enfrentar adecuadamente estos dos retos son un factor de éxito empresarial.

Los sistemas de tecnología de la información se desarrollaron para responder a la creciente complejidad de los procesos [1]. Los chatbots como sistemas tecnológicos permiten la comunicación de las personas y se constituyen como una interfaz de usuario empresa. Los agentes de chat son una evolución de los chatbots.

Actualmente, en la industria se usan chatbots para la entrada de datos en producción [2], que pueden integrarse con inteligencia artificial para reunir información rápidamente que facilite y soporte la toma de decisiones en fabricación [3]. En logística constituyen una de las tecnologías fundamentales aplicadas para la interacción hombre-máquina [4]. También en el sector servicios los chatbots ayudan a los usuarios a programar citas, hacer reservas y establecer recordatorios [5], en general, permiten orientar y ayudar a usuarios internos (trabajadores) y externos (clientes) [6].

La industria 4.0 busca obtener productos inteligentes, mediante la fabricación, la gestión de la cadena de suministro y el manejo a cliente, aspectos desarrollados por sistemas inteligentes. Adicional a ello la Industria 5.0. involucra el uso de sistemas centrados en el ser humano [7], lo que ha conducido a las organizaciones a buscar la integración exitosa de las tecnologías digitales con las personas.

Por ello es posible entender un sistema multiagente enfocado en procesos de inteligencia artificial distribuida [8] como una herramienta digital que facilita la relación con el cliente como en el caso de aplicación del presente artículo, en donde se usa para la captura virtual y el ruteo de pedidos. Como lo proponen en [9] como una herramienta que ayuda a los trabajadores a realizar tareas y tomar decisiones. Un chatbot tradicional responde a una estructura de árbol, que guía una interacción por respuestas numéricas u opciones simples por respuestas tipo si o no.

En [10] se diseña un agente logístico tipo chat para gestionar el transporte de material que se encarga de entregar la información al programador responsable de elegir el vehículo adecuado. También en [11] se presenta un diseño de un sistema de gestión de operaciones de fabricación que incluye una interfaz de chatbot soportada en inteligencia artificial con un sistema de predicción que recupera información en vivo de la base de datos sobre la producción.

Sin embargo, [5] postula que usando chatbots y asistentes virtuales los desarrollos de innovación ciencia y tecnología cuentan con el potencial de suplantar a los empleados. Pero el enfoque aquí expuesto coincide con el de [12] en el sentido rescatar el potencial del trabajo colaborativo entre humanos e IA, en el del marco de la Industria 5.0 pensando en un desarrollo centrado en el ser humano.

Los modelos de lenguaje grandes (LLM en inglés) funcionan bien en tareas de propósito general, pero presentan fallas en áreas específicas que se

actualizan. En [13] integran la generación aumentada por recuperación (RAG) con agentes de razonamiento y acción (ReAct por sus siglas en inglés) con el objeto de evitar fallas en los LLM al generar diseños biomiméticos.

Basado en el análisis del estado del arte, este artículo tiene como objetivo exponer el diseño de un sistema multiagente enfocado en procesos de inteligencia artificial cuyo aporte consiste en orientar de forma automática la toma de productos y su ruteo para el despacho de pedidos, basado solo en la interacción cliente-agente. Su validación se lleva a cabo mediante la aplicación a una empresa de producción y distribución de productos de aseo.

El documento está estructurado en cinco secciones: la primera corresponde a la introducción, en la segunda parte se presenta la metodología, en la tercera se expone el desarrollo, en la cuarta se reportan los principales resultados y en la última se exponen las conclusiones alcanzadas.

2. METODOLOGÍA

Para el desarrollo del Sistema multiagente enfocado en procesos de inteligencia artificial distribuida que permite la captura y posterior ruteo de pedidos se definieron cinco fases.

En la primera fase, se selecciona el agente de razonamiento y acción (Re-Act) [14] que, para reducir los posibles fallos en el LLM, solicita al usuario elegir una acción de una lista de opciones. El agente se potencia con el modelo de inteligencia artificial Gemini de Google [15], que proporciona capacidades avanzadas para gestionar flujos conversacionales complejos, mejorando la experiencia del usuario en la creación de órdenes de pedido.

En la segunda fase se usan bases de datos vectoriales [16] con dos propósitos: realizar búsquedas eficientes de los nombres de los clientes e identificar productos. En este caso utilizando la base de datos vectorial de código abierto Chroma [17], se almacenan representaciones vectoriales de los nombres de los clientes y de los nombres y descripciones de los productos, esto permite búsquedas aproximadas que son robustas frente a errores ortográficos o variaciones en la consulta. Adicionalmente en el caso de los productos permite hacer sugerencias de artículos similares para facilitar al usuario la selección de la opción de su preferencia. El enfoque de las bases de datos vectoriales mejora la precisión y eficiencia de las búsquedas, superando las limitaciones de sistemas tradicionales basados en texto exacto.

En la tercera fase se utiliza PostgreSQL [18] para gestionar bases de datos relacionales del inventario de cada producto, los pedidos, los clientes e incluso las rutas logísticas asociadas al sistema. Este componente asegura la integridad y consistencia de la información operativa. PostgreSQL permite el funcionamiento del sistema relacional para:

- Gestión de inventarios: Registrar y monitorear el estado de los productos.
- Control de clientes y pedidos: Mantener un historial estructurado y confiable.
- Cálculo de rutas: Integrar la matriz de distancias y las órdenes de pedido para generar rutas optimizadas.

Se emplea LlamaIndex [19] para integrar las bases de datos vectoriales y relacionales con el modelo de IA Gemini [15], esto hace que las consultas de los usuarios se traduzcan en acciones concretas, como buscar productos, verificar inventarios o consultar datos de clientes, de forma rápida y eficiente.

En la cuarta fase se lleva a cabo la optimización de rutas dado que la entrega de los pedidos depende de la definición de dichas rutas, para ello se utiliza una matriz de distancias que resume las posibles conexiones entre los puntos de origen y destino. El cálculo de rutas considera simultáneamente dos criterios que pueden ponderarse: importancia del cliente (prioridad del pedido) y distancia recorrida, utilizando mapas de Kohonen [20] para encontrar soluciones de costo mínimo. Este enfoque garantiza la selección de rutas óptimas, que permiten maximizar satisfacción de los clientes según su relevancia, minimizar los tiempos de entrega, o satisfacer de manera simultánea estos dos objetivos de manera ponderada.

Finalmente, en la quinta fase se usa el marco Streamlit [21] para crear una interfaz amigable que permita a los usuarios interactuar fácilmente con el sistema multiagente conversacional, visualizar datos relevantes, como recomendaciones de productos, rutas logísticas y tiempos estimados de entrega. Además de mostrar en tiempo real el estado de los inventarios, las órdenes de pedido y las rutas óptimas generadas.

3. DESARROLLO

Según la Ley de IA de la Comisión Europea los chatbots de bajo riesgo están obligados a cumplir

con las obligaciones de transparencia [22], por ello el desarrollo de la solución se centró en la integración de múltiples tecnologías avanzadas para construir un sistema conversacional robusto y eficiente, cubierto en el marco de la mencionada ley. Como se mencionó se implementó un agente ReAct [14] potenciado por el modelo de lenguaje Gemini de Google [15], orientando su implementación a la interacción natural con los usuarios para gestionar órdenes de pedido.

El sistema multiagente diseñado integra cinco agentes Re-Act [14]: el agente general potenciado con el modelo de lenguaje de inteligencia artificial Gemini de Google que lleva el flujo de la conversación y redirecciona las tareas a los otros agentes, un agente que identifica clientes usando la base de datos vectorial Chroma [17], un agente que reconoce los productos también mediante la base de datos vectorial Chroma, un agente que gestiona bases de datos relacionales del inventario usando PostgreSQL [18] para generar los pedidos y llevar el historial y un agente que rutea utilizando mapas de Kohonen [20]. Se usa Streamlit para crear la interfaz con los usuarios. En la Fig. 1. se presenta la estructuración del sistema multiagente y las distintas herramientas que soportan su funcionamiento.

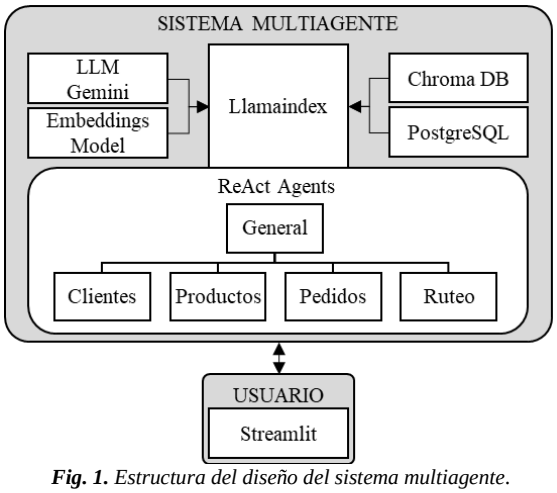

Fig. 1. *Estructura del diseño del sistema multiagente.*

El sistema multiagente enfocado en procesos de inteligencia artificial distribuida para la captura y el ruteo de pedidos funciona basado en el diagrama de flujo mostrado en la Fig. 2. El proceso representado en el diagrama de flujo le permite al multiagente manejar el rumbo de la conversación completamente, evaluando los datos y la información con la que cuenta para redirigirlo si es necesario y solicitar la información pertinente al usuario.

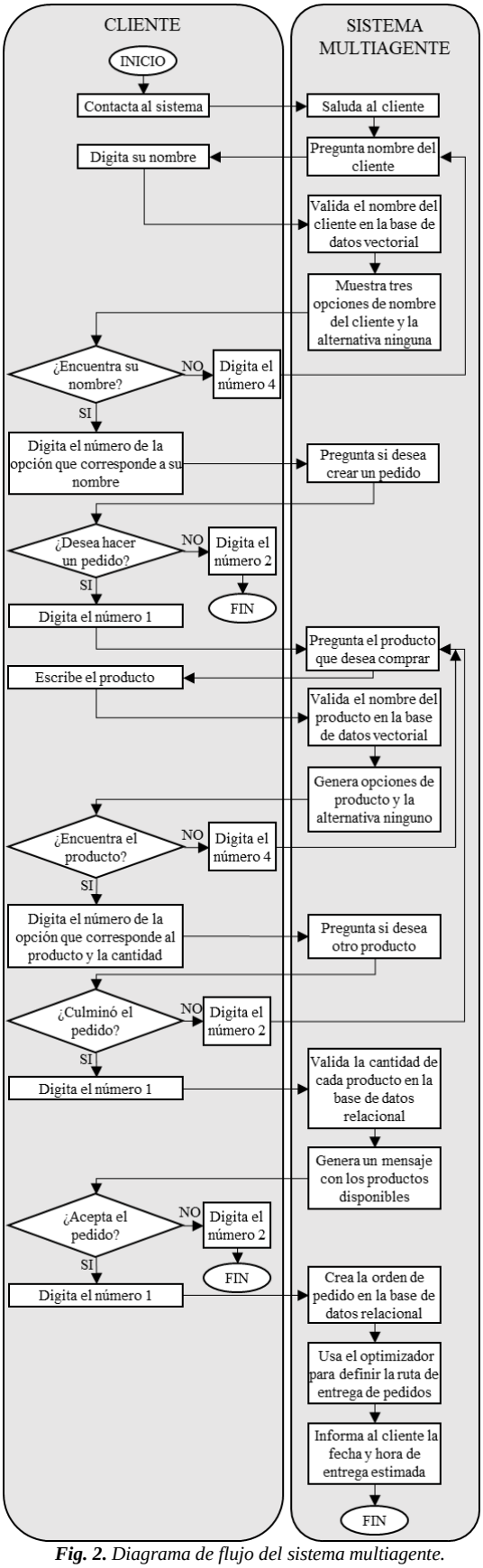

Fig. 2. *Diagrama de flujo del sistema multiagente.*

Respecto a las bases de datos vectoriales usando Chroma [17], se realizó un preprocesamiento principalmente en los datos de los clientes, eliminando palabras recurrentes como: Conjunto, Edificio, Multifamiliar, etc, con el objetivo de mejorar los resultados al realizar las búsquedas semánticas con la entrada del usuario. El mismo preprocesamiento se realizó con los datos de los productos.

Para su implementación se requirió almacenar los datos de productos, clientes, distancias, pedidos, y rutas en bases de datos relacionales del entorno de validación, para mantener una exactitud y manejo adecuado de los datos. Los campos que contienen las tablas en cada una de las bases de datos y su tipo se exponen en la Fig. 3.

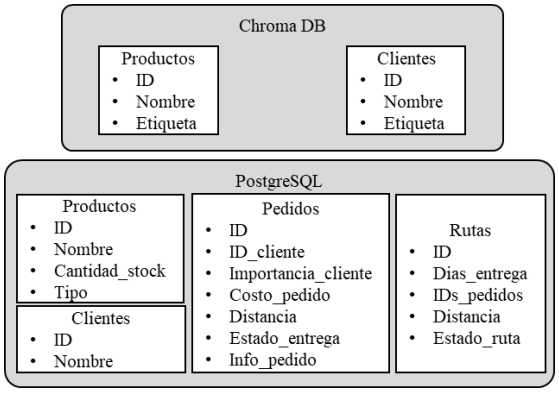

Fig. 3. *Información de las bases de datos.*

La combinación de datos relacionales y vectoriales permite que el sistema multiagente enfocado en procesos de inteligencia artificial distribuida para la captura y el ruteo de pedidos sea versátil y efectivo en la toma de decisiones.

El cálculo de rutas se realizó mediante el uso del algoritmo de mapas de Kohonen [20] debido a su enfoque de aprendizaje competitivo y cooperativo formando relaciones entre las distancias de los sitios a definir la ruta, además a través de la normalización de las distancias se puede incluir otro factor de peso para el cálculo eficiente de las rutas, como lo es la importancia, dejando un valor variable de ponderación de 0 a 1 entre el criterio distancia y el criterio importancia para evaluar los resultados obtenidos. A continuación, se presentan las 10 ecuaciones usadas para el cálculo de distancia.

- Vector dinámico de entrada:

$$S_0 = [0, d_{01}, d_{02}, \dots, d_{0n}]$$
$$S_1 = [d_{10}, 0, d_{12}, \dots, d_{1n}]$$
$$S_2 = [d_{20}, d_{21}, 0, \dots, d_{2n}]$$
$$S_n = [d_{n0}, d_{n1}, d_{n2}, \dots, 0]$$

Donde S_i representa el vector que agrupa los datos de las distancias (d_{ij}) desde el origen i hasta el destino j.

- Matriz dinámica de pesos:

$$M = \begin{bmatrix} w_{00} & w_{01} & \cdots & w_{0n} \\ \vdots & & \ddots & \vdots \\ w_{n0} & w_{n1} & \cdots & w_{nn} \end{bmatrix}$$

Donde W_{ij} corresponde al peso del origen i al destino j.

- Escalado de distancias incluyendo el factor importancia.

$$d_s = \left(\frac{d_s - d_{min}}{d_{max} - d_{min}} \times W_d\right) + \left(\frac{I_s - I_{min}}{I_{max} - I_{min}} \times W_i\right)$$

Donde W_d representa la ponderación del criterio distancia recorrida y W_i la ponderación de la importancia del cliente, es importante tener presente que se debe garantizar que en todo caso se cumpla con:

$$\sum W_d + W_i = 1$$

- Distancia Euclidiana:

$$d_{euc} = \sqrt[2]{\sum (x_{i0} - x_{i1})^2}$$

Donde X_{ij} representa la ubicación de cada cliente.

- Mejor Unidad Coincidente MUC:

$$MUC = min([d_{euc0}, d_{euc1}, \dots, d_{eucn}])$$

- Radio de actualización de pesos:

$$r(t) = r_o e^{-\frac{t}{\lambda}}$$

Donde t representa el periodo de tiempo.

- Factor de producto:

$$\beta_t = e^{-\frac{d^2}{2r_t^2}}$$

- Tasas de aprendizaje:

$$\lambda_0 = \frac{T}{\ln(ro)} \approx 0.5$$

$$\lambda(t) = \lambda_o e^{-\frac{t}{1000}}$$

- Actualización de pesos:

$$M_t = M_{t-1} + \lambda_t \times \beta_t (d_{si} - M_{t-1})$$

En la Fig. 4. se puede observar el diseño de la interfaz desarrollada desde la óptica del cliente (usuario externo).

Fig. 4. *Interfaz usuario.*

Finalmente, en la Fig. 5. se presenta el registro de las operaciones que puede ser consultado en la interfaz por parte del asesor comercial (usuario interno).

Fig. 5. *Interfaz de operaciones.*

4. RESULTADOS

4.1. Validación identificación de clientes

Para la validación de la identificación de los clientes se generó una base de datos con información de los clientes que fue modificada para incluir tres escenarios: uno con la información completa y correcta, uno con información incompleta y otro con información mal digitada, que corresponden a las imprecisiones que los diseñadores consideran pueden presentarse por parte de los clientes.

Tras realizar la búsqueda semántica del nombre de distintos clientes en el Sistema multiagente enfocado en procesos de inteligencia artificial distribuida para la captura y el ruteo de pedidos, se obtienen los resultados que se presentan en la Tabla 1.

Tabla 1: *Resultados pruebas identificación de clientes*

Información ingresada	Porcentaje de acierto
Nombre completo	100,00%
Nombre incompleto	57,14%
Nombre mal digitado	75,00%
Promedio	77,38%

De la Tabla 1 se destaca que la identificación de clientes en el primer escenario obtuvo el resultado esperado que corresponde a un acierto del 100%. Cuando se realiza la búsqueda semántica de nombre de clientes que no estaban completos se acertó en el 57.14% de los casos. Mientras que la búsqueda semántica de nombre de clientes ingresados por el usuario con errores de digitación se obtiene un 75% de exactitud. Este permite calcular un porcentaje de acierto por promedio simple de los tres escenarios evaluados en la prueba del 77,38%.

Para mejorar los resultados es recomendable etiquetar mejor los nombres de los clientes, usar un modelo de incrustaciones de palabras [23] con mayores dimensiones y/o el aprendizaje por refuerzo [24] para el entrenamiento de la inteligencia artificial.

4.2. Validación identificación de productos

Para la validación de la identificación de los productos se aplicó el mismo procedimiento que para la identificación de clientes, en el cual se modificó la base de datos con información de los productos para contar con los mismos tres escenarios: uno con información completa y correcta, uno con información incompleta y otro con información mal digitada. En la tabla 2 se presentan

los resultados obtenidos para las pruebas de productos en los tres escenarios considerados.

Tabla 2: *Resultados pruebas identificación de productos*

Información ingresada	Porcentaje de acierto
Nombre completo	100,00%
Nombre incompleto	65,71%
Nombre mal digitado	100,00%
Promedio	88,57%

Como se observa en la Tabla 2 en el primer escenario nuevamente se logró un porcentaje de acierto del 100%, y también se logró ese mismo resultado cuando el usuario ingresó los datos mal digitados. Solo se presentaron fallas cuando la información de los productos no estaba completa, es este escenario se obtiene un 65,71% de éxito. El porcentaje de acierto por promedio simple de los tres escenarios evaluados en la prueba del 88,57%.

Se puede inferir también que con un mejor etiquetado y un modelo con mayores dimensiones de incrustaciones de palabras [23] podría mejorar la exactitud de la identificación cuando la información ingresada por el usuario es incompleta, además es posible usar técnicas de RAG [13] para realizar varias búsquedas por sinónimo de los productos.

4.2. Validación cálculo de rutas

Para la validación del cálculo de las rutas óptimas se consideraron seis escenarios usando los mapas de Kohonen con distintas ponderaciones de los dos criterios de decisión establecidos: importancia del cliente y distancia recorrida). En la tabla 3 se relacionan los resultados obtenidos del cálculo de rutas.

Tabla 3: *Distancia promedio recorrida usando distintas ponderaciones de los criterios*

Ponderación		Distancia recorrida (Km)			
Importancia	Distancia	Día 1	Día 2	Día 3	Promedio
0%	100%	78,90	88,94	152,55	106,795
100%	0%	96,36	101,27	199,24	132,288
60%	40%	63,99	84,47	153,22	100,560
40%	60%	92,13	91,23	153,98	112,444
80%	20%	96,66	82,29	153,22	110,724
20%	80%	78,90	78,68	156,44	104,672

Como se puede observar en la tabla 3, el escenario con mejores resultados se obtiene cuando se da una ponderación del 40% a la distancia total recorrida y del 60% a la importancia del cliente, que presenta valores menores de distancia y posiciona los sitios más importantes al inicio de la ruta.

La Fig. 6. presenta un ejemplo de la ruta inicial sin optimizar, los números indican el orden de entrega de los pedidos, esta recorre una distancia total de 26.4 Km.

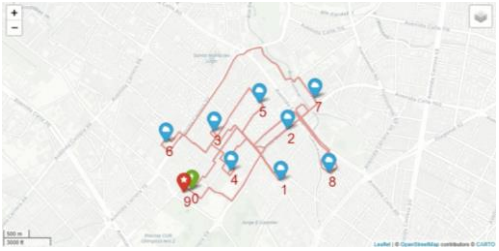

Fig. 6. *Ruta inicial*

En la Fig. 7., se incluyen imágenes que muestran cómo, durante algunas de las 5000 épocas que dura el entrenamiento, las neuronas se van auto organizando en los distintos sitios de entrega teniendo en cuenta las distancias con coordenadas de longitud y latitud.

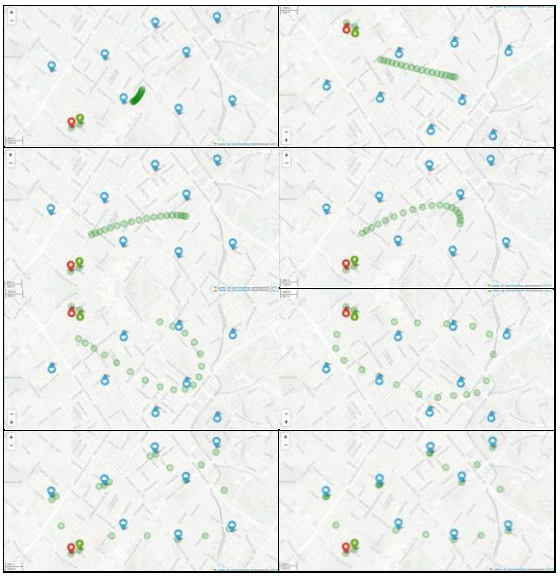

Fig. 7. *Auto organización de ruta*

En la Fig. 8, se presenta la ruta optimizada con una distancia total de 17.1 Km, con una reducción del 35,23% de recorrido.

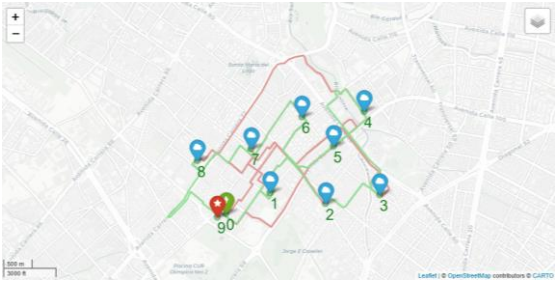

Fig. 8. *Ruta final optimizada*

Es importante señalar que en las Fig. 6 a 8, se tiene un error de hasta un 1 Km en la ubicación, ese margen de error está dado debido a que las distancias se calculan usando las latitudes y longitudes de cada sitio de entrega, pero no es tenido en cuenta la distancia real que se obtendría al conocer con exactitud las calles que permiten llegar a los puntos de entrega.

5. CONCLUSIONES

Se logró evidenciar que las bases de datos vectoriales permiten identificar de manera aceptable (77,38%) los clientes aún en escenarios con ambigüedades, es decir que el desarrollo opera con un margen de errores en la digitación de búsqueda, lo cual es favorable. También permiten identificar productos de forma adecuada (88,57%) lo que sirve para recomendar al usuario opciones similares según las preferencias. Chroma mejora la experiencia del usuario al buscar precisión y rapidez en las búsquedas, incluso en consultas imprecisas o mal redactadas.

La integración de tecnologías avanzadas de procesamiento del lenguaje natural, bases de datos vectoriales y relacionales, y métodos de optimización de rutas proporciono un diseño del Sistema multiagente enfocado en procesos de inteligencia artificial distribuida que se considera robusto y eficiente para la gestión de pedidos, en las pruebas realizadas. A su vez se evidencio que es capaz de adaptarse a las necesidades del usuario y optimizar procesos logísticos, mejorando la experiencia del cliente y reduciendo costos operativos asociados con la toma y entrega final de pedidos.

Como trabajo futuro se propone considerar la interacción vía voz para facilitar la conversación entre el usuario y el agente de chat inteligente. Mejorar la identificación de clientes y productos cuando la información ingresada por el cliente es incompleta, por técnicas como el aprendizaje por refuerzo, incrustaciones de palabras y RAG. También tener presentes otros criterios de optimización para el establecimiento de las rutas de entrega de pedidos.

AGRADECIMIENTOS

Producto derivado del proyecto de investigación titulado “Mejoramiento de procesos de control de inventario y alistamiento de pedidos en el marco de la industria 4.0" INV-ING-3970 financiado por la vicerrectoría de investigaciones de la Universidad Militar Nueva Granada, vigencia 2024.

REFERENCIAS

[1] A. Börold, D. Schweers y M. Freitag, «Towards Multimodal Information Systems for Assisting Humans in Production and Logistics Processes,» *Procedia CIRP*, vol. 120, pp. 1089-1094, 2023. https://doi.org/10.1016/j.procir.2023.09.130.

[2] A. Dörner, P. Gerald, S. Michal y B. Marek, «Use of Lean Management Methods based on Eye-Tracking Information to make User Interfaces in Production more Human-centered,» vol. 128, pp. 514-519, 2024. https://doi.org/10.1016/j.procir.2024.04.014.

[3] S. Vernim, H. Bauer, E. Rauch, M. T. Ziegler y S. Umbrello, «A value sensitive design approach for designing AI-based worker assistance systems in manufacturing,,» *Procedia Computer Science*, vol. 200, pp. 505-516, 2022. https://doi.org/10.1016/j.procs.2022.01.248.

[4] M. Pellicelli, «Chapter Two - A long road to maximizing efficiency,» de *The Digital Transformation of Supply Chain Management*, Elsevier, 2023. https://doi.org/10.1016/B978-0-323-85532-7.00009-8, pp. 37-61.

[5] I. Ghannouchi, «Examining the dynamic nexus between industry 4.0 technologies and sustainable economy: New insights from empirical evidence using GMM estimator across 20 OECD nations,» *Technology in Society*, vol. 75, p. 102408, 2023. https://doi.org/10.1016/j.techsoc.2023.102408

[6] A. Salis, A. Marguglio, G. D. Luca, S. Razzetti, W. Quadrini y S. Gusmeroli, «An Edge-Cloud based Reference Architecture to support cognitive solutions in Process Industry,» *Procedia Computer Science*, vol. 217, pp. 20-30, 2023. https://doi.org/10.1016/j.procs.2022.12.198.

[7] T. A. Aniceski, L. T. Pontes Miranda, O. Canciglieri Junior y G. Brittes Benitez, «The Four Smarts of Industry 4.0 and barriers for technology deployment: A TOE perspective,» *Computers & Industrial Engineering*, vol. 193, p. 110345, 2024. https://doi.org/10.1016/j.cie.2024.110345.

[8] J. S. Dhatterwal, K. S. Kaswan y R. P. Ojha, «Chapter 13. The role of multiagent system in

industry 4.0.,» de *A Roadmap for Enabling Industry 4.0 by Artificial Intelligence*, 2022. https://doi.org/10.1002/9781119905141.ch13, pp. 227-246..

[9] B. G. Mark, E. Rauch y D. T. Matt, «Systematic selection methodology for worker assistance systems in manufacturing,» *Computers & Industrial Engineering*, vol. 166, p. 107982, 2022. https://doi.org/10.1016/j.cie.2022.107982.

[10] G. Bitsch, P. Senjic y J. Askin, «Dynamic adaption in cyber-physical production systems based on ontologies,» *Procedia Computer Science*, vol. 200, pp. 577-584, 2022. https://doi.org/10.1016/j.procs.2022.01.255.

[11] S. Mantravadi, C. Møller, C. LI y R. Schnyder, «Design choices for next-generation IIoT-connected MES/MOM: An empirical study on smart factories,» *Robotics and Computer-Integrated Manufacturing*, vol. 73, p. 102225, 2022. https://doi.org/10.1016/j.rcim.2021.102225.

[12] X. Sun y Y. Song, «Unlocking the Synergy: Increasing productivity through Human-AI collaboration in the industry 5.0 Era,» *Computers & Industrial Engineering*, vol. 200, p. 110657, 2025. https://doi.org/10.1016/j.cie.2024.110657.

[13] C. Toukmaji y A. Tee, «Retrieval-Augmented Generation and LLM Agents for Biomimicry Design Solutions,» de *Proceedings of the AAAI Symposium Series*, 2024. https://doi.org/10.1609/aaaiss.v3i1.31210.

[14] S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. Narasimhan y Y. Cao, «ReAct: Synergizing Reasoning and Acting in Language Models,» 2023. https://doi.org/10.48550/arXiv.2210.03629.

[15] H. R. Saeidnia, «Welcome to the Gemini era: Google DeepMind and the information industry,» *Library Hi Tech News*, 2023. https://doi.org/10.1108/LHTN-12-2023-0214.

[16] Z. Jing, Y. Su y Y. Han, «When large language models meet vector databases: A survey,» 2024. https://doi.org/10.48550/arXiv.2402.01763 .

[17] P. N. Singh, S. Talasila y S. V. Banakar, «Analyzing Embedding Models for Embedding Vectors in Vector Databases,» de *2023 IEEE International Conference on ICT in Business Industry & Government (ICTBIG)*, Indore, India, 2023. https://doi.org/10.1109/ICTBIG59752.2023.10455990.

[18] A. Makris, K. Tserpes, G. Spiliopoulos, D. Zissis y D. Anagnostopoulos, «MongoDB Vs PostgreSQL: A comparative study on performance aspects,» *Geoinformatica*, vol. 25, p. 243–268, 2021. https://doi.org/10.1007/s10707-020-00407-w.

[19] R. K. Malviya, V. Javalkar y R. Malviya, «Scalability and Performance Benchmarking of LangChain, LlamaIndex, and Haystack for Enterprise AI Customer Support Systems,» de *IJGIS Fall of 2024 Conference. The New World Foundation.*, 2024. https://doi.org/10.21428/e90189c8.43aeb06e.

[20] Q. Guan, X. Hong, W. Ke, L. Zhang, G. Sun y Y. Gong, «Kohonen Self-Organizing Map based Route Planning: A Revisit,» de *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Prague, Czech Republic, 2021. https://doi.org/10.1109/IROS51168.2021.9636025.

[21] Y. Akkem, B. S. Kumar y A. Varanasi, «Streamlit Application for Advanced Ensemble Learning Methods in Crop Recommendation Systems – A Review and Implementation,» *Indian Journal of Science and Technology*, vol. 16, nº 48, pp. 4688-4702, 2023. https://doi.org/10.17485/IJST/v16i48.2850.

[22] G. Schuh, M.-F. Stroh y L. Johanning, «Framework To Design Compliance Rules For Digital Technologies In Manufacturing Companies,» *Procedia CIRP*, vol. 119, pp. 103-108, 2023. https://doi.org/10.1016/j.procir.2023.04.005.

[23] M. Habib, M. Faris, A. Alomari y H. Faris, «AltibbiVec: A Word Embedding Model for Medical and Health Applications in the Arabic Language,» *IEEE Access*, vol. 9, pp. 133875-133888, 2021. https://doi.org/10.1109/ACCESS.2021.3115617.

[24] Y. Matsuo, Y. LeCun, M. Sahani, D. Precup, D. Silver, M. Sugiyama, E. Uchibe y J. Morimoto, «Deep learning, reinforcement learning, and world models,» *Neural Networks*, vol. 152, pp. 267-275, 2022. https://doi.org/10.1016/j.neunet.2022.03.037.