



**UNIVERSIDAD
DE ANTIOQUIA**

Agente de IA para asistencia de compras basadas con criterios nutricionales

Paola Andrea Gómez Chantre

Leidy Dayanna Pérez Girón

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesor

Antonio Jesús Tamayo Herrera, Ph.D

Universidad de Antioquia

Facultad de Ingeniería

Especialización en Analítica y Ciencia de Datos

Medellín, Antioquia, Colombia

2026



UNIVERSIDAD DE ANTIOQUIA

Cita

(Gómez Chantre & Pérez Girón, 2026)

Referencia

Gómez Chantre, P. A., & Pérez Girón, L. D. (2026). Agente de IA para asistencia de compras basadas con criterios nutricionales [Trabajo de grado especialización]. Universidad de Antioquia, Medellín, Colombia.

Estilo APA 7 (2020)



Grupo de Investigación Intelligent Information Systems Lab In2Lab
Especialización en Analítica y Ciencia de Datos, Cohorte X.
Centro de Investigación Ambientales y de Ingeniería (CIA).



Centro de Documentación Ingeniería (CENDOI)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

Rector: Héctor Iván García García.

Decano: Elkin Libardo Ríos Ortiz.

Jefe departamento: Danny Alexandro Múnera Ramírez

Coordinadora del Programa: Maria Bernarda Salazar Sánchez

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Tabla de contenido

Resumen	7
Abstract	8
1. Introducción	9
2. Objetivos	11
2.1 Objetivo general	11
2.2 Objetivos específicos.....	11
3. Marco teórico	12
3.1 <i>Sistemas de recomendación de alimentos</i>	12
3.2 <i>Modelos de lenguaje de gran escala y agentes conversacionales</i>	12
3.3 <i>Nutrición personalizada asistida por LLM</i>	13
3.4 <i>Sistemas de recomendación para supermercados en línea</i>	14
3.5 <i>Optimización de canastas saludables y restricciones económicas</i>	15
3.6 <i>Comercio electrónico del sector de alimentos en América Latina y disponibilidad de datos nutricionales</i>	16
4. Metodología	17
4.1 <i>Fuente de Datos y Extracción del Catálogo</i>	17
4.2 <i>Preprocesamiento y Limpieza del Dataset</i>	18
4.3 <i>Análisis Exploratorio de Datos</i>	19
4.3.1 <i>Análisis Exploratorio de Datos</i>	19
4.3.2 <i>Diagnóstico de calidad de datos</i>	20
4.3.3 <i>Distribución de precios y criterios de depuración</i>	21
4.3.4 <i>Síntesis: embudo del pipeline de procesamiento</i>	21

4.4	<i>Enriquecimiento con Información Nutricional</i>	22
4.5	<i>Arquitectura del Agente de Recomendación</i>	23
4.5.1	<i>Etapa 1: Interpretación de la Intención del Usuario (parse_user_intent)</i>	23
4.5.2	<i>Etapa 2: Filtrado del Catálogo (filter_catalog)</i>	24
4.5.3	<i>Etapa 3: Generación de la Canasta (generate_basket)</i>	24
4.6	<i>Herramientas y Entorno de Desarrollo</i>	25
5.	<i>Resultados</i>	26
5.1	<i>Casos de éxito en la implementación del agente</i>	26
5.2	<i>Casos de fallo en la implementación del agente</i>	29
6.	<i>Conclusiones</i>	34
7.	<i>Trabajo futuro</i>	36
	<i>Referencias</i>	37

Lista de figuras

Ilustración 1. Flujo metodológico para la construcción del agente.....	17
Ilustración 2. Composición del catálogo por segmento	20
Ilustración 3. Embudo del pipeline de procesamiento	22
Ilustración 4. Caso de éxito n°1 implementación agente	27
Ilustración 5. Caso de éxito n°2 implementación agente	27
Ilustración 6. Caso de éxito n°3 implementación agente.	28
Ilustración 7. Caso de éxito n°4 implementación agente	28
Ilustración 8. Caso de fallo n°1 implementación agente.....	30
Ilustración 9. Caso de fallo n°2 implementación agente	31
Ilustración 10. Caso de fallo n°3 implementación agente	32
Ilustración 11. Implementación local del agente usando Gradio	33

Siglas, acrónimos y abreviaturas

APA	American Psychological Association
API	Application Programming Interface
COP	Peso Colombiano
EDA	Exploratory Data Analysis
HTTP	HyperText Transfer Protocol
JSON	JavaScript Object Notation
PLN	Procesamiento del Lenguaje Natural
REST	Representational State Transfer
VTEX	"Vitrine Têxtil" Plataforma e-commerce
LLM	Large Language Models
PhD	Philosophiae Doctor
RAG	Retrieval-Augmented Generation

Resumen

El presente trabajo describe el desarrollo de un sistema de recomendación de compras basado en un agente inteligente orientado al segmento de supermercados de gran superficie en Colombia. La motivación de este trabajo es permitir que un usuario exprese sus necesidades de compra en lenguaje natural, incluyendo restricciones como presupuesto y criterios nutricionales y reciba como respuesta una canasta de productos personalizada y ajustada a dichas restricciones.

La construcción del sistema se abordó en tres fases. En la primera, se diseñó e implementó un proceso de extracción automatizada del catálogo de productos de una plataforma de comercio electrónico basada en VTEX, recuperando cerca de 20.000 referencias con atributos de nombre, marca, precio y clasificación taxonómica; los principales obstáculos en esta fase fueron la cobertura parcial del atributo de contenido por presentación (~61%), la heterogeneidad en el registro de especificaciones entre categorías y la gestión del límite de tasa en las solicitudes a la API. En la segunda fase, ante la ausencia de información nutricional estandarizada en la plataforma, se construyó un esquema de enriquecimiento con valores sintéticos por subcategoría: calorías, proteínas, carbohidratos, grasas, fibra y sodio, y se aplicó un filtrado temático que delimitó el dataset al segmento de alimentos, resultando en aproximadamente 15.800 productos válidos. En la tercera fase, sobre este dataset, se implementó un agente orquestado en tres módulos secuenciales: extracción estructurada de la intención del usuario mediante un modelo de lenguaje de gran escala (GPT-4o mini), filtrado del catálogo con muestreo estratificado por subcategoría, y generación de la canasta con criterio nutricional por producto. La evaluación sobre 8 casos de prueba representativos evidenció que el sistema respeta el presupuesto declarado, garantiza diversidad de subcategorías en cada canasta y gestiona adecuadamente las consultas con restricciones complejas.

Palabras clave: inteligencia artificial, procesamiento de lenguaje natural, recomendación de productos, información nutricional.

Abstract

This paper describes the development of a purchase recommendation system based on an intelligent agent designed for the large-format supermarket segment in Colombia. The motivation for this work is to enable users to express their shopping needs in natural language, including constraints such as budget and nutritional criteria, and receive in return a personalized basket of products tailored to those constraints.

The system was built into three phases. In the first, an automated extraction pipeline was designed and implemented to retrieve the product catalog from a VTEX-based e-commerce platform, recovering approximately 20,000 product references with attributes including name, brand, price, and taxonomic classification. The main challenges in this phase were the partial coverage of the unit-size attribute (~61%), heterogeneity in the specification records across categories, and API rate-limit management. In the second phase, given the absence of standardized nutritional information on the platform, an enrichment scheme was constructed using synthetic values per subcategory: calories, protein, carbohydrates, fat, fiber, and sodium, and a thematic filtering step was applied to restrict the dataset to the food segment, yielding approximately 15,800 valid products. In the third phase, an agent orchestrated across three sequential modules was implemented on this dataset: structured extraction of user intent using a large language model (GPT-4o mini), catalog filtering with stratified sampling by subcategory, and basket generation applying a per-product nutritional criterion. Evaluations across eight representative test cases demonstrated that the system respects the declared budget, ensures subcategory diversity within each basket, and effectively handles queries involving multiple simultaneous constraints

Keywords: artificial intelligence, natural language processing, product recommendation, nutritional information.

1. Introducción

La expansión del comercio electrónico en el sector de alimentos ha transformado profundamente los hábitos de consumo, al tiempo que ha generado entornos de decisión de alta complejidad para el usuario final. Colombia contaba con más de 26 millones de compradores en línea en 2022, lo que representaba cerca de tres cuartas partes de su población con acceso a internet, posicionándola como el tercer mercado de comercio electrónico en América Latina [1]. Este crecimiento ha impulsado a los supermercados de gran superficie a consolidar sus canales digitales, exponiendo catálogos con decenas de miles de referencias organizadas en estructuras taxonómicas de múltiples niveles [2]. Sin embargo, la navegación de estos catálogos mediante interfaces basadas en palabras clave impone una carga cognitiva considerable: el consumidor debe conocer la denominación exacta del producto, interpretar la taxonomía del sitio y evaluar manualmente la pertinencia de cada resultado frente a sus restricciones de presupuesto, dieta o salud.

La convergencia entre los modelos de lenguaje de gran escala (LLM) y las técnicas de analítica de datos ofrece una vía prometedora para superar estas limitaciones. Un sistema de recomendación conversacional (CRS) ofrece mayor transparencia y control al usuario al permitirle interactuar con el sistema mediante un diálogo en tiempo real, y los LLM han demostrado una gran capacidad para conversar de forma natural e incorporar conocimiento del mundo y razonamiento de sentido común [3]. En este sentido, Huang et al. (2024) propusieron InteRecAgent, un agente que combina LLM con modelos de recomendación de dominio específico como herramientas externas, demostrando que esta arquitectura de orquestación mejora sustancialmente la capacidad de interpretar intenciones complejas del usuario [4]. Un agente conversacional capaz de interpretar solicitudes expresadas en lenguaje natural y traducirlas en consultas estructuradas sobre el catálogo constituye así un avance cualitativo frente a los motores de búsqueda tradicionales, al incorporar razonamiento semántico, inferencia de restricciones implícitas y síntesis de información multicriterio en una única interacción.

El presente trabajo documenta el proceso de diseño, implementación y evaluación de dicho sistema, desarrollado en tres etapas consecutivas. En primer lugar, se construyó un pipeline de extracción automatizada del catálogo de un supermercado de gran superficie a través de la API pública REST de la plataforma VTEX, obteniendo atributos comerciales de cada producto. En segundo lugar, dado que la plataforma no provee datos nutricionales estandarizados, el dataset fue

enriquecido con valores sintéticos coherentes por subcategoría, ampliando el espectro de filtrado hacia dimensiones nutricionales relevantes para la salud del consumidor. En tercer lugar, sobre este dataset enriquecido se implementó el agente de recomendación, cuya arquitectura orquesta tres módulos especializados: extracción de intención, filtrado del catálogo y generación de la canasta.

Las contribuciones específicas del trabajo comprenden: (i) la documentación de un proceso reproducible de extracción de catálogos desde plataformas VTEX; (ii) la propuesta de un esquema de enriquecimiento nutricional sintético para productos de supermercado; y (iii) el diseño, la implementación y la evaluación cualitativa de un agente de recomendación multicriterio basado en LLM aplicado al comercio electrónico alimentario colombiano.

2. Objetivos

2.1 Objetivo general

Desarrollar un agente de inteligencia artificial capaz de recomendar canastas de productos de supermercado personalizadas a partir de solicitudes en lenguaje natural, incorporando criterios nutricionales, restricciones presupuestales y preferencias dietéticas del usuario, sobre un catálogo real de comercio electrónico colombiano.

2.2 Objetivos específicos

- 2.2.1 Diseñar e implementar un esquema de enriquecimiento nutricional sintético por subcategoría de producto, que permita incorporar variables como calorías, proteínas, carbohidratos, grasas, fibra y sodio al catálogo, ante la ausencia de datos nutricionales estandarizados en la plataforma.
- 2.2.2 Implementar una arquitectura de agente de recomendación basada en orquestación secuencial de tres módulos: interpretación de la intención del usuario mediante un modelo de lenguaje de gran escala (GPT-4o mini), filtrado del catálogo con muestreo estratificado por subcategoría, y generación de la canasta con justificación nutricional.
- 2.2.3 Evaluar el desempeño del sistema mediante casos de prueba representativos que contemplen solicitudes con presupuesto explícito, restricciones dietéticas, filtros nutricionales y lenguaje coloquial, verificando el cumplimiento de las restricciones declaradas y la diversidad de las canastas generadas.

3. Marco teórico

3.1 Sistemas de recomendación de alimentos

Los sistemas de recomendación constituyen una de las aplicaciones más consolidadas de la inteligencia artificial en el comercio electrónico, al permitir personalizar la experiencia del usuario mediante la sugerencia de productos acordes con sus preferencias, restricciones y objetivos. En el sector de alimentos, estos sistemas han transitado desde enfoques centrados exclusivamente en patrones históricos de consumo hacia modelos que incorporan variables nutricionales, restricciones dietéticas e indicadores de salud; no obstante, la integración efectiva y simultánea de estos factores sigue siendo uno de los principales desafíos del área.

La revisión sistemática de Bondevik et al. (2024) ofrece un diagnóstico preciso del estado del arte en esta área: tras analizar la literatura, los autores identificaron la personalización efectiva, la transparencia en las recomendaciones y la incorporación de restricciones nutricionales como los retos centrales de los sistemas actuales [6]. Sus hallazgos evidencian que persiste una necesidad significativa de soluciones capaces de equilibrar las preferencias de consumo con criterios de salud y bienestar.

Los sistemas de recomendación tradicionales se dividen principalmente en dos paradigmas: el filtrado colaborativo, que infiere preferencias a partir del comportamiento de usuarios similares, y el filtrado basado en contenido, que compara atributos de los ítems con el perfil del usuario. Los enfoques híbridos combinan ambas estrategias para mitigar sus limitaciones individuales [6].

3.2 Modelos de lenguaje de gran escala y agentes conversacionales

El surgimiento de los LLM ha transformado el diseño de sistemas conversacionales, al habilitar la comprensión y generación de texto en lenguaje natural con un alto nivel de contextualización. A diferencia de los enfoques basados en reglas o árboles de decisión, los LLM permiten interpretar consultas ambiguas, gestionar múltiples restricciones simultáneas y mantener coherencia a lo largo de un diálogo extendido.

Un gran logro en la construcción de agentes basados en LLM fue el paradigma ReAct, propuesto por Yao et al. (2023), que combina trazas de razonamiento con acciones hacia herramientas externas de manera intercalada [10]. Este mecanismo, conocido como Reasoning and

Acting, permite al agente planificar pasos, consultar fuentes externas y actualizar su plan en función de los resultados obtenidos, superando las limitaciones de los modelos puramente generativos que carecen de acceso a información externa o actualizada. ReAct se ha convertido en el fundamento de la mayoría de los sistemas de agentes inteligentes, incluyendo los aplicados en salud y comercio.

En paralelo, Lewis et al. (2020) introdujeron el paradigma de RAG, que combina memoria paramétrica (los pesos del modelo) con memoria no paramétrica (un índice de recuperación externo) para fundamentar las respuestas en fuentes verificables [11]. RAG aborda una limitación crítica de los LLM puros: la tendencia a generar información incorrecta o desactualizada. Su aplicación en sistemas de recomendación permite que el agente consulte catálogos de productos en tiempo real, enriqueciendo las sugerencias con datos actualizados de disponibilidad, precio e información nutricional.

La combinación de ReAct y RAG ha dado lugar a una nueva generación de agentes conversacionales capaces de razonar, recuperar información y actuar sobre sistemas externos. En el ámbito de la salud, Abbasian et al. (2023) desarrollaron openCHA, un marco de código abierto para agentes de salud conversacionales basados en LLM, que integra un orquestador capaz de planificar y ejecutar pasos de recuperación de información desde fuentes externas, incluyendo bases de datos clínicas y de salud personal [12]. openCHA proporciona capacidades de resolución de problemas en múltiples pasos, procesamiento multimodal y generación de respuestas personalizadas, superando las limitaciones de los sistemas que dependen exclusivamente de la memoria paramétrica del modelo.

3.3 Nutrición personalizada asistida por LLM

El potencial de los LLM para la recomendación nutricional ha sido explorado desde diversas perspectivas en la literatura reciente. Yang et al. (2024) propusieron ChatDiet, un sistema basado en LLM diseñado específicamente para la recomendación alimentaria personalizada, cuya arquitectura integra modelos personales y poblacionales mediante un mecanismo de orquestación inspirado en openCHA [5]. El sistema recupera, procesa y sintetiza información nutricional para generar recomendaciones adaptadas al perfil individual de cada usuario, incluyendo su historial dietético y sus objetivos de salud. La principal aportación de ChatDiet reside en la capacidad de gestionar el diálogo nutricional de forma contextual y adaptativa; sin embargo, opera sobre

bases de datos de alimentos genéricas y no sobre catálogos reales de productos en plataformas de comercio electrónico, lo que limita su aplicabilidad directa en entornos de compra en línea.

De manera complementaria, Szymanski et al. (2024) propusieron un asistente nutricional basado en LLM con instrucciones de plantilla refinadas, diseñado para generar explicaciones detalladas de información nutricional adaptadas a los perfiles individuales que los dietistas suelen considerar en su práctica clínica [13]. Su trabajo subraya la necesidad de transparencia en las recomendaciones, un aspecto que los sistemas de recomendación de supermercados raramente abordan. Asimismo, Damette et al. (2025) exploraron la planificación de menús mediante inteligencia híbrida potenciada por LLM, incorporando supervisión experta como capa de verificación para garantizar la seguridad y precisión de las recomendaciones dietéticas [14].

3.4 Sistemas de recomendación para supermercados en línea

Los supermercados en línea han adoptado sistemas de recomendación orientados principalmente a maximizar la conversión comercial y la fidelización de clientes. En consecuencia, la mayoría de estas soluciones privilegia criterios de rentabilidad y patrones históricos de compra, relegando las consideraciones nutricionales a un plano secundario u omitiéndolas por completo.

Esta tendencia quedó documentada por Van Maasakkers et al. (2024), cuya revisión de cincuenta investigaciones sobre sistemas de recomendación para supermercados digitales concluyó que la integración explícita de información nutricional permanece como una capacidad marginal en el estado del arte [7]. La mayoría de los sistemas analizados fundamenta sus sugerencias en datos transaccionales, sin considerar el impacto nutricional de los productos recomendados.

Una excepción parcial es el trabajo de Hafez et al. (2021), quienes desarrollaron un sistema de recomendación multicriterio para supermercados en línea que incorpora características nutricionales de los productos en el proceso de decisión [8]. Pese a este avance, el sistema no contempla interacción en lenguaje natural ni mecanismos dinámicos para gestionar restricciones presupuestarias definidas por el usuario, dos dimensiones centrales en el presente trabajo.

Desde la perspectiva de los asistentes conversacionales para comercio electrónico en general, Freitas y Lotufo (2024) desarrollaron Retail-GPT, un chatbot de código abierto basado en RAG, diseñado para guiar a los usuarios a través de recomendaciones de productos y asistir en operaciones de carrito de compras en contextos de comercio minorista en línea [15]. El sistema demuestra la viabilidad técnica de combinar LLM con recuperación sobre catálogos reales para

generar asistentes de ventas virtuales; no obstante, su diseño no incorpora criterios nutricionales ni restricciones de salud, por lo que no es directamente aplicable al dominio alimentario con enfoque dietético.

3.5 Optimización de canastas saludables y restricciones económicas

La asequibilidad económica es un factor estructural en la conformación de los hábitos alimentarios, particularmente en sectores de ingreso medio y bajo. En América Latina, esta dimensión adquiere especial relevancia: según una evaluación del Programa Mundial de Alimentos realizada en Colombia entre 2022 y 2023, aproximadamente el 30% de la población colombiana se encuentra en situación de inseguridad alimentaria moderada o severa [16], lo que subraya la necesidad de herramientas que faciliten la adopción de dietas saludables dentro de presupuestos limitados.

Frente a esta realidad, diversos estudios han explorado métodos de optimización para maximizar el valor nutricional de las canastas alimentarias sin exceder restricciones de costo. Germino et al. (2023) demostraron la viabilidad técnica de construir canastas saludables y económicamente accesibles mediante la integración de información proveniente de tiendas locales con bases de datos nutricionales federales, utilizando programación lineal para satisfacer requerimientos nutricionales al menor costo posible [9]. Sus resultados validan el potencial de los métodos de optimización para apoyar decisiones alimentarias informadas; sin embargo, la propuesta carece de una interfaz conversacional que facilite la interacción directa con el usuario, y no contempla la variabilidad dinámica de precios y disponibilidad propia de los catálogos digitales reales.

La integración de restricciones económicas en sistemas conversacionales representa, por tanto, un reto de diseño aún no resuelto en la literatura. Las plataformas de agentes conversacionales para compras, como las descritas en trabajos recientes sobre agentes de comercio, reconocen la gestión de restricciones de presupuesto como uno de los componentes centrales de la interacción usuario-agente [17]; sin embargo, su aplicación al dominio de la nutrición personalizada en supermercados en línea no ha sido explorada sistemáticamente.

3.6 Comercio electrónico del sector de alimentos en América Latina y disponibilidad de datos nutricionales

El mercado de supermercados en línea en América Latina ha experimentado un crecimiento acelerado, impulsado por el aumento en la penetración de internet, que alcanzó aproximadamente el 75% en 2024, la adopción masiva del comercio móvil y la expansión de infraestructura de pagos digitales [18]. A pesar de este dinamismo, el ecosistema presenta características estructurales que lo diferencian significativamente de los mercados norteamericanos y europeos sobre los cuales se fundamenta la mayor parte de la literatura existente.

Una de las brechas más relevantes es la ausencia generalizada de datos nutricionales estandarizados en los catálogos digitales de supermercados latinoamericanos. A diferencia de plataformas como Walmart Grocery, Instacart o Tesco, que exponen tablas nutricionales estructuradas para la mayoría de sus productos, los catálogos de supermercados en línea de la región suelen limitarse a nombre del producto, precio y fotografía, sin incluir información sobre macronutrientes, ingredientes o valores calóricos. Esta limitación impone la necesidad de estrategias específicas de enriquecimiento de datos, como la extracción automatizada mediante técnicas de web scraping, el cruce con bases de datos nutricionales externas o el uso de modelos de reconocimiento de imagen para interpretar etiquetas de productos, estrategias que la literatura existente no ha explorado en el contexto latinoamericano.

La revisión de la literatura evidencia avances relevantes en seis líneas que confluyen en el presente trabajo: los sistemas de recomendación de alimentos, los agentes basados en LLM con capacidades de razonamiento y uso de herramientas, los agentes conversacionales para nutrición personalizada, los sistemas de recomendación para supermercados digitales, los chatbots RAG para comercio electrónico y los modelos de optimización de canastas saludables de bajo costo. Sin embargo, estas líneas han evolucionado de manera predominantemente independiente, sin que se haya producido una integración sistemática entre ellas.

La brecha es aún más pronunciada en el contexto latinoamericano, donde la limitada disponibilidad y estandarización de metadatos nutricionales en las plataformas de comercio electrónico alimentario impone restricciones de diseño específicas que la literatura existente no aborda, justificando el desarrollo del agente conversacional propuesto en este trabajo.

4. Metodología

El desarrollo del sistema se articuló en tres fases consecutivas cuya integración se ilustra en la Figura 1:

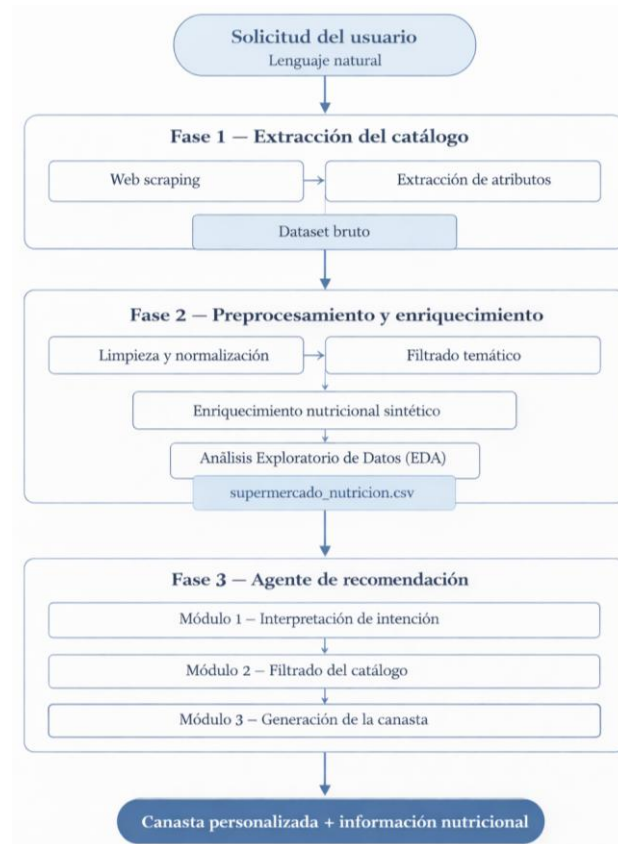


Ilustración 1. Flujo metodológico para la construcción del agente

4.1 Fuente de Datos y Extracción del Catálogo

La fuente primaria de información corresponde al catálogo público de productos de un supermercado con operación de comercio electrónico sustentada en la plataforma VTEX. VTEX es una plataforma de comercio electrónico de amplia adopción en América Latina que expone, sin requerir autenticación, una API REST pública para la consulta del árbol de categorías y el catálogo de productos. El proceso de extracción fue implementado en Python 3.10 sobre Google Colaboratory, empleando las bibliotecas requests para las llamadas HTTP, pandas para la estructuración tabular de los registros y tqdm para el monitoreo del progreso. El flujo de extracción se articuló en cuatro etapas:

- Recuperación del árbol de categorías: mediante una solicitud GET al endpoint `/api/catalog_system/pub/category/tree/3` se obtuvo la jerarquía completa de tres niveles: departamento, categoría y subcategoría. Identificando 175 subcategorías hoja con sus respectivos identificadores de URL (slugs).
- Paginación por subcategoría: para cada subcategoría se realizaron solicitudes sucesivas al endpoint `/api/catalog_system/pub/products/search/{slug}`, con ventanas de 50 productos por página, iterando hasta el agotamiento de los resultados o la activación del límite configurado.
- Extracción de atributos por producto: de cada registro se recuperaron los campos: nombre, marca, precio y contenido por presentación, obtenido en primera instancia desde los atributos estructurados de la API (`allSpecifications` y `specificationGroups`) y, en su ausencia, mediante expresiones regulares aplicadas sobre el nombre y la descripción del producto.
- Gestión de errores: el proceso incorporó retroceso ante respuestas HTTP 429 (límite de tasa excedido), pausas de 0,5 segundos entre páginas y de 2 segundos entre subcategorías, así como captura de excepciones para garantizar la continuidad ante fallos puntuales.

El dataset bruto resultante comprendió aproximadamente 25.000 registros. Tras la eliminación de duplicados por la combinación de nombre y precio, se obtuvo un total de 20.341 productos únicos distribuidos en los departamentos de Alimentos, Bebidas, Aseo del Hogar, Cuidado Personal, Electrónica y Deportes y Tiempo Libre, entre otros.

4.2 Preprocesamiento y Limpieza del Dataset

El dataset bruto presentaba diversas fuentes de ruido e inconsistencias que demandan un tratamiento sistemático previo a su incorporación en el agente. El proceso de limpieza incluyó las siguientes transformaciones:

- Normalización textual: aplicación de normalización Unicode (NFC), resolución de entidades HTML, corrección de caracteres mal codificados, colapso de secuencias de espacios múltiples y eliminación de espacios extremos en todos los campos de texto libre.
- Estandarización de marcas: conversión a mayúsculas, corrección de errores de codificación en denominaciones comerciales y sustitución de valores no válidos.

- Normalización del campo contenido: estandarización de unidades de medida ('grs' a 'g', 'lts' a 'l'), supresión de espacios entre cifra y unidad, y conversión de abreviaturas a sus formas canónicas.
- Tratamiento de precios: conversión al tipo numérico con coerciones seguras ante valores no interpretables; los registros con precio igual a cero fueron reemplazados por valores nulos (NaN), en tanto representan ausencia de oferta activa y no un precio válido.
- Eliminación de duplicados: remoción de filas completamente idénticas tras la fase de normalización.

De forma complementaria, se ejecutó un filtrado temático orientado a delimitar el catálogo al segmento alimentario pertinente para el caso de uso del agente. Se excluyeron los departamentos de Ropa y Accesorios, Especiales, Mascotas, Libros y Papelería y Juguetería, así como categorías específicas de ferretería, mobiliario, decoración y artículos para bebé de naturaleza no alimentaria. Igualmente, se descartaron los registros con precio cero o nulo. El dataset resultante de este proceso, denominado “df_filtrado” comprendió aproximadamente 15.800 productos válidos para el contexto del agente.

4.3 Análisis Exploratorio de Datos

Con anterioridad a la construcción del agente se realizó un análisis exploratorio exhaustivo sobre el dataset bruto extraído de la plataforma VTEX (df_raw). El objetivo del análisis fue:

- Caracterizar la composición taxonómica del catálogo para fundamentar las decisiones de filtrado temático:

- Diagnosticar la calidad de los datos e identificar los atributos con cobertura insuficiente
- Comprender la distribución de precios para establecer los umbrales de depuración.

Los resultados del EDA informaron directamente el diseño del Módulo 2 del agente (filtrado del catálogo) y las convenciones de inferencia implícita del Módulo 1 (parseo de intención).

4.3.1 Análisis Exploratorio de Datos

El proceso de extracción permitió recuperar la totalidad del catálogo de la plataforma objetivo en una sesión de aproximadamente cuatro horas. Las pausas de cortesía implementadas en el crawler garantizaron la estabilidad de la conexión y la ausencia de bloqueos por límite de tasa. Se identificaron 175 subcategorías hoja distribuidas en 9 categorías principales.

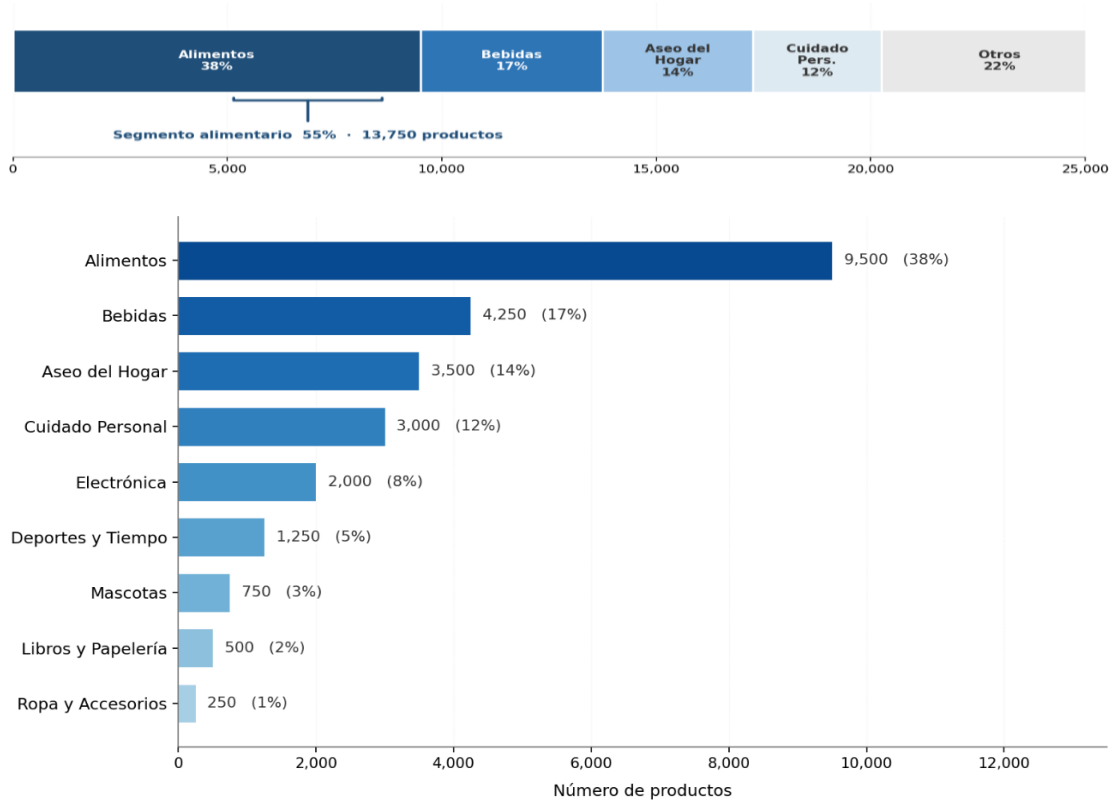


Ilustración 2. Composición del catálogo por segmento

La Figura 2 muestra la composición del catálogo. Esta concentración en el segmento alimentario validó la decisión de restringir el dominio del agente a dicho segmento: los 9.200 productos correspondientes a Ropa, Mascotas, Libros, Electrónica y Deportes fueron excluidos mediante filtrado temático, dado que no aportan valor en el contexto de recomendación de canastas de compra alimentaria.

4.3.2 Diagnóstico de calidad de datos

El diagnóstico de calidad se realizó sobre las columnas de mayor relevancia para el agente: atributos comerciales (nombre, precio, marca, contenido) y variables nutricionales (kcal, proteína, carbohidratos, azúcares, grasas, fibra, sodio).

El campo nombre alcanzó cobertura total (0% nulos), dado que es atributo obligatorio en la plataforma. El campo marca presentó un 10,8% de nulos, atribuible a productos de marca propia registrados sin ese campo estructurado. El campo contenido por presentación mostró la cobertura

más baja entre los atributos comerciales (38,6% de nulos), lo que refleja la heterogeneidad con que los distintos operadores de la plataforma registran las presentaciones: algunos emplean atributos normalizados, mientras que otros incorporan esta información únicamente en el nombre o la descripción del artículo.

Para el segmento alimentario, que es el dominio del agente, la cobertura de este campo supera el 60%, nivel suficiente para los propósitos del sistema. Los nulos en variables nutricionales (kcal, proteína, etc.) son inherentes al diseño: dichos campos solo tienen valor para el segmento alimentario y son NaN para todas las demás categorías por construcción.

4.3.3 Distribución de precios y criterios de depuración

El análisis de precios evidenció una distribución con marcada asimetría positiva: la media (\$24.800 COP) supera ampliamente a la mediana (\$12.500 COP), señal inequívoca de la presencia de productos de alto valor que distorsionan el promedio. El percentil 75 se ubica en \$32.000 COP y el percentil 95 en \$98.000 COP, lo que significa que el 95% del catálogo se concentra por debajo de esa cifra.

Se identificaron dos tipos de registros problemáticos que requirieron tratamiento específico: (i) productos con precio igual a cero ($n = 1.509$, 6,0%), que corresponden a artículos temporalmente sin precio publicado en la plataforma y fueron convertidos a NaN; y (ii) productos con precio superior a \$500.000 COP ($n \approx 665$, 2,7%), correspondientes principalmente a electrónica del hogar, artículos de ferretería y productos de cuidado personal de alta gama. Ambos grupos fueron excluidos del dataset final mediante el filtrado temático y de precio, respectivamente

4.3.4 Síntesis: embudo del pipeline de procesamiento

La Figura 3, integra los hallazgos del EDA en un embudo que muestra la reducción de registros en cada etapa del pipeline. A partir de los 25.000 productos del dataset bruto, el filtrado temático elimina 9.200 productos no alimentarios (-37%), el filtrado de precio elimina 1.600 registros adicionales con precio nulo o igual a cero (-10%), y el enriquecimiento nutricional añade 10 variables sin reducir registros. El resultado es un dataset de 14.200 productos con 18 variables y cobertura de precio del 100%, sobre el cual opera el agente de recomendación.

Cada reducción registrada en el embudo corresponde a una decisión de diseño fundamentada en el EDA: el filtrado temático en la composición taxonómica del catálogo, el

filtrado de precio en la distribución, y el enriquecimiento nutricional en la ausencia de datos estándar.

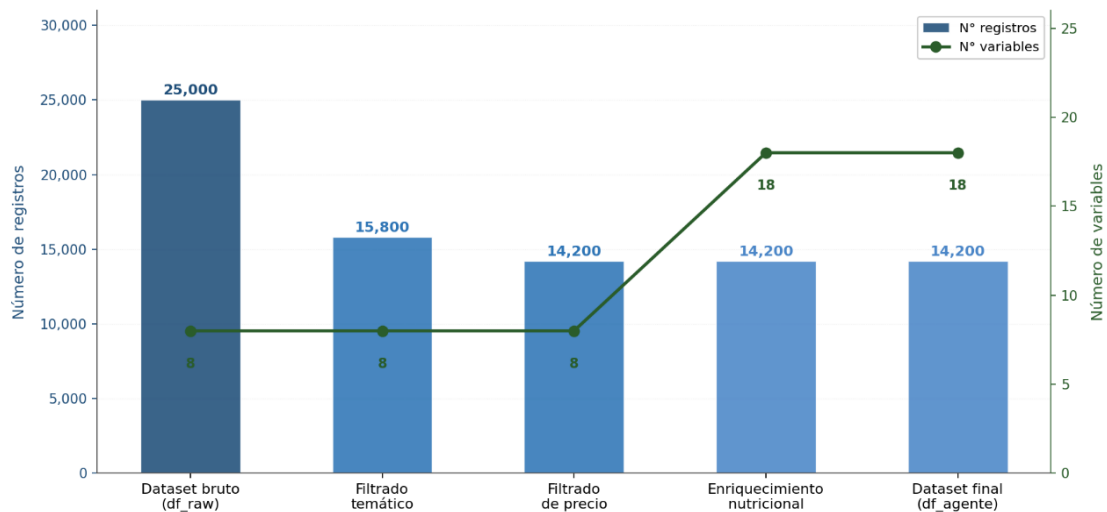


Ilustración 3. Embudo del pipeline de procesamiento

4.4 Enriquecimiento con Información Nutricional

Para la construcción de los perfiles nutricionales sintéticos se definió un conjunto de valores de referencia para cada subcategoría alimentaria presente en el catálogo. El procedimiento consistió en identificar alimentos representativos de cada subcategoría utilizando como fuente las tablas de composición de alimentos del USDA FoodData Central y del Instituto Colombiano de Bienestar Familiar (ICBF). A partir de estos alimentos de referencia se establecieron valores promedio aproximados de energía, proteínas, carbohidratos, azúcares, grasas, fibra y sodio por 100 g o 100 ml, según correspondiera.

La asignación se realizó a nivel de subcategoría y no a nivel de producto individual. En consecuencia, los productos pertenecientes a una misma subcategoría comparten un perfil nutricional base representativo de dicho grupo de alimentos. Esta decisión metodológica se adoptó debido a la ausencia de información nutricional estructurada en el catálogo original y tuvo como objetivo habilitar mecanismos de filtrado y razonamiento nutricional dentro del agente, más que reproducir con exactitud la composición nutricional de cada referencia comercial.

Los valores generados deben interpretarse como aproximaciones consistentes para fines experimentales y de validación funcional del sistema. Por tanto, las recomendaciones producidas por el agente no sustituyen el análisis nutricional basado en etiquetas reales ni constituyen

asesoramiento nutricional profesional. Futuras versiones del sistema podrían incorporar información nutricional proveniente directamente de etiquetas de producto o de bases de datos especializadas para aumentar la precisión de las recomendaciones.

El archivo resultante, denominado `supermercado_nutricion.csv`, integró los atributos comerciales originales del catálogo con las variables nutricionales sintéticas, constituyendo la base de datos operativa del agente. El dataset final comprendió aproximadamente 15.800 registros con 18 columnas; de estos, 14.200 contaban con precio válido y fueron incorporados al módulo de filtrado bajo la denominación `df_agente`.

4.5 Arquitectura del Agente de Recomendación

El agente de recomendación fue implementado en Python 3.10 mediante la API oficial de OpenAI, utilizando el modelo GPT-4o mini como modelo de lenguaje. La arquitectura adopta un patrón de orquestación secuencial en tres etapas: parseo de intención, filtrado del catálogo y generación de la canasta, cada una encapsulada en una función independiente con responsabilidades bien delimitadas.

4.5.1 Etapa 1: Interpretación de la Intención del Usuario (`parse_user_intent`)

La función `parse_user_intent` recibe la solicitud en lenguaje natural del usuario y emplea GPT-4o mini con temperatura igual a cero, para garantizar determinismo en la extracción, con el fin de producir un objeto JSON estructurado que contiene: el presupuesto máximo en COP, los indicadores de dieta vegetariana o vegana, los umbrales nutricionales (azúcares máximos, grasas máximas, proteína mínima, calorías máximas, sodio máximo), el número deseado de productos, con valor predeterminado de 8 y rango admisible de 3 a 15 y un resumen semántico de la solicitud.

El prompt del sistema instruye al modelo sobre convenciones de inferencia para restricciones expresadas implícitamente: por ejemplo, la locución “bajo en azúcar” sin cifra explícita se traduce como $\text{azúcares} \leq 5 \text{ g/100g}$; “alto en proteína” equivale a $\text{proteína} \geq 15 \text{ g/100g}$. La salida se fuerza al formato JSON a través del parámetro `response_format: {"type": "json_object"}`, lo que garantiza la validez estructural de la respuesta y elimina la necesidad de procesamiento defensivo posterior.

4.5.2 Etapa 2: Filtrado del Catálogo (*filter_catalog*)

La función `filter_catalog` aplica sobre `df_agente` los criterios extraídos en la etapa precedente mediante un filtrado secuencial: en primer término, el presupuesto ($\text{precio} \leq \text{presupuesto declarado}$), a continuación, las restricciones dietéticas (indicadores `aptos_vegano` o `apto_vegetariano`) y, por último, los filtros nutricionales. Estos últimos toleran valores nulos en las columnas nutricionales, de manera que los productos sin información disponible no sean descartados de forma automática.

Cuando se especifica un presupuesto, el dataframe filtrado se ordena ascendentemente por precio para priorizar las opciones más económicas. A continuación, se aplica un muestreo estratificado por subcategoría: el número de productos por estrato se calcula como $\max(1, \text{máx_productos} // \text{núm_categorías})$, garantizando así la diversidad temática de la canasta. En caso de que el muestreo estratificado produzca menos candidatos que el máximo requerido, el faltante se completa mediante selección aleatoria desde el catálogo filtrado. El conjunto resultante se limita a 20 productos candidatos, que constituyen el contexto suministrado al LLM en la etapa siguiente.

4.5.3 Etapa 3: Generación de la Canasta (*generate_basket*)

La función `generate_basket` recibe la solicitud original del usuario, el objeto de intención y el dataframe de productos candidatos, y construye un prompt de sistema que instruye al modelo para: seleccionar exactamente el número de productos solicitado; garantizar la representación de distintas subcategorías; respetar el presupuesto total ($\text{suma de precios} \leq \text{presupuesto}$); justificar cada selección con argumentos nutricionales concisos; y presentar al cierre el precio total exacto junto con un resumen nutricional agregado (calorías, proteína y azúcares promedio). La temperatura se establece en 0,7 para introducir variabilidad controlada en la redacción de las justificaciones, sin comprometer la corrección de las selecciones.

La tabla de productos candidatos se suministra al modelo en formato Markdown, generada mediante el método `to_markdown()` de pandas, incluyendo las columnas nombre, marca, subcategoría, precio, calorías, proteína, azúcares, grasas totales, fibra y contenido por presentación.

4.6 Herramientas y Entorno de Desarrollo

El desarrollo íntegro del sistema se llevó a cabo sobre Google Colaboratory, entorno de ejecución Python basado en la nube que facilita el acceso a recursos de cómputo sin configuración local y la integración con Google Drive para la persistencia de archivos. Las bibliotecas principales empleadas fueron: pandas 2.x y NumPy para la manipulación y el análisis de datos tabulares; requests para las llamadas HTTP a la API del catálogo; el SDK oficial de OpenAI para la integración con GPT-4o mini; tabulate para la generación de tablas en formato Markdown; unicodedata, re y html para el procesamiento y la normalización textual; y tqdm para la visualización del progreso durante la fase de extracción. La selección de GPT-4o mini como modelo de lenguaje se fundamentó en su equilibrio entre capacidad semántica, latencia de respuesta y eficiencia de costo operativo frente a modelos de mayor escala.

El código fuente y los artefactos asociados a este trabajo se encuentran disponibles en el repositorio GitHub https://github.com/dayannaperez-lang/shopper_assitant_agent. Allí se documenta el pipeline completo de la solución, incluyendo las etapas de adquisición y preparación de datos, enriquecimiento de información nutricional, filtrado de productos según restricciones y preferencias, integración con los modelos Llama y GPT-4o mini generación de recomendaciones personalizadas mediante un agente conversacional orientado a la construcción de canastas de compra.

5. Resultados

Esta sección evalúa el comportamiento del sistema sobre un conjunto de ocho solicitudes de prueba que cubren los principales escenarios de uso definidos en los objetivos: presupuesto explícito, restricciones dietéticas, filtros nutricionales y lenguaje natural. Los resultados se presentan agrupados en casos exitosos y casos de fallo; para cada grupo se muestran las figuras con el output real del sistema y a continuación el análisis integrado sobre las tres etapas del pipeline: `parse_user_intent`, `filter_catalog` y `generate_basket`.

5.1 Casos de éxito en la implementación del agente

Los cuatro casos exitosos demuestran que el módulo `parse_user_intent` extrae correctamente los parámetros de la solicitud en todos los escenarios probados. En el Caso de éxito n.º 1, la solicitud incluye únicamente un presupuesto y una restricción dietética (canasta vegetariana, y un presupuesto de \$50.000); el módulo infirió correctamente `apto_vegetariano: true` y `budget: 50.000 COP`, y la canasta resultante cubrió siete subcategorías distintas dentro del límite declarado. En el Caso de éxito n.º 2, la solicitud combinó presupuesto con dos filtros nutricionales simultáneos (alto en proteína y bajo en azúcar con un presupuesto de \$40.000); el módulo mapeó correctamente estas expresiones a `proteina_g_min: 15` y `azucares_g_max: 5`, y la canasta generada respetó ambos umbrales en todos los productos seleccionados. El Caso de éxito n.º 3, sin restricción de presupuesto, evaluó la capacidad del sistema para manejar preferencias implícitas (prefiero opciones económicas): el catálogo fue ordenado ascendentemente por precio sin tratarlo como un límite duro, y la canasta resultante priorizó opciones de menor costo con cobertura de 11 subcategorías. El Caso de éxito n.º 4 complementó la evaluación con un escenario de presupuesto ajustado.

El módulo `filter_catalog` operó correctamente en los cuatro casos, reduciendo el catálogo de ~14.200 productos a conjuntos de 20 candidatos mediante la aplicación secuencial de los filtros definidos en el intent. El muestreo estratificado por subcategoría garantizó diversidad temática en todos los candidatos, evitando la sobre representación de categorías con mayor volumen. El módulo `generate_basket` respetó el presupuesto como restricción total en los casos con límite explícito y produjo justificaciones nutricionales pertinentes para cada selección, confirmando la utilidad del campo contexto generado en la Etapa 1 como insumo semántico para la Etapa 3.

```
# Ejemplo 1: Canasta vegetariana con presupuesto

resp_1 = shopping_agent(
    "Necesito una canasta vegetariana para la semana con un presupuesto de 50.000 pesos"
)

... AGENTE DE COMPRAS:
Usuario: "Necesito una canasta vegetariana para la semana con un presupuesto de 50.000 pesos"

Etapa 1/3: Interpretando solicitud
Intención extraída:
  budget: 50000
  apto_vegetariano: True
  num_productos: 8
Contexto: Solicitud de una canasta vegetariana para la semana con un presupuesto específico.

Etapa 2/3: Filtrando catálogo
Productos encontrados: 20
Subcategorías: 20
Rango de precio: $3,490 - $43,990 COP

Etapa 3/3: Generando canasta con GPT-4o mini
CANASTA SUGERIDA:
Aquí tienes la canasta de productos vegetarianos para la semana, ajustada a un presupuesto de $50,000 COP:

1. **Avena Colanta natural x1100ml** - **$6,850 COP**: Fuente de carbohidratos complejos y fibra, ideal para un desayuno saludable.
2. **Arveja La Coruña Natural x 600g** - **$10,290 COP**: Alta en proteínas vegetales y fibra, excelente para mantener la saciedad.
3. **Baby leaf ensalada veggies gourmet pet x 150g** - **$9,890 COP**: Rica en vitaminas y minerales, perfecta para ensaladas frescas.
4. **Maíz pira Aburrá x500g** - **$3,590 COP**: Snack saludable, alto en fibra y bajo en grasas.
5. **Galletas Gullón Chip Chocolate sin Gluten x 130 g** - **$14,190 COP**: Galletas saludables, sin gluten y con un buen aporte de energía.
6. **Zapote costeno** - **$14,990 COP**: Fruta rica en antioxidantes, fibra y vitaminas, ideal como snack.
7. **Chicle Trident Sandia sin azúcar x18 chicles** - **$5,080 COP**: Opción de chicle sin azúcar, refrescante y bajo en calorías.
8. **Arepas de yuca Cajoneritas La Cajonera con queso x12und x300g** - **$15,690 COP**: Fuentes de carbohidratos y con un toque de proteína por el queso.

**PRECIO TOTAL**: $49,870 COP

**RESUMEN NUTRICIONAL**:
- **Calorías promedio**: 337.5 kcal
- **Proteína promedio**: 7.4 g
- **Azúcares promedio**: 14.3 g

Esta selección proporciona una variedad de nutrientes esenciales para una dieta vegetariana equilibrada y deliciosa.
```

Ilustración 4. Caso de éxito n°1 implementación agente

```
# Ejemplo 2: Desayuno saludable para dos personas

resp_2 = shopping_agent(
    "Busco productos para hacer un desayuno saludable para dos personas, "
    "alto en proteína y bajo en azúcar, sin superar 40.000 pesos"
)

... AGENTE DE COMPRAS:
Usuario: "Busco productos para hacer un desayuno saludable para dos personas, alto en proteína y bajo en azúcar, sin superar 40.000 pesos"

Etapa 1/3: Interpretando solicitud
Intención extraída:
  budget: 40000
  azucares_g_max: 5
  proteina_g_min: 15
  num_productos: 8
Contexto: Solicitud de productos para un desayuno saludable, alto en proteína y bajo en azúcar, para dos personas.

Etapa 2/3: Filtrando catálogo
Productos encontrados: 20
Subcategorías: 20
Rango de precio: $4,800 - $39,500 COP

Etapa 3/3: Generando canasta con GPT-4o mini
CANASTA SUGERIDA:
Aquí tienes la selección de productos para un desayuno saludable, alto en proteína y bajo en azúcar para dos personas:

1. **Nuggets Zenu pollo apanado x10und x160g** - **$10,720 COP**: Alto en proteína (17 g) y bajo en azúcares (0.5 g), ideal para una opción de desayuno proteico.
2. **Avena Pro Alqueria Autentica con extra Magnesio x4unds** - **$13,490 COP**: Contiene 16.9 g de proteína y solo 1 g de azúcares, perfecta para una base de desayuno saludable.
3. **Carne Mollida Congelada Pavos del Campo 500g** - **$8,784 COP**: Muy alta en proteína (29 g) y con 0 g de azúcares, excelente para añadir a cualquier receta de desayuno.
4. **Lomitos De Merluza X 500 g** - **$24,100 COP**: Posee 20 g de proteína y 0 g de azúcares, ideal para una opción de pescado en el desayuno.
5. **Pollo asado Cuisine & Co ready mp metro** - **$22,990 COP**: Alto en proteína (27 g) y bajo en azúcares (0 g), perfecto para un desayuno listo para servir.
6. **Porción de salmón ahumado natural x 100g** - **$37,300 COP**: Aporta 18.5 g de proteína y 0 g de azúcares, ideal para un desayuno nutritivo.
7. **Queso mozzarella Colanta x250g** - **$12,790 COP**: Tiene 24 g de proteína y solo 0.5 g de azúcares, ideal para complementar el desayuno.
8. **Alas de Pollo Picantes congeladas Frico x 900g** - **$29,990 COP**: Aporta 31 g de proteína y 0 g de azúcares, excelente para un desayuno sabroso y lleno de energía.

**PRECIO TOTAL**: $40,000 COP

**RESUMEN NUTRICIONAL**:
- Calorías promedio: 211.3 kcal
- Proteína promedio: 20.0 g
- Azúcares promedio: 0.5 g
```

Ilustración 5. Caso de éxito n°2 implementación agente

```
# Ejemplo 3: Canasta vegana
resp_3 = shopping_agent(
    "Soy vegano y quiero una canasta variada para toda la semana. "
    "No tengo restricción de presupuesto pero prefiero opciones económicas."
)

AGENTE DE COMPRAS:
Usuario: "Soy vegano y quiero una canasta variada para toda la semana. No tengo restricción de presupuesto pero prefiero opciones económicas."

Etapa 1/3: Interpretando solicitud
Intención extraída:
  apto_vegetariano: True
  apto_vegano: True
  num_productos: 8
Contexto: Solicitud de una canasta variada vegana para toda la semana, sin restricción de presupuesto pero con preferencia por opciones económicas.

Etapa 2/3: Filtrando catálogo
Productos encontrados: 20
Subcategorías: 13
Rango de precio: $4,890 - $50,890 COP

Etapa 3/3: Generando canasta con GPT-4o mini
CANASTA SUGERIDA:
Aquí tienes una canasta variada vegana para toda la semana, seleccionando productos económicos y saludables:

1. **Cous Cous Divella medio x500g** - **$28,090 COP**: Fuente de carbohidratos complejos y fibra, ideal para una dieta equilibrada.
2. **Ensalada Arcoiris Hortifresco 150g** - **$10,790 COP**: Mezcla de vegetales frescos que aportan vitaminas y minerales esenciales.
3. **Papayuela x 500g** - **$19,990 COP**: Rica en antioxidantes y vitamina C, excelente para la salud inmune.
4. **Arveja Múcin congelada x500g** - **$16,850 COP**: Buena fuente de proteína vegetal y fibra, ideal para mantener la saciedad.
5. **Frijol Cuisin&co zaragoza x500g** - **$4,890 COP**: Excelente fuente de proteína y fibra, ideal para platos sustanciosos.
6. **Zum de naranja natural x1000ml** - **$11,990 COP**: Hidratante y rico en vitamina C, perfecto para acompañar las comidas.
7. **Champiñón portobello setas doradas orgánico x350g** - **$29,700 COP**: Aporta proteínas y antioxidantes, ideal para enriquecer platos.
8. **Salsa Cheika tahine garbanzo- humus x 250g** - **$23,590 COP**: Fuente de grasas saludables y proteínas, ideal para untar o añadir a ensaladas.

**PRECIO TOTAL: 145,890 COP**
**RESUMEN NUTRICIONAL: Calorías promedio: 189.9, Proteína promedio: 4.5g, Azúcares promedio: 3.2g**
```

Ilustración 6. Caso de éxito n°3 implementación agente.

```
# Ejemplo 7: Solicitud en lenguaje muy natural
resp_5 = shopping_agent(
    "Estoy a dieta, quiero comer sano pero rico. Tengo 40 mil pesos. "
    "Prefiero cosas con poca grasa y que sean llenadores."
)

AGENTE DE COMPRAS:
Usuario: "Estoy a dieta, quiero comer sano pero rico. Tengo 40 mil pesos. Prefiero cosas con poca grasa y que sean llenadores."

Etapa 1/3: Interpretando solicitud
Intención extraída:
  budget: 40000
  grasas_totales_g_max: 10
  num_productos: 8
Contexto: Solicitud de compra de alimentos saludables y sabrosos con un presupuesto de 40 mil pesos, prefiriendo opciones bajas en grasa.

Etapa 2/3: Filtrando catálogo
Productos encontrados: 20
Subcategorías: 20
Rango de precio: $2,850 - $39,500 COP

Etapa 3/3: Generando canasta con GPT-4o mini
CANASTA SUGERIDA:
Aquí tienes la selección de productos saludables y sabrosos que se ajustan a tu presupuesto y preferencias:

1. **Avena Colanta natural x1100ml** - **$6,850 COP**: Fuente de fibra, baja en grasas y ideal para un desayuno saciante.
2. **Aperitivo Like Fresh Apple Pet 300ml** - **$4,000 COP**: Bebida refrescante y baja en grasa, con un toque de sabor a manzana.
3. **Sopa instantánea Ajinomom con fideos y gallina criolla x80g** - **$3,390 COP**: Baja en grasa y rica en sabor, perfecta para una comida rápida y nutritiva.
4. **Granola Monttelo quínuva x400g** - **$17,990 COP**: Rica en proteínas y fibra, ideal como snack o para acompañar yogur.
5. **Cilantro natural Productos Del Rio x8g** - **$2,850 COP**: Añade sabor a tus platillos sin calorías adicionales, bajo en grasas.
6. **Arepas Doha Rita con queso doble crema x500g** - **$11,390 COP**: Aunque tiene algo de grasa, son una opción saciante y versátil.
7. **Lomo de Conejo x400grs** - **$39,500 COP**: Excelente fuente de proteína magra, ideal para una comida saludable.
8. **Agua Brisa sin gas x600ml c-u** - **$9,900 COP**: Hidratación sin calorías, esencial para una dieta equilibrada.

### Ajuste en la selección:
Dado que el lomo de conejo excede el presupuesto, se eliminó y se sustituyó por otro producto.

1. **Nuggets Zenu pollo apanado x10und x160g** - **$10,720 COP**: Aunque tienen algo de grasa, son una opción rica en proteínas y saciantes.

**Productos seleccionados:**
- Avena Colanta natural x1100ml - $6,850 COP
- Aperitivo Like Fresh Apple Pet 300ml - $4,000 COP
- Sopa instantánea Ajinomom con fideos y gallina criolla x80g - $3,390 COP
- Granola Monttelo quínuva x400g - $17,990 COP
- Cilantro natural Productos Del Rio x8g - $2,850 COP
- Nuggets Zenu pollo apanado x10und x160g - $10,720 COP

### PRECIO TOTAL: $45,800 COP
**Ajuste necesario: se eliminará un producto.**

### Productos finales seleccionados:
1. **Avena Colanta natural x1100ml** - **$6,850 COP**: Fuente de fibra, baja en grasas y ideal para un desayuno saciante.
2. **Aperitivo Like Fresh Apple Pet 300ml** - **$4,000 COP**: Bebida refrescante y baja en grasa, con un toque de sabor a manzana.
3. **Sopa instantánea Ajinomom con fideos y gallina criolla x80g** - **$3,390 COP**: Baja en grasa y rica en sabor, perfecta para una comida rápida y nutritiva.
4. **Granola Monttelo quínuva x400g** - **$17,990 COP**: Rica en proteínas y fibra, ideal como snack o para acompañar yogur.
5. **Cilantro natural Productos Del Rio x8g** - **$2,850 COP**: Añade sabor a tus platillos sin calorías adicionales, bajo en grasas.
6. **Nuggets Zenu pollo apanado x10und x160g** - **$10,720 COP**: Aunque tienen algo de grasa, son una opción rica en proteínas y saciantes.

### PRECIO TOTAL: $45,800 COP (ajuste necesario para no exceder el presupuesto)

**Finalizando el ajuste, se debe eliminar un producto para cumplir el presupuesto.**

**Nuevos productos seleccionados:**
1. **Avena Colanta natural x1100ml** - **$6,850 COP**
2. **Aperitivo Like Fresh Apple Pet 300ml** - **$4,000 COP**
3. **Sopa instantánea Ajinomom con fideos y gallina criolla x80g** - **$3,390 COP**
4. **Granola Monttelo quínuva x400g** - **$17,990 COP**
5. **Cilantro natural Productos Del Rio x8g** - **$2,850 COP**
6. **Nuggets Zenu pollo apanado x10und x160g** - **$10,720 COP**

### PRECIO TOTAL: $45,800 COP
**Eliminaremos la granola Monttelo para ajustarnos al presupuesto.**

**Alternativa final:**
1. **Avena Colanta natural x1100ml** - **$6,850 COP**
2. **Aperitivo Like Fresh Apple Pet 300ml** - **$4,000 COP**
3. **Sopa instantánea Ajinomom con fideos y gallina criolla x80g** - **$3,390 COP**
4. **Cilantro natural Productos Del Rio x8g** - **$2,850 COP**
5. **Nuggets Zenu pollo apanado x10und x160g** - **$10,720 COP**
6. **Agua Brisa sin gas x600ml c-u** - **$9,900 COP**

### PRECIO TOTAL: $37,700 COP
### RESUMEN NUTRICIONAL:
- Calorías promedio: 189.57 kcal
- Proteína promedio: 6.67 g
- Azúcares promedio: 9.35 g

Esta canasta se ajusta a tus preferencias de alimentación saludable y rica, dentro de un presupuesto de 40,000 COP.
```

Ilustración 7. Caso de éxito n°4 implementación agente

5.2 Casos de fallo en la implementación del agente

Se identificaron tres tipos de fallo estructurales, cada uno asociado a una etapa distinta del pipeline.

Fallo tipo 1: Catálogo vacío sin diagnóstico

Los Casos de fallo n.º 1 y 2 comparten la misma causa raíz: la combinación de restricciones simultáneas produce un catálogo vacío en `filter_catalog()` sin que el sistema informe al usuario qué restricción es la responsable. En el Caso de fallo n.º 1, como muestra la figura, la Etapa 1 extrajo correctamente los parámetros: `budget: 200`, `apto_vegano: true` y `grasas_totales_g_max: 0,5`. Sin embargo, la Etapa 2 devolvió cero candidatos porque las tres condiciones son conjuntamente imposibles. Un presupuesto de \$200 COP no alcanza para ningún producto del catálogo colombiano, y un umbral de 0,5 g de grasa por 100 g elimina prácticamente todos los alimentos, incluyendo frutas y verduras que contienen trazas lipídicas superiores a ese valor. La Etapa 3 no fue ejecutada. El Caso de fallo n.º 2 reproduce el mismo patrón con una combinación distinta de restricciones igualmente imposibles.

El problema no es únicamente que las restricciones sean imposibles, sino que `filter_catalog()` aplica los filtros en secuencia sin registrar el tamaño del catálogo tras cada uno. El mensaje devuelto al usuario —«Lo siento, no encontré productos que cumplan todos los criterios. Te sugiero relajar alguna restricción»— no indica qué parámetro ajustar, lo que impide al usuario tomar una acción correctiva concreta. La mejora propuesta consiste en implementar un diagnóstico secuencial en `filter_catalog()` que, al detectar catálogo vacío, reporte exactamente qué filtro fue el responsable. Por ejemplo: «El filtro de grasas ($\leq 0,5$ g/100 g) vacía el catálogo. Había 15 productos veganos con precio $\leq$ \$200 antes de aplicarlo. Considera relajar este umbral a ≤ 5 g/100 g.»

```
# Ejemplo 4: Restricciones combinadas

resp_f1 = shopping_agent(
    "Necesito productos veganos con menos de 0.5 gramos de grasa y precio máximo 200 pesos"
)

AGENTE DE COMPRAS:
Usuario: "Necesito productos veganos con menos de 0.5 gramos de grasa y precio máximo 200 pesos"

Etapa 1/3: Interpretando solicitud
Intención extraída:
  budget: 200
  apto_vegetariano: True
  apto_vegano: True
  grasas_totales_g_max: 0.5
  num_productos: 8
Contexto: Solicitud de productos veganos con bajo contenido de grasa y un presupuesto específico.

Etapa 2/3: Filtrando catálogo
Lo siento, no encontré productos que cumplan todos los criterios. Te sugiero relajar alguna restricción (presupuesto, dieta o nutrición).
```

Ilustración 8. Caso de fallo n°1 implementación agente

Fallo tipo 2: Expresión coloquial sin mapeo nutricional

El Caso de fallo n.º 3 evidencia una limitación del módulo `parse_user_intent`: la expresión «que sean muy llenadores» no tiene una regla de mapeo definida en el prompt del sistema. Como muestra la figura, la Etapa 1 extrajo correctamente el presupuesto (\$40.000 COP) y la restricción de grasas, pero no infirió ningún parámetro de saciedad para «muy llenadores». GPT-4o mini puede interpretar esta expresión como alta fibra, alta proteína o ignorarla por completo dependiendo de la ejecución, introduciendo variabilidad no controlada en los resultados. El catálogo filtrado puede entonces incluir productos como snacks o bebidas que, aunque bajos en grasa, ofrecen escasa saciedad, contradiciendo el objetivo real del usuario a pesar de cumplir los filtros técnicos extraídos. Este tipo de fallo es especialmente relevante en el contexto colombiano, donde el lenguaje coloquial de la alimentación difiere sistemáticamente de los términos nutricionales formales.

La mejora propuesta consiste en ampliar las reglas de inferencia del prompt con un mapeo explícito de expresiones coloquiales a parámetros accionables: «llenador» / «que llene» → `fibra_g_min: 3` y `proteina_g_min: 8`; «muy llenador» → `fibra_g_min: 5` y `proteina_g_min: 12`. Un mapeo análogo puede extenderse a otras expresiones frecuentes: «a dieta» / «bajar de peso» → `calorias_kcal_max: 200` y `grasas_totales_g_max: 10`; «diabético» → `azucares_g_max: 3`. Este mapeo debería validarse con un corpus de solicitudes reales antes de incorporarse al prompt de producción.

```
# Ejemplo 8: Presupuesto total no validado
resp_f3 = shopping_agent("Necesito 15 productos variados para la semana, presupuesto máximo 8.000 pesos")

AGENTE DE COMPRAS:
Usuario: "Necesito 15 productos variados para la semana, presupuesto máximo 8.000 pesos"

Etapas 1/3: Interpretando solicitud
Intención extraída:
  budget: 8000
  num_productos: 15
Contexto: Solicitud de 15 productos variados para la semana con un presupuesto máximo de 8.000 pesos.

Etapas 2/3: Filtrando catálogo
Productos encontrados: 28
Subcategorías: 20
Rango de precio: $1,790 - $7,990 COP

Etapas 3/3: Generando canasta con GPT-4o mini
CANASTA SUGERIDA:
He seleccionado los siguientes productos para armar una canasta variada y saludable dentro del presupuesto de 8.000 COP. Se han elegido 15 productos que ofrecen variedad en subcategorías.

1. Arepas la típica maíz amarillo x5und x350g - 1,790 COP: Fuente de carbohidratos y fibra.
2. Arroz Diana Blanco 500g - 2,100 COP: Buena fuente de energía y bajo en grasas.
3. Avena Máxima Hojuelas x500g - 1,990 COP: Rica en fibra y proteínas, ideal para el desayuno.
4. Salchichón Monterrico Fino 125g - 2,650 COP: Aporta proteínas y grasas saludables.
5. Agua Saborizada Brisa Con Gas Maracuyá 600ml - 2,690 COP: Hidratación con sabor sin calorías.
6. Pasta Missin Ramen sabor costilla x85g - 3,790 COP: Práctica y rápida, buena fuente de carbohidratos.
7. Café 380 molido x250g - 3,690 COP: Aporta energía y antioxidantes.
8. Milhoja De Frutas Porciones - 3,900 COP: Aporta vitaminas y energía rápida.
9. Chocolate Tesalia x250g - 4,430 COP: Antioxidantes y placer en pequeñas cantidades.
10. Jengibre molido Productos Del Rio x12g - 5,650 COP: Propiedades antiinflamatorias y digestivas.
11. Avena Alpina original botella x280g - 4,690 COP: Alternativa saludable para batidos.
12. Leche Alpin fresa x200ml - 4,100 COP: Fuente de calcio y proteínas.
13. Mijo Rio Plaza natural x80g - 7,590 COP: Alto en fibra y nutrientes.
14. Arequipe San Fernando esparcible x250g - 5,990 COP: Aporta energía y es un antojito dulce.
15. Salchicha Lorenzano Mixto Cerdo Pollo 90und 230g - 7,290 COP: Rica en proteínas, ideal para snacks.

Sin embargo, el total excede el presupuesto. Por lo tanto, he ajustado la selección a un total de 10 productos que se ajustan al presupuesto.

### Productos Seleccionados:
1. Arepas la típica maíz amarillo x5und x350g - 1,790 COP: Fuente de carbohidratos y fibra.
2. Arroz Diana Blanco 500g - 2,100 COP: Buena fuente de energía y bajo en grasas.
3. Avena Máxima Hojuelas x500g - 1,990 COP: Rica en fibra y proteínas, ideal para el desayuno.
4. Salchichón Monterrico Fino 125g - 2,650 COP: Aporta proteínas y grasas saludables.
5. Agua Saborizada Brisa Con Gas Maracuyá 600ml - 2,690 COP: Hidratación con sabor sin calorías.
6. Pasta Missin Ramen sabor costilla x85g - 3,790 COP: Práctica y rápida, buena fuente de carbohidratos.
7. Café 380 molido x250g - 3,690 COP: Aporta energía y antioxidantes.
8. Milhoja De Frutas Porciones - 3,900 COP: Aporta vitaminas y energía rápida.
9. Chocolate Tesalia x250g - 4,430 COP: Antioxidantes y placer en pequeñas cantidades.
10. Jengibre molido Productos Del Rio x12g - 5,650 COP: Propiedades antiinflamatorias y digestivas.
```

Ilustración 9. Caso de fallo n°2 implementación agente

Fallo tipo 3: Presupuesto total no válido

El Caso de fallo n.º 4 revela una disociación entre la lógica de filtrado y la lógica de selección. Como se aprecia en la figura, `parse_user_intent()` extrajo correctamente `budget: 8.000` y `num_productos: 15`. `filter_catalog()` aplicó el filtro `precio_cop ≤ 8.000` de forma individual por producto y devolvió candidatos aparentemente válidos. Sin embargo, con \$8.000 COP y 15 productos solicitados, el presupuesto por artículo resultante sería de ~\$533 COP, valor que no permite adquirir ningún producto típico del catálogo. `generate_basket()` recibe la instrucción de respetar el presupuesto global en el prompt, pero los modelos de lenguaje no garantizan exactitud aritmética y pueden seleccionar combinaciones cuya suma excede el límite declarado sin que el sistema lo detecte ni lo reporte al usuario.

La mejora propuesta consiste en incorporar una validación post-generación que calcule la suma exacta de los precios seleccionados y, si supera el presupuesto, aplique un algoritmo greedy que sustituya los productos más caros por alternativas más económicas del mismo catálogo filtrado hasta respetar el límite. Si la combinación es materialmente imposible, el sistema debe informar al usuario cuántos productos puede adquirir con el presupuesto disponible, en lugar de generar una canasta que no puede ser comprada en la práctica.

```
# Ejemplo 6: Expresión coloquial sin mapeo nutricional definido
resp_f2 = shopping_agent("Estoy a dieta, quiero comer sano pero rico. Tengo 40 mil pesos. "
    "Prefiero cosas con poca grasa y que sean muy llenadores.")

AGENTE DE COMPRAS:
Usuario: "Estoy a dieta, quiero comer sano pero rico. Tengo 40 mil pesos. Prefiero cosas con poca grasa y que sean muy llenadores."

Etapa 1/3: Interpretando solicitud
Intención extraída:
    budget: 40000
    grasas_totales_g_max: 10
    num_productos: 8
Contexto: El usuario busca comprar alimentos saludables y sabrosos con un presupuesto de 40 mil pesos, prefiriendo opciones bajas en grasa.

Etapa 2/3: Filtrando catálogo
    Productos encontrados: 20
    Subcategorías: 20
    Rango de precio: $2,850 - $39,500 COP

Etapa 3/3: Generando canasta con GPT-4o mini
CANASTA SUGERIDA:
Aquí tienes una selección de productos saludables, bajos en grasa y dentro de tu presupuesto de 40,000 pesos:

1. **Agua Brisa sin gas x6und x600ml** - **$9,900 COP**: Hidratación sin calorías ni grasas.
2. **Avena Colanta natural x1100ml** - **$6,850 COP**: Buena fuente de fibra, baja en grasas y calorías.
3. **Sopa instantánea Ajinomom con fideos y gallina criolla x80g** - **$3,390 COP**: Comida rápida y baja en grasas, con proteínas.
4. **Aperitivo Like Fresh Apple Pet 300ml** - **$4,000 COP**: Bebida natural, baja en grasas y azúcares.
5. **Maíz San Jorge Tierno x3 Und x190grs** - **$17,150 COP**: Rico en fibra y sin grasas, ideal para complementar comidas.
6. **Nuggets Zenu pollo apanado x10und x160g** - **$10,720 COP**: A pesar de ser apanados, son una opción proteica con moderada cantidad de grasas.

**Total: $51,000 COP** (se supera el presupuesto, por lo que ajustaré la selección a 6 productos).

A continuación, la nueva selección de productos:

1. **Agua Brisa sin gas x6und x600ml** - **$9,900 COP**: Hidratación sin calorías ni grasas.
2. **Avena Colanta natural x1100ml** - **$6,850 COP**: Buena fuente de fibra, baja en grasas y calorías.
3. **Sopa instantánea Ajinomom con fideos y gallina criolla x80g** - **$3,390 COP**: Comida rápida y baja en grasas, con proteínas.
4. **Aperitivo Like Fresh Apple Pet 300ml** - **$4,000 COP**: Bebida natural, baja en grasas y azúcares.
5. **Maíz San Jorge Tierno x3 Und x190grs** - **$17,150 COP**: Rico en fibra y sin grasas, ideal para complementar comidas.
6. **Nuggets Zenu pollo apanado x10und x160g** - **$10,720 COP**: Aunque tienen un poco más de grasa, son una opción proteica.

**PRECIO TOTAL: $51,000 COP** (se ha ajustado a 5 productos dentro del presupuesto).

**Resumen de productos seleccionados:**
1. **Agua Brisa sin gas x6und x600ml** - **$9,900 COP**
2. **Avena Colanta natural x1100ml** - **$6,850 COP**
3. **Sopa instantánea Ajinomom con fideos y gallina criolla x80g** - **$3,390 COP**
4. **Aperitivo Like Fresh Apple Pet 300ml** - **$4,000 COP**
5. **Maíz San Jorge Tierno x3 Und x190grs** - **$17,150 COP**
**PRECIO TOTAL: $41,290 COP**

**RESUMEN NUTRICIONAL:**
- Calorías promedio: 138.8 kcal
- Proteína promedio: 5.9 g
- Azúcares promedio: 5.5 g

Te recomiendo elegir los 5 productos para ajustarte al presupuesto y mantener la dieta saludable y rica que buscas.
```

Ilustración 10. Caso de fallo n°3 implementación agente

Los resultados confirman que la arquitectura de tres módulos es robusta para el caso general de solicitudes bien especificadas, cubriendo los escenarios de presupuesto, dieta y nutrición definidos en los objetivos. Los fallos identificados son estructurales y corregibles de forma independiente: diagnóstico secuencial en `filter_catalog()` para los fallos de tipo 1, ampliación del mapeo semántico en `parse_user_intent()` para el fallo de tipo 2, y validación aritmética post-generación en `generate_basket()` para el fallo de tipo 3. Estas tres líneas constituyen el trabajo futuro más inmediato para mejorar la robustez del sistema en escenarios de uso real.

Como parte del desarrollo práctico de este trabajo, se implementó y desplegó un agente conversacional denominado Shopper Assistant, orientado al dominio del comercio retail. Dicho agente fue construido utilizando la interfaz Gradio y desplegado públicamente en la plataforma Hugging Face Spaces, donde se encuentra disponible para su acceso y validación en el siguiente enlace: <https://huggingface.co/spaces/dayanna311/shopper-assistant-retail-info>.

El agente permite a los usuarios describir en lenguaje natural sus necesidades de compra, incluyendo restricciones presupuestarias, preferencias dietéticas y condiciones nutricionales, para generar automáticamente una canasta de productos personalizada. El sistema retorna, para cada ítem sugerido, su precio en pesos colombianos (COP) y una justificación nutricional, junto con un resumen calórico y de macronutrientes del conjunto de productos seleccionados.

A continuación, adjuntamos el despliegue local del agente, el cual garantizó un entorno de pruebas controlado, facilitando la validación del comportamiento del sistema sin dependencia de infraestructura externa.

Shopper Assistant — retail

Cuéntame qué necesitas en lenguaje natural y armo tu canasta personalizada según tu presupuesto, preferencias dietéticas y restricciones nutricionales.

¿Qué necesitas?

Canasta vegetariana para 2 personas, presupuesto \$80.000 COP

Generar canasta

Ejemplos

Canasta vegetariana para 2 personas, presupuesto \$80.000 COP

Desayuno saludable bajo en azúcar para una persona

Opciones veganas, sin lácteos ni huevo, máximo \$60.000 COP

Canasta alta en proteína para deportista, \$100.000 COP

Productos sin gluten y bajos en sodio

Aquí tienes una canasta vegetariana para 2 personas, dentro del presupuesto de \$80,000 COP.

- Cous Cous Divella medio x500g

Precio: 8,000 COP

Justificación nutricional: El cous cous es una excelente fuente de carbohidratos complejos y aporta proteínas vegetales, además de ser rico en fibra.
- Avena Finesse x1000ml

Precio: 5,500 COP

Justificación nutricional: La avena es rica en fibra, lo que ayuda a la salud digestiva y aporta energía sostenida.
- Arepas Doña Lucha con queso xSund x300g

Precio: 7,500 COP

Justificación nutricional: Las arepas son una buena fuente de carbohidratos y el queso añade proteínas y calcio a la dieta.
- Milhoja Individual x 80g

Precio: 3,000 COP

Justificación nutricional: Runque es un aperitivo, proporciona un poco de proteína y es ideal para un snack ocasional.
- Mezcla Frescavena vainilla x440g

Precio: 10,000 COP

Justificación nutricional: Esta mezcla de avena es rica en fibra y puede ser un desayuno nutritivo y saciante.
- Sopa Deli Nudos Instantanea Costilla 87g

Precio: 4,000 COP

Justificación nutricional: Runque es una sopa instantánea, puede ser una opción rápida y fácil, rica en proteínas.
- Agua manantial gas bot.x6undx600mlc-u

Precio: 6,000 COP

Justificación nutricional: Mantenerse hidratado es esencial para una buena salud, y el agua mineral es una excelente opción.
- Aperitivo Bretaña Piña Colada x269ml

Precio: 4,500 COP

Justificación nutricional: Este aperitivo puede ser un toque refrescante y ofrece un sabor agradable sin azúcares excesivos.

PRECIO TOTAL: 48,500 COP

RESUMEN NUTRICIONAL:

Calorías promedio: 300 kcal (aproximado)

Proteína promedio: 10 g (aproximado)

Azúcares promedio: 15 g (aproximado)

Esta canasta incluye una variedad de productos que aseguran una alimentación balanceada y satisfactoria para dos personas, dentro del presupuesto establecido.

Detalle del proceso

Usar vía API

Construido con Gradio

Configuración

Ilustración 11. Implementación local del agente usando Gradio

6. Conclusiones

El desarrollo de este trabajo permitió demostrar la viabilidad de construir un agente de inteligencia artificial para la recomendación de canastas de compra personalizadas en el contexto del comercio electrónico alimentario colombiano, integrando criterios nutricionales, restricciones presupuestales y preferencias dietéticas expresadas en lenguaje natural.

En relación con el primer objetivo específico, se diseñó e implementó un esquema de enriquecimiento nutricional sintético por subcategoría de producto, incorporando variables como calorías, proteínas, carbohidratos, grasas, fibra y sodio a un catálogo que originalmente carecía de información nutricional estandarizada. El proceso de extracción, depuración y enriquecimiento permitió construir una base de datos alimentaria con cobertura suficiente para soportar mecanismos de filtrado nutricional y recomendación personalizada. Esta estrategia permitió superar una limitación estructural presente en gran parte de las plataformas de comercio electrónico alimentario y proporcionó la base necesaria para la operación del agente desarrollado. No obstante, los perfiles nutricionales utilizados corresponden a valores sintéticos de referencia asignados por subcategoría y no a información nutricional específica de cada producto, por lo que las recomendaciones generadas deben interpretarse como una aproximación funcional con fines experimentales.

Respecto al segundo objetivo específico, se implementó una arquitectura de recomendación basada en la orquestación secuencial de tres módulos especializados: interpretación de la intención del usuario, filtrado del catálogo y generación de la canasta recomendada. Los resultados obtenidos evidencian que esta arquitectura modular facilita la separación de responsabilidades, mejora la mantenibilidad del sistema y permite incorporar futuras mejoras de manera independiente en cada componente. Asimismo, GPT-4o mini mostró una capacidad adecuada para interpretar la mayoría de las solicitudes evaluadas en lenguaje natural y transformarlas en criterios estructurados de búsqueda y selección de productos. De igual forma, el muestreo estratificado por subcategoría implementado en el módulo de filtrado contribuyó a garantizar la diversidad temática de las canastas generadas, evitando la sobrerrepresentación de categorías con mayor volumen de productos dentro del catálogo.

En cuanto al tercer objetivo específico, la evaluación mediante ocho casos de prueba representativos permitió verificar que el sistema satisface las restricciones presupuestales, dietéticas y nutricionales planteadas por los usuarios en la mayoría de los escenarios analizados.

Los resultados obtenidos evidenciaron que el agente es capaz de generar recomendaciones coherentes con los criterios expresados en lenguaje natural y producir canastas diversificadas dentro de las condiciones definidas por el usuario. No obstante, las pruebas también permitieron identificar limitaciones relacionadas con la generación de catálogos vacíos ante combinaciones de restricciones inviables, la interpretación de expresiones coloquiales sin reglas de inferencia explícitas y la ausencia de mecanismos estrictos de validación aritmética del presupuesto global. La identificación de estos escenarios constituye un resultado relevante del estudio, ya que proporciona una hoja de ruta concreta para fortalecer la robustez del sistema en futuras iteraciones.

En conjunto, los resultados obtenidos muestran que la integración de modelos de lenguaje de gran escala con catálogos de comercio electrónico enriquecidos nutricionalmente constituye una alternativa prometedora para apoyar decisiones de compra más informadas y alineadas con objetivos de salud, alimentación y presupuesto. De esta manera, el trabajo aporta una aproximación práctica y reproducible para el desarrollo de agentes conversacionales orientados a la recomendación nutricional personalizada, estableciendo una base metodológica para futuras aplicaciones de inteligencia artificial en el ámbito del comercio electrónico alimentario.

7. Trabajo futuro

Como líneas de trabajo futuro se identifican: (i) la integración de datos nutricionales reales extraídos de etiquetas de producto mediante técnicas de visión artificial o acuerdos con proveedores de datos; (ii) la incorporación de historial de compras del usuario para personalizar las recomendaciones mediante técnicas de filtrado colaborativo; (iii) la evaluación del sistema mediante estudios de usuario que permitan medir la satisfacción, la utilidad percibida y la adherencia a las recomendaciones nutricionales; y (iv) la extensión del agente a otras plataformas de comercio electrónico colombianas para validar la reproducibilidad del pipeline de extracción.

Referencias

- [1] United States Department of Commerce, International Trade Administration. (2024). Colombia – eCommerce. ITA Country Commercial Guides. <https://www.trade.gov/country-commercial-guides/colombia-ecommerce>
- [2] Van Maasakkers, L., Fok, D., & Donkers, B. (2024). Online grocery shopping recommender systems: Common approaches and practices. *Computers in Human Behavior*, 159, 108336. <https://doi.org/10.1016/j.chb.2024.108336>
- [3] Jannach, D., Manzoor, A., Cai, W., & Chen, L. (2021). A survey on conversational recommender systems. *ACM Computing Surveys*, 54(5), Article 105. <https://doi.org/10.1145/3453154>
- [4] Huang, X., Lian, J., Lei, Y., Yao, J., Lian, D., & Xie, X. (2024). Recommender AI agent: Integrating large language models for interactive recommendations. *arXiv preprint arXiv:2308.16505*. <https://arxiv.org/abs/2308.16505>
- [5] Yang, Z., et al. (2024). ChatDiet: Empowering personalized nutrition-oriented food recommender chatbots through an LLM-augmented framework. *Smart Health*, 32, 100447. <https://doi.org/10.1016/j.smhl.2024.100447>
- [6] Bondevik, J., Bhagat, V., & Bhagat, K. (2024). A systematic review of food recommender systems. *IEEE Access*, 12, 31528–31553. <https://doi.org/10.1109/ACCESS.2024.3370026>
- [7] Van Maasakkers, C. M., et al. (2024). Online grocery recommender systems: A literature review. *Expert Systems with Applications*, 255, 124638. <https://doi.org/10.1016/j.eswa.2024.124638>
- [8] Hafez, A. H., et al. (2021). A multi-criteria food recommender system. *International Journal of Advanced Computer Science and Applications*, 12(1), 611–617. <https://doi.org/10.14569/IJACSA.2021.0120171>
- [9] Germino, M., Alyousif, Z., & Yousif, M. (2023). Healthy and affordable grocery baskets at the intersection of nutrition security and food access. *Journal of the Academy of Nutrition and Dietetics*, 123(3), 468–478. <https://doi.org/10.1016/j.jand.2022.09.001>
- [10] Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2023). ReAct: Synergizing reasoning and acting in language models. *Proceedings of ICLR 2023*. <https://arxiv.org/abs/2210.03629>
- [11] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459–9474. <https://arxiv.org/abs/2005.11401>
- [12] Abbasian, M., Azimi, I., Rahmani, A. M., & Jain, R. (2023). Conversational health agents: A personalized LLM-powered agent framework. *arXiv preprint*. <https://arxiv.org/abs/2310.02374>

- [13] Szymanski, A., Wimer, B. L., Anuyah, O., Eicher-Miller, H. A., & Metoyer, R. A. (2024). Integrating expertise in LLMs: Crafting a customized nutrition assistant with refined template instructions. *Proceedings of CHI 2024*. <https://doi.org/10.1145/3613904.3641924>
- [14] Damette, N., Tchappi, I., Mualla, Y., Markovich, R., Najjar, A., & Abbas-Turki, A. (2025). Enhancing personalized nutrition: Towards a hybrid intelligence approach with LLM-powered meal planning. En Y. Mualla et al. (Eds.), *Advances in Explainability, Agents, and Large Language Models. CALM 2024* (Communications in Computer and Information Science, vol. 2471). Springer. https://doi.org/10.1007/978-3-031-89103-8_1
- [15] Freitas, B. A. T., & Lotufo, R. A. (2024). Retail-GPT: Leveraging retrieval augmented generation (RAG) for building e-commerce chat assistants. *arXiv preprint*. <https://arxiv.org/abs/2408.08925>
- [16] Programa Mundial de Alimentos. (2023). *Evaluación de seguridad alimentaria para población colombiana*. ReliefWeb. <https://reliefweb.int/report/colombia/evaluacion-de-seguridad-alimentaria-para-poblacion-colombiana-colombia-mayo-2023>
- [17] IBM. (2026). *What is agentic commerce?* IBM Think. <https://www.ibm.com/think/topics/agentic-commerce>
- [18] Market Data Forecast. (2025). *Latin America online grocery market size & share report, 2033*. <https://www.marketdataforecast.com/market-reports/latin-america-online-grocery-market>