ARQUITECTURAS SUBAGÉNTICAS DE IA: HACIA LA ESPECIALIZACIÓN Y LA ORQUESTACIÓN JERÁRQUICA

Olda Bustillos Ortega
https://orcid.org/0000-0003-2822-3428
obustillos@uia.ac.cr
Escuela de Ingeniería Informática,
Universidad Internacional de las Américas, Costa Rica

Jorge Murillo Gamboa
https://orcid.org/0000-0001-5548-8283
jmurillo@uia.ac.cr
Escuela de Ingeniería Informática,
Universidad Internacional de las Américas, Costa Rica

Olman Núñez Peralta
https://orcid.org/0000-0001-6780-022X
onunez@edu.uia.ac.cr
Escuela de Ingeniería Informática,
Universidad Internacional de las Américas, Costa Rica

Fabián Rodríguez Sibaja
https://orcid.org/0009-0008-3276-9865
frodriguezsi@edu.uia.ac.cr
Escuela de Ingeniería Informática,
Universidad Internacional de las Américas, Costa Rica

Daniel Mena Bocker
https://orcid.org/0009-0002-1062-6675
dfmenab@edu.uia.ac.cr
Escuela de Ingeniería Informática,
Universidad Internacional de las Américas, Costa Rica

Recibido: 27 de marzo del 2026 / Aceptado: 18 de abril del 2026
doi: https://doi.org/10.26439/interfases2026.n023.8714

RESUMEN. Durante el segundo semestre del 2025, la investigación en inteligencia artificial (IA) ha evidenciado una transición hacia arquitecturas multiagente, en las cuales los subagentes adquieren un rol estructural central. A diferencia de los enfoques monolíticos, estas arquitecturas permiten descomponer tareas complejas en unidades funcionales

especializadas, coordinadas por un agente orquestador, lo que favorece la escalabilidad, el paralelismo y una gestión más eficiente del contexto. Se definen los sistemas multiagente, agentes web y subagentes. Se analiza el marco conceptual *multi-agent data mining* (MADM), la arquitectura del árbol de tareas complejas y la metáfora de la cocina. Se identifican beneficios en términos de eficiencia, resiliencia y adaptabilidad, tanto en entornos científicos como en plataformas comerciales. No obstante, persisten desafíos asociados a la orquestación, la seguridad, los costos computacionales, formación académica y la gobernanza ética. En este contexto, se examina el Índice Latinoamericano de Inteligencia Artificial (ILIA 2025) como guía para integrar arquitecturas subagénticas a modelos educativos con enfoque ético, interdisciplinario y sostenible, orientadas a impulsar una innovación regional más responsable e inclusiva.

PALABRAS CLAVE: agentes / arquitecturas / ILIA / inteligencia artificial / subagentes

SUB-AGENTIC AI ARCHITECTURE: A CONCEPTUAL REVIEW OF SPECIALIZATION AND HIERARCHICAL ORCHESTRATION

ABSTRACT. During the second half of 2025, artificial intelligence (AI) research has shown a shift towards multi-agent architectures, in which sub-agents acquire a central structural role. Unlike monolithic approaches, these architectures allow complex tasks to be broken down into specialized functional units, coordinated by an orchestrating agent, which promotes scalability, parallelism, and more efficient context management. This article defines multi-agent architectures, web agents, and sub-agents. The conceptual framework Multi-Agent Data Mining (MADM), the architecture of the complex task tree, and the kitchen metaphor are analyzed. Benefits in terms of efficiency, resilience, and adaptability are identified, both in scientific environments and on commercial platforms. However, challenges remain related to orchestration, security, computational costs, academic training, and ethical governance. In this context, the Latin American Artificial Intelligence Index (ILIA 2025) is examined as a guide for integrating sub-agentic architectures into educational models with an ethical, interdisciplinary, and sustainable focus, aimed at fostering more responsible and inclusive regional innovation.

KEYWORDS: agents / architecture / artificial inteligence / ILIA / subagents

INTRODUCCIÓN

Multiagentes

En el ámbito de las ciencias de la computación, los sistemas multiagente (*multi-agent data mining*, MAS) constituyen una subárea de la inteligencia artificial (IA) cuyo propósito es resolver problemas que superan las capacidades de un solo agente. Estos sistemas se han consolidado como uno de los campos más relevantes de investigación en IA, ya que permiten abordar tareas complejas mediante la cooperación, la coordinación o, incluso, la competencia entre agentes, lo que permite una mayor eficacia y adaptabilidad del sistema en su conjunto (Elmahalawy, 2015; Gonçalves et al., 2015).

De acuerdo con Carrascosa et al. (2005), un sistema MAS puede adoptar distintos comportamientos o asumir diferentes roles. En el ejemplo de un entorno típico de comercio electrónico, un mismo agente puede actuar tanto como comprador o como vendedor, alternando entre estos dos comportamientos en función del tipo de transacción, incluso cuando sus objetivos y tareas resulten opuestos o potencialmente contradictorios.

Una de las principales razones de este enfoque radica en la creciente necesidad de automatizar procesos complejos, que abarcan desde la investigación científica y el análisis de grandes volúmenes de datos hasta la atención al cliente. Plataformas recientes evidencian que la coordinación jerárquica de agentes, mediante orquestadores que supervisan conjuntos de subagentes especializados, permite superar limitaciones asociadas al manejo del contexto, la latencia y la eficiencia cognitiva de los modelos.

De acuerdo con Rother et al. (2025), y respecto del entrenamiento de estos multiagentes, surgió el aprendizaje por refuerzo multiagente (*multi-agent reinforcement learning*, MARL), aplicado en entornos con agentes previamente definidos y a través de tareas principalmente cooperativas. El aprendizaje se apoyó inicialmente en esquemas de entrenamiento o control centralizado, aunque con un alcance limitado. Posteriormente, se desarrollaron enfoques descentralizados que permitieron a cada agente aprender su propia política, como el Independent Q-Learning. No obstante, la mayoría de estas técnicas asumen conjuntos de agentes con tareas relativamente fijas. En contraste, enfoques más recientes plantean la incorporación o eliminación dinámica de agentes y la adaptación a tareas cambiantes, lo que amplía el alcance del MARL.

WebAgents o agentes web

A medida que la World Wide Web (WWW), o red informática mundial, evoluciona, el acceso a la información y a los servicios digitales se transforma de forma sustancial. Actividades cotidianas como realizar compras y pagos en línea, registrar nuevas cuentas o completar formularios requieren la introducción reiterada de datos personales, lo que incrementa la carga operativa de los usuarios. En respuesta a este escenario, han surgido

los WebAgents como sistemas basados en técnicas de IA diseñados para automatizar tareas rutinarias mediante la emulación de procesos cognitivos humanos, con el objetivo de optimizar la productividad y mejorar la calidad de vida.

De acuerdo con Ning et al. (2025), la interacción entre el usuario y un WebAgent puede realizarse mediante una única instrucción en lenguaje natural, por ejemplo, solicitar una reunión: "Programa una reunión con León en Starbucks el día 23 de noviembre de 2025 a las 4:00 p. m. mediante correo electrónico". Estos agentes son capaces de ejecutar de manera autónoma flujos completos de trabajo, desde obtener la dirección electrónica de León, redactar el mensaje, incluir la reunión en la agenda y, finalmente, enviarlo por correo. El WebAgent lleva a cabo la interacción con aplicaciones hasta la ejecución de acciones finales, lo que incrementa significativamente la eficiencia y la comodidad del usuario. No obstante, su adopción plantea desafíos asociados a la transparencia, la propagación de sesgos, la protección de la privacidad y el desempeño frente a situaciones imprevistas.

De acuerdo con Chevrot et al. (2025), los agentes autónomos web (*autonomous web agents*, AWA) constituyen una alternativa prometedora para reducir la fragilidad y el alto costo de mantenimiento de la automatización de tareas tradicionales. Estos agentes interactúan con aplicaciones web mediante ciclos iterativos de observación, análisis y acción guiados por descripciones de tareas en lenguaje natural, sin requerir conocimiento previo del sistema.

Mediante la construcción y el ajuste dinámico de *prompts*, los AWA son capaces de interpretar los escenarios de prueba como tareas autónomas, ejecutar las acciones correspondientes y aplicar modificaciones de manera adaptativa. Esta capacidad de interpretación contribuye a fortalecer la robustez y la flexibilidad de los procesos al reducir la dependencia de *scripts* estáticos y mitigar la vulnerabilidad de estos frente a cambios en la interfaz o en la lógica de las aplicaciones.

Subagentes de IA

Las arquitecturas subagénticas de IA pueden entenderse como aquellos sistemas distribuidos compuestos por agentes especializados que cooperan tanto en los procesos de toma de decisiones como en la ejecución de tareas y que están articulados en torno a modelos de lenguaje LLM (*large language models*). Estas arquitecturas se fundamentan en la delegación estructurada de tareas desde un agente orquestador hacia subagentes especializados, los cuales operan de manera autónoma y colaborativa para cumplir objetivos específicos dentro del sistema distribuido.

De acuerdo con Claude Code Docs (2025), los subagentes posibilitan una resolución de problemas más eficiente al operar con configuraciones de tareas específicas que incluyen indicaciones de sistema personalizadas, herramientas dedicadas y ventanas de

contexto independientes. En plataformas de programación como Claude Code[1], estos subagentes se asemejan a asistentes de IA especializados que pueden ser invocados para gestionar tipos concretos de tareas.

A modo de ejemplo, cuando el orquestador de Claude Code identifica una tarea alineada con la especialización de un subagente, procede a delegarla a dicho subagente, el cual ejecuta de manera autónoma las acciones necesarias y retorna los resultados correspondientes. Entonces, una vez preconfigurados, los subagentes reciben encargos definidos y ejecutan de manera autónoma las acciones correspondientes dentro de su ámbito de especialización:

- Tienen un propósito específico y un área de experiencia funcional.
- Utilizan su propia ventana de contexto separada de la conversación principal.
- Pueden configurarse para que utilice herramientas específicas.
- Incluyen un indicador personalizado que guía su comportamiento.

Este enfoque de subagentes de IA plantea una mejora en el rendimiento técnico y favorece la modularidad, escalabilidad y adaptabilidad de los sistemas, lo que posiciona a los agentes y subagentes como un recurso estratégico para el desarrollo productivo. Sin embargo, su impacto efectivo depende de la adecuada gestión de desafíos asociados a la orquestación, la seguridad y la gobernanza ética.

Para organizar e integrar las funciones de subagentes de IA, Zhu et al. (2025) proponen tres categorías: identificadores de agentes, monitoreo en tiempo real y registro de actividades. Estas están destinadas a incrementar la visibilidad operativa de los sistemas de IA, medidas que pueden interpretarse como funciones especializadas dentro de una arquitectura subagéntica.

En una arquitectura jerárquica, las funciones de identificación, monitoreo y registro pueden delegarse a subagentes especializados, lo que permite una supervisión coordinada y una rendición de cuentas efectiva. Esta distribución de funciones permite coordinar de manera coherente la supervisión, el control y la rendición de cuentas del sistema, lo que refuerza la tesis de que las arquitecturas subagénticas, al asignar responsabilidades específicas a módulos coordinados jerárquicamente, ofrecen un marco robusto para la transparencia, la seguridad y la gestión fiable del comportamiento de agentes autónomos.

[1] Claude Code es un entorno especializado de programación y asistencia al desarrollo creado por Anthropic con IA enfocada en tareas de codificación, depuración, análisis de archivos y ejecución controlada de código, similar a un espacio de trabajo optimizado para programadores. Para visitar la página web, puede ingresar aquí: https://website.claude.com/product/overview

Aplicaciones con subagentes de IA

- Movilidad y transporte. De acuerdo con Samdani et al. (2025), la aplicación de subagentes de IA en movilidad y transporte, especialmente en vehículos autónomos, ha transformado la interacción entre sistemas inteligentes, infraestructura digital y seres humanos, con el objetivo principal de reducir el error humano, una de las causas predominantes de accidentes viales. En este contexto, los subagentes permiten la especialización funcional en tareas como percepción del entorno, razonamiento, planificación y control, integrando enfoques basados en reglas y aprendizaje automático dentro de arquitecturas distribuidas.
- Inversiones financieras. De acuerdo con Babaei y Giudici (2025), las arquitecturas subagénticas permiten mejorar el análisis de inversiones financieras mediante la coordinación de agentes especializados, lo que aumenta la trazabilidad, la explicabilidad y la precisión de las decisiones. En este enfoque jerárquico, el modelo principal de riesgo-retorno se complementa con módulos explicativos basados en valores. Así, la especialización funcional y la coordinación multinivel incrementan tanto la precisión como la transparencia, lo que consolida a las arquitecturas subagénticas como un marco relevante para la toma de decisiones financieras robustas y verificables.
- Sector de salud. De acuerdo con Samdani et al. (2025), en el ámbito de la salud, las arquitecturas subagénticas de IA pueden potenciar las capacidades del personal clínico y administrativo mediante la especialización de tareas clínicas y no clínicas. No obstante, debido a la sensibilidad del dominio sanitario, su adopción exige una estrecha colaboración interdisciplinaria y una gobernanza ética sólida que garantice un uso responsable de sistemas híbridos entre lo humano y la IA.
- Sector académico. En este sector, al introducir estas arquitecturas de multiagente, WebAgents y subagentes, los estudiantes de distintos niveles, desde la secundaria hasta la universitaria, tienen acceso a plataformas educativas adaptativas, lo que reduce las brechas de aprendizaje y potencia el desarrollo de competencias en áreas como Matemáticas, Programación o Ciencias Aplicadas. Los docentes pueden apoyarse en agentes y subagentes para automatizar sus tareas administrativas, tales como generar material didáctico individualizado, analizar múltiples referencias y recibir recomendaciones de parte de un agente de IA en torno a estrategias pedagógicas basadas en datos del desempeño estudiantil (Babaei & Giudici, 2025).

En síntesis, las arquitecturas subagénticas de IA representan un modelo evolutivo de inteligencia distribuida, capaz de potenciar la productividad y la sostenibilidad, siempre

que su desarrollo se acompañe de estrategias coordinadas en orquestación técnica, seguridad digital, financiamiento de la infraestructura y regulación ética.

En este artículo, se examina el marco conceptual del *multi-agent data mining framework* (MADM), el cual integra técnicas de minería de datos (*data mining*, DM) con sistemas multiagente (*multi-agent systems*, MAS), con el objetivo de automatizar, distribuir y potenciar el proceso de descubrimiento de conocimiento en grandes volúmenes de información.

Otra referencia utilizada para ilustrar la dinámica entre agentes y subagentes de IA es la metáfora de la brigada de cocina profesional. En este contexto, el *chef de cuisine* asume la responsabilidad de coordinar de manera integral las actividades de la brigada, por lo que desempeña un rol análogo al del agente orquestador dentro de una arquitectura de IA. Por su parte, Los *chefs de partie* representan a los agentes especializados encargados de áreas o tareas específicas, mientras que los roles de apoyo —como los *cuisiniers* y *commis*— pueden asociarse a subagentes que ejecutan funciones operativas de asistencia bajo un esquema de colaboración jerárquica.

El árbol de tareas complejas constituye otra referencia conceptual —la cual será analizada más adelante— que permite comprender el ciclo de búsqueda de información (*information retrieval*) orientado a la resolución de tareas de alta complejidad. Este enfoque modela dicho proceso mediante una estructura jerárquica en forma de árbol, sustentada en el uso coordinado de agentes y subagentes de IA.

Asimismo, se incluye la iniciativa del Índice Latinoamericano de Inteligencia Artificial 2025 (ILIA 2025) como un esfuerzo pionero orientado a medir de forma sistemática el avance de la inteligencia artificial en América Latina y el Caribe. La experiencia de los países de la región sugiere que la articulación entre talento humano, una institucionalidad sólida y la adopción responsable de la IA puede posicionar a América Latina como un entorno favorable para el desarrollo de sistemas de IA escalables, transparentes y orientados al bien común.

OBJETIVOS DE LA INVESTIGACIÓN

El objetivo principal fue analizar el desarrollo y la relevancia de las arquitecturas subagénticas de inteligencia artificial, destacando sus fundamentos conceptuales, aplicaciones emergentes, desafíos técnicos y consideraciones éticas, con el fin de comprender su impacto en la evolución de los sistemas agénticos y en la automatización de procesos complejos.

Por su parte, los objetivos específicos de la investigación son los siguientes:

- Definir el concepto de subagente y diferenciarlo de los modelos de agente único y de las arquitecturas multiagente tradicionales

- Examinar los principales enfoques arquitectónicos que sustentan a los subagentes, incluyendo jerarquía, paralelismo e interacción colaborativa
- Identificar aplicaciones prácticas de los subagentes en investigación científica, educación, industria y robótica
- Analizar los desafíos técnicos asociados a la coordinación, eficiencia computacional, seguridad y gobernanza de los sistemas subagénticos
- Explorar las perspectivas futuras de la investigación en subagentes y su potencial integración con otras tecnologías disruptivas

Asimismo, se buscó responder a la siguiente pregunta de investigación: ¿cómo contribuyen las arquitecturas subagénticas de IA a impulsar el desarrollo productivo, considerando los desafíos en la orquestación, la seguridad y la gobernanza ética?

METODOLOGÍA

La presente investigación se desarrolló mediante un enfoque de revisión sistemática de la literatura, orientado a analizar las arquitecturas subagénticas en IA y su impacto en diversos sectores de aplicación. El proceso metodológico se diseñó siguiendo lineamientos de revisiones en ingeniería y ciencias computacionales, a fin de garantizar transparencia y reproducibilidad.

La estrategia de búsqueda de información se realizó entre enero del 2020 y marzo del 2026, con el objetivo de capturar los avances más recientes en el área. Para ello, se consultaron bases de datos académicas indexadas de alto impacto, incluyendo ACM Digital Library, IEEE Xplore, Google Académico, ProQuest Dissertations and Theses y Academia.edu.

Las ecuaciones de búsqueda se construyeron combinando operadores booleanos y términos clave en español y en inglés:

- “sub-agent architectures” OR “multi-agent systems”
- “AI agents” AND “coordination”
- “hierarchical agents” OR “agent orchestration”
- “autonomous agents” AND “AI systems design”
- “distributed artificial intelligence” AND “applications”

Estas expresiones se adaptaron a la sintaxis específica de cada base de datos. De igual manera, como criterios de inclusión se establecieron los siguientes:

- Publicaciones entre 2020 y 2026
- Artículos en revistas indexadas (Q1-Q3) y congresos reconocidos en informática (por ejemplo, IEEE, ACM)

- Estudios que aborden fundamentos teóricos, arquitecturas, aplicaciones o implicaciones éticas de sistemas basados en subagentes
- Documentos en idioma inglés o español

Por su lado, los criterios de exclusión se definieron de la siguiente manera:

- Estudios centrados exclusivamente en herramientas específicas sin aporte arquitectónico
- Publicaciones no revisadas por pares (salvo informes institucionales relevantes)
- Trabajos enfocados únicamente en ChatGPT, lenguajes de programación o bases de datos sin relación directa con arquitecturas subagénticas

Como fuentes complementarias y con el fin de contextualizar tendencias industriales y aplicaciones reales, se incorporaron informes recientes de organizaciones y empresas tecnológicas, tales como McKinsey & Company, Deloitte, Gartner, World Economic Forum, IBM y Microsoft. Asimismo, se analizaron estudios de caso asociados a plataformas como Amazon Bedrock (AgentCore) y Claude, así como el documento estratégico ILIA 2025 y sus recomendaciones sobre IA para la región latinoamericana.

Los estudios seleccionados fueron analizados mediante un enfoque de síntesis cualitativa temática, clasificando los hallazgos en tres dimensiones principales: fundamentos y modelos arquitectónicos; aplicaciones y casos de uso; e implicaciones éticas, organizacionales y técnicas.

RESULTADOS

Marco conceptual *multi-agent data mining* (MADM)

Para comprender el funcionamiento de los sistemas multiagente de IA, se retoma el marco conceptual MADM propuesto por Chaimontree et al. (2010). Este integra técnicas de minería de datos (*data mining*, DM) con sistemas multiagente (*multi-agent systems*, MAS). Este enfoque tiene como objetivo automatizar, distribuir y potenciar de manera inteligente el proceso de descubrimiento de conocimiento en grandes volúmenes de datos.

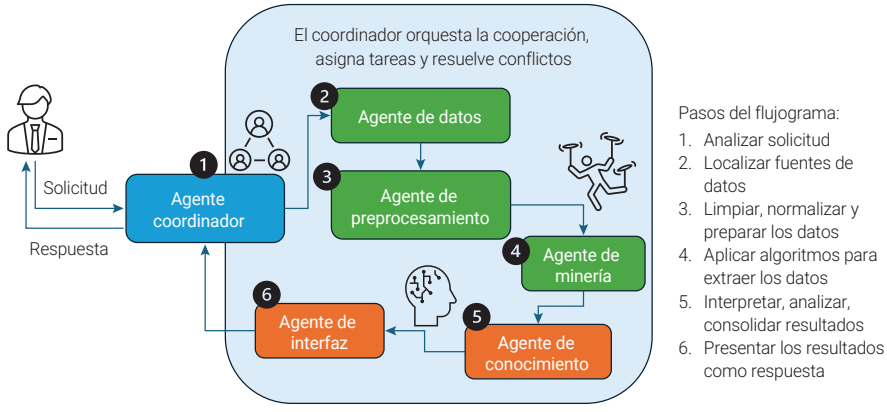
En este entorno, un conjunto de agentes coopera para abordar tareas de minería de datos específicas, con el fin de mejorar el proceso de acceso a los datos mediante agentes inteligentes cooperativos, los cuales pueden ejecutar de manera autónoma, distribuida y coordinada las distintas etapas del proceso, el cual se encuentra subdividido en tres partes:

- Preprocesamiento (limpieza, integración y selección de datos)
- Minería propiamente dicha (aplicación de algoritmos de aprendizaje o *clustering*)
- Posprocesamiento (evaluación e interpretación de los resultados)

Para facilitar la comprensión de su funcionamiento, en la Figura 1 se muestra un flujograma que ilustra la interacción entre el usuario y el sistema: el usuario formula una solicitud al agente, el orquestador distribuye y coordina las tareas entre los agentes especializados, y, finalmente, se integra y devuelve la respuesta correspondiente.

Figura 1

Flujograma de interacción de multiagentes de IA basado en modelo MADM



Nota. Los datos proceden de "Multi-Agent Based Clustering: Towards Generic Multi-Agent Data Mining", por S. Chaimontree, K. Atkinson y F. Coenen, 2010, en P. Perner (Ed.), *Advances in Data Mining: Applications and Theoretical Aspects*, Springer (https://doi.org/10.1007/978-3-642-14400-4_9).

El análisis de este flujograma de seis pasos es el siguiente:

1. El agente coordinador analiza la solicitud, orquesta la cooperación entre los demás agentes, asigna tareas y resuelve conflictos dentro de su área de acción.
2. Los agentes de datos gestionan las fuentes de datos, acceden a información relevante usando distintas bases y repositorios.
3. Los agentes de preprocesamiento limpian, transforman, normalizan y preparan los datos para ser extraídos apropiadamente.
4. Los agentes de minería se encargan en extraer los datos aplicando algoritmos apropiados según el tipo de análisis.

5. Los agentes de conocimiento interpretan y analizan patrones, y consolidan los resultados como información final.
6. Los agentes de interfaz, como especializados en la interacción con el usuario, son los que dan formato a la información y presentan los resultados al coordinador, quien, finalmente, envía la respuesta al solicitante.

La secuencia inicia desde el momento en que el usuario formula la solicitud. Luego, el proceso se desarrolla de manera cooperativa, adaptativa y paralela, bajo la supervisión de un agente coordinador que actúa como orquestador de todo el ciclo. Conforme al modelo MADM, y tomando como referencia el proceso de minería de datos, diversos tipos de agentes especializados intervienen con funciones específicas, cuya interacción es gestionada por un coordinador general (Tabla 1).

Tabla 1

Funciones de multiagentes según modelo MADM

Tipo de agente	Función principal
Agente de coordinación (*coordination agent*)	Orquesta la cooperación entre los demás agentes, asigna tareas y resuelve conflictos.
Agente de datos (*user agent*)	Gestiona fuentes de datos distribuidas, accede y selecciona la información relevante. Proporciona una interfaz gráfica de usuario entre los agentes de usuario y los agentes de tarea. Conduce hacia la minería de datos genérica multiagéntica.
Agente de preprocesamiento (*task agent*)	Es responsable de gestionar y programar las solicitudes de minería de datos. Facilitan la ejecución de las tareas de minería de datos: limpian, transforman y normalizan los datos.
Agente de datos (*data agent*)	Posee metadatos sobre un conjunto de datos específico, lo que les permite acceder a ellos. Existe una relación directa entre los agentes de datos y los conjuntos de datos.
Agente minero o analítico (*mining agent*)	Es responsable de realizar la minería de datos y generar resultados. El resultado se devuelve a un agente de tareas. Aplica algoritmos de minería de datos (por ejemplo, árboles de decisión, redes neuronales, *clustering*, reglas de asociación).
Agente de validación (*validation agent*)	Es responsable de evaluar y valorar la validez de los resultados de minería de datos. Identifica varios subtipos de agentes de validación, como subtipos identificados y asociados con diferentes tareas genéricas de minería de datos.
Agente de conocimiento (*knowledge agent*)	Interpreta y almacena los patrones descubiertos, y gestiona la base de conocimiento.
Agente de interfaz o usuario (*user interface agent*)	Facilita la interacción con el analista o usuario final.

Nota. Los datos proceden de "Multi-Agent Based Clustering: Towards Generic Multi-Agent Data Mining", por S. Chaimontree, K. Atkinson y F. Coenen, 2010, en P. Perner (Ed.), *Advances in Data Mining: Applications and Theoretical Aspects*, Springer (https://doi.org/10.1007/978-3-642-14400-4_9).

Desde esta perspectiva, las arquitecturas multiagente de IA contribuyen al desarrollo productivo cuando permiten la automatización y distribución inteligente de tareas, tal como lo demuestra el marco conceptual MADM, orientado a optimizar procesos analíticos en entornos complejos y de gran escala. En este contexto, los multiagentes cooperan para ejecutar tareas específicas de minería de datos, lo que facilita el acceso, el análisis y el procesamiento de la información mediante una ejecución autónoma, distribuida y coordinada de las distintas etapas del proceso (Tabla 1).

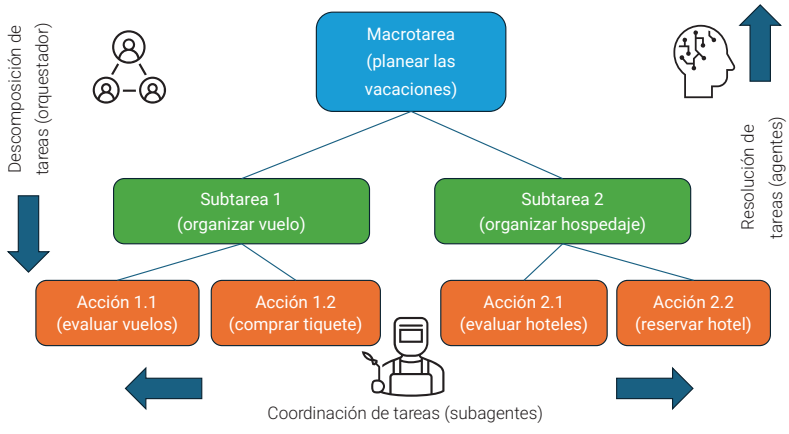
Árbol de tareas complejas

Los motores de búsqueda han existido durante décadas y desempeñan un papel fundamental al ofrecer acceso casi inmediato a respuestas y recursos para una amplia gama de consultas. Sin embargo, el fortalecimiento de los procesos de búsqueda mediante agentes de inteligencia artificial exige situar a las tareas como el eje central del *information retrieval*, en la medida en que estas articulan los objetivos, las acciones y los contextos que guían la interacción entre los usuarios y los sistemas de búsqueda.

De acuerdo con White (2024), los sistemas tradicionales continúan basándose, principalmente, en la observación de acciones explícitas del usuario, lo que limita la inferencia de los objetivos subyacentes, especialmente en escenarios caracterizados por una elevada complejidad. Estas tareas se caracterizan por ser mal definidas, multifásicas y distribuidas a lo largo de múltiples consultas y sesiones, además de exigir elevados niveles de esfuerzo cognitivo, estrechamente vinculados con la experiencia del usuario cuando interactúa con la IA. No obstante, su integración con arquitecturas de agentes de IA organizadas jerárquicamente permite trascender los modelos tradicionales centrados en consultas aisladas.

Figura 2

Árbol de tareas complejas utilizando agentes y subagentes de IA



Nota. Los datos proceden de "Advancing the Search Frontier with AI Agents", por R. W. White, 2024, *Communications of the ACM*, *67*(9), p. 3 (https://doi.org/10.1145/3655615).

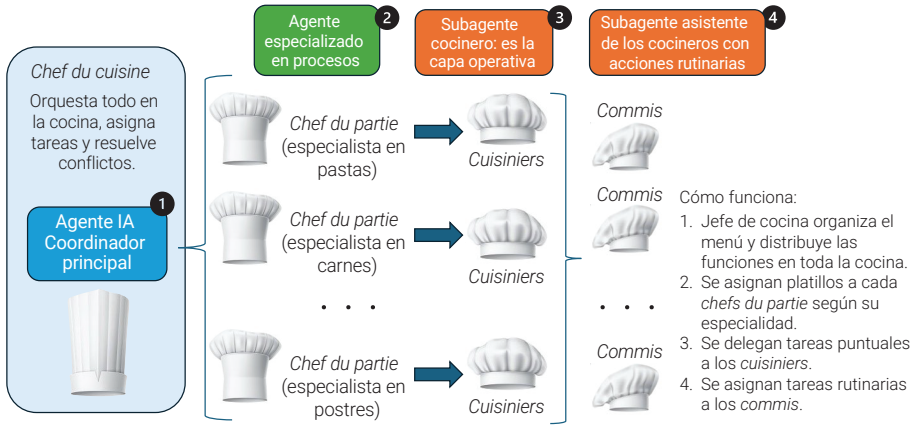
Un agente orquestador, apoyado por subagentes especializados, permite una comprensión más precisa de las intenciones del usuario y la coordinación de acciones avanzadas para la resolución progresiva de tareas complejas. En la Figura 2, se muestra un proceso de búsqueda de información orientado a la resolución de tareas complejas, modelado mediante una estructura jerárquica en forma de árbol.

Esta propuesta permite la descomposición de tareas complejas en macrotareas, subtareas y acciones, que representan objetivos de alto nivel (planear las vacaciones), componentes intermedios (evaluar el vuelo y el hospedaje) y, finalmente, operaciones observables (comprar un tiquete y reservar un hotel). Para abordar la complejidad inherente a tareas de búsqueda, White (2024) sugiere el modelo de árbol de tareas, sustentado en agentes y subagentes coordinados jerárquicamente, en el que los agentes especializados asumen la ejecución de subtareas específicas bajo la supervisión de un agente orquestador (subtareas 1 y 2). Este enfoque facilita la descomposición estructurada de dichas subtareas (acciones 1.1, 1.2, 2.1 y 2.2).

La coordinación eficiente y la resolución progresiva de tareas complejas consolidan a las arquitecturas subagénticas como un mecanismo fundamental para mejorar la efectividad de los sistemas de búsqueda y para apoyar la toma de decisiones (White, 2024).

Figura 3

Metáfora de la cocina utilizando agentes y subagentes de IA



Nota. Los datos proceden de "Creando asistentes, agentes y plataformas colaborativas de inteligencia artificial", por O. Bustillos Ortega, J. Murillo Gamboa, O. Nuñez Peralta y F. Rodríguez Sibaja, 2025, *Interfases*, (21), p. 46 (https://doi.org/10.26439/interfases2025.n021.7801).

Metáfora de la cocina

Otra forma de comprender el funcionamiento de los agentes y asistentes de IA es a través de la organización jerárquica de una brigada de cocina tradicional, considerando al

personal compuesto por un chef principal, chefs especializados, cocineros y asistentes de cocina. Esta metáfora permite ilustrar cómo los distintos roles humanos pueden asociarse conceptualmente con agentes y subagentes de IA, cada uno encargado de ejecutar tareas diferenciadas y coordinadas dentro de un entorno operativo común (Figura 3).

El ciclo operacional de la cocina, según la Figura 3, se articula de la siguiente manera:

1. *Chef de cuisine*: define el menú del día y establece las directrices generales relativas para la preparación y presentación de los platillos. Luego, descompone estas directrices en objetivos y tareas específicas, que son asignadas a los *chefs* conforme a su especialización funcional.
2. *Chefs de partie* especializados: coordinan con sus respectivos subagentes (*cuisiniers* y *commis*), quienes ejecutan las operaciones concretas requeridas para la elaboración de cada platillo, a fin de asegurar la coherencia, la eficiencia y la alineación con los objetivos globales del sistema (Bustillos Ortega et al. 2025).
3. *Cuisiniers*: se enfocan en elaborar cada componente del platillo.
4. *Commis*: asisten y ayudan a los *cuisiniers* y a los *chefs de partie*.

En esta metáfora, el *chef de cuisine* actúa como el responsable de la coordinación global de todas las actividades, por lo que cumple un rol equivalente al de un orquestador en un sistema de IA. Los *chefs de partie* son los responsables de la preparación de platillos específicos —como pastas, carnes o postres—, por lo que pueden considerarse agentes de IA especializados, cada uno focalizado en un dominio funcional o en un conjunto de tareas bien delimitadas. Los cocineros (*cuisiniers*) y los ayudantes de cocina (*commis*) conforman la capa operativa de soporte, que es análoga a la de los subagentes, encargados de ejecutar acciones más acotadas, rutinarias y de menor nivel de abstracción, siguiendo las directrices establecidas por los agentes. principales.

Soluciones con agentes y subagentes de IA

Los agentes de IA pueden, entre otras cosas, ayudar a los usuarios a alcanzar objetivos, maximizar la utilidad y realizar funciones automatizadas. Por ejemplo, los asistentes digitales personales Siri de Apple y Alexa de Amazon, que pueden responder preguntas y ayudar con la gestión de tareas personales.

Con respecto a las plataformas de desarrollo de soluciones de IA, existe GitHub Copilot (Github, s. f.), el cual reduce el esfuerzo requerido al desarrollador de aplicaciones, facilita la ejecución exitosa de las tareas y acelera significativamente la finalización de soluciones automatizadas. También está el Auto-GPT (s. f.), un agente totalmente autónomo que puede descomponer las tareas en subtareas y ejecutarlas de forma independiente en nombre del usuario para facilitar la consecución de objetivos.

En ambos casos, la capa de orquestación de IA gestiona los flujos de información internos, activa, establece y ejecuta herramientas o complementos, y procesa sus respuestas, entre otras cosas. Todo esto se ejecuta sobre una infraestructura de IA a gran escala alojada en la nube en plataformas, tales como Microsoft Azure, Google Cloud y Amazon Web Services. Se sustenta en un fuerte compromiso con una IA responsable que garantiza la seguridad, protección y transparencia de los agentes (White, 2024).

Índice Latinoamericano de IA (ILIA2025)

De acuerdo con Durán et al. (2025), la iniciativa del Índice Latinoamericano de Inteligencia Artificial (ILIA) es una iniciativa elaborada en el 2023 por el Centro Nacional de Inteligencia Artificial (CENIA) de Chile y el Observatorio de Desarrollo Digital de la Comisión Económica para América Latina y el Caribe (CEPAL), con el apoyo de diversas organizaciones académicas, públicas y privadas. El ILIA identifica una transición en la región, basado en el análisis de ochenta y cinco indicadores en doce áreas clave para la transformación digital y con la visión de incorporar de manera estratégica estas tecnologías emergentes para acelerar un desarrollo productivo, inclusivo y sostenible, y asegurar, al mismo tiempo, su uso ético, responsable y orientado al bien común.

Los resultados del ILIA 2025 evidencian que América Latina avanza progresivamente desde una fase de adopción inicial hacia la consolidación de capacidades propias en investigación, formación avanzada y desarrollo de talento especializado en inteligencia artificial, aunque este progreso se caracteriza por marcadas asimetrías entre países. El conocimiento y la producción científica continúan concentrándose en un grupo reducido de naciones —principalmente Brasil, México, Chile, Colombia y Argentina—, mientras que gran parte de la región mantiene ecosistemas científicos incipientes y una limitada proyección internacional. No obstante, experiencias como las de Costa Rica y Perú demuestran que la articulación entre talento humano, institucionalidad sólida y adopción responsable de la IA puede sentar las bases para ecosistemas escalables, auditables y orientados al bien común, en los que la inteligencia distribuida actúe como motor del desarrollo sostenible. En el ámbito de la formación académica, se observa un crecimiento sostenido de los programas de maestría y doctorado en IA, junto con la expansión de iniciativas de ciencia abierta y colaboración regional que favorecen el desarrollo de soluciones locales y transparentes. Sin embargo, persiste un "embudo formativo": mientras se amplía la alfabetización en IA en los niveles básico y técnico, continúa siendo insuficiente la generación de desarrolladores e investigadores altamente especializados (Durán et al. 2025; Soto Arriaza & Durán Rojas, 2023).

En tal sentido, se considera que la formación futura deberá integrar ética, explicabilidad, sostenibilidad y multidisciplinariedad, incorporando la IA generativa como herramienta pedagógica y de democratización del conocimiento. Asimismo, el informe recomienda fortalecer centros regionales de excelencia, programas interdisciplinarios de

posgrado y consorcios de infraestructura compartida en cómputo y datos abiertos. Solo así la región latinoamericana podrá pasar de ser consumidora a productora de innovación responsable, lo que garantizará soberanía digital y desarrollo sostenible.

El ILIA 2025 destaca el avance de América Latina hacia ecosistemas de IA más integrados, en los que los sistemas compuestos por agentes cooperativos y autónomos emergen como un instrumento clave para impulsar el desarrollo productivo sostenible. Estas arquitecturas distribuyen funciones de percepción, razonamiento y acción, lo que mejora el desempeño y la escalabilidad de los procesos industriales, educativos y públicos mediante la optimización de recursos y la adaptabilidad operativa. Sin embargo, su implementación enfrenta desafíos en orquestación, seguridad, costo computacional y gobernanza ética.

Se proyecta que el desarrollo futuro de la IA en América Latina estará condicionado por la capacidad de sus sistemas educativos y científicos para formar talento avanzado, promover investigación abierta y aplicada, y articular redes regionales de conocimiento. En este sentido, el tránsito desde una alfabetización básica hacia una especialización profunda en IA se configura como un punto decisivo, ya que solo a través de una educación con enfoque ético, interdisciplinario y sostenible la región podrá superar su rol de consumidora de tecnología y consolidarse como productora de innovación responsable (Durán et al. 2025; Soto Arriaza & Durán Rojas, 2023). En la Tabla 2, se presentan recomendaciones clave agrupadas según el eje estratégico del ILIA2025.

Tabla 2

Recomendaciones estratégicas para la academia según ILIA 2025

Eje estratégico	Recomendación clave	Objetivo
Política educativa	Incluir IA en todos los niveles educativos, con enfoque en equidad y ética	Democratizar el acceso y generar interés temprano
Investigación aplicada	Vincular universidades con sectores productivos y públicos	Alinear la IA con desafíos nacionales (salud, sostenibilidad, educación)
Formación avanzada	Fortalecer programas de maestría y doctorado interdisciplinarios	Elevar la especialización científica y tecnológica
Infraestructura académica	Crear consorcios regionales de cómputo y datos abiertos	Garantizar soberanía digital e investigación de frontera
Colaboración internacional	Participar en redes globales y normas ISO/IEEE de IA	Aumentar visibilidad y transferencia de conocimiento

Nota. Los datos proceden de *Índice Latinoamericano de Inteligencia Artificial (ILIA) 2025*, por R. Durán, A. Moreno, S. Adasme, S. Rovira, V. Jordán y L. Poveda (Coords.), 2025, Comisión Económica para América Latina y el Caribe (https://www.cepal.org/es/publicaciones/82514-indice-latinoamericano-inteligencia-artificial-ilia-2025).

El ILIA 2025 confirma que América Latina transita de una etapa de adopción incipiente hacia la construcción de capacidades propias en investigación y formación avanzada en inteligencia artificial. Sin embargo, este proceso avanza de manera desigual y todavía enfrenta limitaciones estructurales en inversión, especialización y articulación científica (Durán et al. 2025; Soto Arriaza & Durán Rojas, 2023) (Tabla 3).

Tabla 3

Índice Latinoamericano ILIA 2025 estado actual y tendencias en investigación

ILIA (indicadores)	Estado actual y tendencias
Concentración geográfica del conocimiento	La investigación en IA sigue altamente concentrada. Brasil y México reúnen el 68 % de los investigadores activos y —junto con Colombia, Chile y Argentina— concentran el 90 % de las publicaciones científicas. Solo siete países participan en conferencias internacionales de alto impacto, con predominio de Chile y Brasil, lo que deja a la mayoría con baja visibilidad.
Crecimiento del ecosistema formativo en Latinoamérica	Se observa un aumento de programas de posgrado en IA (magíster y doctorado), especialmente en Chile, Brasil, México, Colombia y emergentes como Costa Rica y Perú, que han incorporado IA en currículos universitarios y de educación técnica. Pese a este avance, once de los diecinueve países aún no ofrecen programas de doctorado específicos en IA.
Apertura y colaboración	El informe destaca el modelo de código abierto y la ciencia colaborativa como motores del desarrollo científico regional. Honduras, El Salvador y Cuba son ejemplos del potencial del *open source* para crear soluciones locales y fomentar comunidades académicas activas. Esta tendencia podría ser clave para democratizar la investigación, superar barreras económicas y promover la transparencia algorítmica.
Desafíos estructurales	Escasa inversión en I+D. La región representa apenas el 1,12 % de la inversión global en IA, pese a concentrar el 8,8 % de la población mundial. Falta de redes interuniversitarias sólidas y limitado intercambio de investigadores impiden generar una masa crítica regional. Poca participación en estándares y foros internacionales, lo que reduce la influencia latinoamericana en la gobernanza global de la IA.
Hacia una alfabetización avanzada	La alfabetización en IA en la región Latinoamericana se ha expandido —ya presente en los currículos escolares de seis países—, pero la formación especializada sigue rezagada. La proporción de profesionales con habilidades técnicas en IA es cuatro veces menor que la de alfabetización general. El informe advierte un "embudo formativo": se forma una base amplia de usuarios, pero pocos desarrolladores o investigadores.
Competencias clave para el futuro	El ILIA 2025 propone impulsar nuevos perfiles académicos centrados en IA responsable y ética, con formación en sesgos, transparencia y explicabilidad, así como en ciencia de datos aplicada a políticas públicas y desarrollo sostenible. Asimismo, se recomienda promover la interdisciplinariedad mediante la integración de ingeniería, ciencias sociales y humanidades en programas de IA, así como desarrollar capacidades para la IA generativa y multimodal, considerada un campo democratizador y de alto impacto.

(*continúa*)

(*continuación*)

ILIA (indicadores)	Estado actual y tendencias
Formación conectada a la soberanía digital	La escasez de infraestructura (HPC y GPU) limita la investigación avanzada; por ello, el ILIA recomienda alianzas regionales para uso compartido de cómputo y datos. La educación debe orientarse a fortalecer capacidades endógenas, a fin de evitar la dependencia tecnológica y sesgos en los modelos importados.
Modelos educativos colaborativos	Se promueve la creación de centros regionales de excelencia en IA y plataformas abiertas de aprendizaje que repliquen experiencias exitosas como "Hazlo con IA" en Chile o el LatamGPT colaborativo. Estas iniciativas demuestran que la formación masiva puede acelerar la adopción responsable y reducir la brecha de talento.

Nota. Los datos proceden de *Índice Latinoamericano de Inteligencia Artificial (ILIA) 2025*, por R. Durán, A. Moreno, S. Adasme, S. Rovira, V. Jordán y L. Poveda (Coords.), 2025, Comisión Económica para América Latina y el Caribe (https://www.cepal.org/es/publicaciones/82514-indice-latinoamericano-inteligencia-artificial-ilia-2025).

Riesgos y ética

Los riesgos relacionados con las tecnologías de IA ponen de manifiesto la necesidad de diseñar arquitecturas que incorporen mecanismos de visibilidad, trazabilidad y control capaces de habilitar una supervisión granular del comportamiento de los agentes. La literatura coincide en que el aumento de la autonomía de los agentes inteligentes incrementa su exposición a amenazas técnicas y éticas, lo que exige el diseño de arquitecturas jerárquicas basadas en subagentes especializados que refuercen la supervisión, la resiliencia y la gobernanza interna.

Para mitigar riesgos asociados a la privacidad, el control, la explicabilidad y la asignación de responsabilidades, resulta fundamental incorporar desde etapas tempranas principios éticos, mecanismos robustos de seguridad, así como procesos de auditoría y trazabilidad que garanticen un despliegue confiable, seguro y responsable de estas arquitecturas en entornos productivos (Deng et al. 2025; Ning et al. 2025; Samdani et al. 2025). De acuerdo con Chan et al. (2024), los agentes de IA presentan riesgos significativamente mayores que los sistemas tradicionales, debido a su capacidad de actuar autónomamente y de ejecutar acciones en el mundo digital sin supervisión constante. En tal sentido, destacan como principales riesgos los siguientes:

- La manipulación de los objetivos del agente, que puede llevar a perseguir fines no alineados con los establecidos por los desarrolladores
- La ejecución de acciones dañinas o inseguras, derivadas de interpretaciones erróneas, instrucciones ambiguas o fallos en sus cadenas de planificación
- La susceptibilidad a ataques externos, en los que actores maliciosos pueden redirigir, engañar o explotar al agente

- La opacidad operativa que dificulta detectar desviaciones de comportamiento, fallos de seguridad o actividades no autorizadas

En conjunto, las arquitecturas subagénticas presentan riesgos significativos asociados a la seguridad, la gobernanza, la propagación de sesgos y los costos computacionales derivados de la coordinación entre múltiples subagentes. Frente a estas vulnerabilidades, resulta esencial promover diseños que integren explícitamente el bienestar humano y fomenten una colaboración humano-IA segura, transparente y responsable. A futuro, los sistemas de IA colaborativos basados en subagentes deberán orientarse a tareas especializadas y centradas en el usuario y avanzar hacia modelos de control compartido humano-IA (Samdani et al. 2025).

Desde una perspectiva de riesgos, los resultados del ILIA 2025 evidencian que la limitada inversión en investigación y desarrollo constituye una vulnerabilidad estructural crítica para el avance de la inteligencia artificial en América Latina. A pesar de albergar el 8,8 % de la población mundial, la región concentra apenas el 1,12 % de la inversión global en IA, lo que incrementa el riesgo de dependencia tecnológica, rezago científico y pérdida de competitividad. Este escenario se ve agravado por la debilidad de las redes interuniversitarias y el escaso intercambio de investigadores, factores que obstaculizan la conformación de una masa crítica regional y limitan la capacidad de generar conocimiento, innovación y liderazgo científico sostenible. Frente a este riesgo, la ausencia de centros regionales de excelencia en inteligencia artificial compromete la formación continua de talento especializado y la adopción coordinada, ética y responsable de arquitecturas subagénticas de IA, lo que incrementa la probabilidad de una implementación fragmentada y poco sostenible de estas tecnologías (Durán et al. 2025; Soto Arriaza & Durán Rojas, 2023).

DISCUSIÓN

Los resultados del análisis evidencian que las arquitecturas subagénticas de inteligencia artificial representan un avance significativo frente a los enfoques monolíticos tradicionales, pues permiten la especialización funcional, la descomposición de tareas complejas y la coordinación jerárquica mediante agentes orquestadores. Esta organización distribuida se traduce en mejoras sustanciales en términos de eficiencia operativa, escalabilidad, paralelismo y adaptabilidad, particularmente en contextos caracterizados por grandes volúmenes de información y procesos multifásicos.

El estudio del marco conceptual MADM demuestra que la integración de técnicas de minería de datos con sistemas multiagente facilita la automatización y optimización del descubrimiento de conocimiento, pues asigna funciones específicas a agentes especializados en preprocesamiento, análisis, validación e interpretación. Este enfoque evidencia que la inteligencia distribuida no solo acelera los flujos analíticos, sino que también

mejora la trazabilidad y la calidad de los resultados, lo que la consolida como un modelo eficaz para entornos productivos y científicos a gran escala.

Asimismo, el modelo del árbol de tareas complejas confirma que la estructuración jerárquica de tareas mediante agentes y subagentes permite una comprensión más precisa de las intenciones del usuario y una resolución progresiva de problemas complejos. La capacidad de descomponer macrotareas en subtareas y acciones observables fortalece los procesos de búsqueda de información y soporte a la toma de decisiones, lo que supera las limitaciones de los sistemas tradicionales basados en consultas aisladas.

La metáfora de la brigada de cocina refuerza conceptualmente estos hallazgos al ilustrar cómo la coordinación entre roles especializados, bajo la supervisión de un orquestador, resulta clave para garantizar la coherencia, la eficiencia y la calidad de los resultados. Esta analogía facilita la comprensión de la dinámica operativa de las arquitecturas subagénticas y subraya la importancia de una orquestación clara para evitar redundancias, conflictos y sobrecarga computacional.

En cuanto a las aplicaciones prácticas, los casos analizados en movilidad y transporte, finanzas, salud y educación evidencian que las arquitecturas subagénticas potencian la productividad al distribuir responsabilidades, aumentar la explicabilidad de los procesos y favorecer la colaboración entre humanos e IA. No obstante, los resultados también ponen de manifiesto desafíos persistentes asociados a la orquestación técnica, la seguridad, los costos computacionales y la gobernanza ética, especialmente en sistemas con altos niveles de autonomía.

La discusión resalta que la autonomía creciente de los subagentes incrementa los riesgos de opacidad, sesgos, vulnerabilidades de seguridad y desviaciones de comportamiento, lo que exige la incorporación de mecanismos de monitoreo, auditoría, trazabilidad y control jerárquico como componentes estructurales del diseño. En este sentido, la delegación de funciones de supervisión y seguridad a subagentes especializados emerge como una estrategia efectiva para reforzar la resiliencia y la confiabilidad de los sistemas.

Finalmente, el análisis del ILIA 2025 contextualiza estos resultados en el ámbito regional, lo que evidencia que América Latina cuenta con avances relevantes en alfabetización y adopción de IA, pero enfrenta limitaciones estructurales en inversión, infraestructura y formación avanzada. Los hallazgos sugieren que las arquitecturas subagénticas pueden convertirse en un catalizador del desarrollo productivo regional, siempre que se articulen con políticas educativas, marcos éticos sólidos y estrategias de cooperación interinstitucional orientadas a la soberanía digital y al bien común.

CONCLUSIONES

Con respecto a la pregunta de investigación —¿cómo contribuyen las arquitecturas subagénticas de IA a impulsar el desarrollo productivo, considerando los desafíos en

la orquestación, la seguridad y la gobernanza ética?—, las arquitecturas subagénticas de inteligencia artificial contribuyen de manera decisiva al desarrollo productivo al posibilitar la descomposición de tareas complejas en unidades funcionales especializadas, coordinadas por un agente orquestador dentro de sistemas distribuidos. Este enfoque incrementa la eficiencia operativa, la escalabilidad y la adaptabilidad de los procesos en sectores clave, tales como la industria, la educación, la salud, las finanzas y la gestión del conocimiento, al permitir la ejecución paralela, la reutilización de capacidades especializadas y una mejor gestión del contexto en entornos intensivos en datos. Marcos como el MADM, el árbol de tareas complejas y las arquitecturas de WebAgents demuestran que la inteligencia distribuida facilita la automatización avanzada y la toma de decisiones informadas en escenarios de alta complejidad.

No obstante, el impacto productivo de estas arquitecturas depende críticamente de la orquestación técnica, ya que la coordinación jerárquica entre agentes y subagentes requiere mecanismos robustos para la asignación de tareas, la resolución de conflictos y la integración coherente de resultados. Una orquestación deficiente puede generar ineficiencias, sobrecostos computacionales y pérdida de control operativo, lo que subraya la necesidad de diseños modulares, monitoreo en tiempo real y estrategias de control distribuido.

En paralelo, los riesgos para la seguridad emergen como un desafío estructural, dado que la autonomía de los subagentes amplía la superficie de ataque y los riesgos asociados a acciones no alineadas, fallos de interpretación o interferencias maliciosas. El documento evidencia que la incorporación de subagentes dedicados a funciones de identificación, monitoreo, auditoría y registro fortalece la resiliencia del sistema y permite aplicar principios de defensa en profundidad, lo que reduce vulnerabilidades sin sacrificar la flexibilidad operativa.

La gobernanza ética se posiciona como un eje transversal para garantizar que el desarrollo productivo impulsado por arquitecturas subagénticas sea sostenible y socialmente legítimo. La necesidad de transparencia, trazabilidad, explicabilidad y rendición de cuentas exige integrar marcos normativos, auditorías continuas y supervisión entre humanos e IA desde las fases iniciales de diseño. En este sentido, el ILIA 2025 destaca que la articulación entre talento humano, institucionalidad sólida y adopción responsable de la IA puede convertir a estas arquitecturas en un motor de innovación inclusiva, especialmente en contextos latinoamericanos caracterizados por brechas de infraestructura y especialización.

El documento concluye que las arquitecturas subagénticas de IA impulsan el desarrollo productivo al combinar especialización funcional, coordinación jerárquica y automatización inteligente, siempre que se acompañen de estrategias integrales de orquestación, seguridad y gobernanza ética. Su verdadero valor no reside únicamente en la eficiencia técnica, sino en su capacidad para habilitar ecosistemas productivos más

transparentes, resilientes y orientados al bien común, sustentados en una formación avanzada y políticas públicas responsables.

En conclusión, el desafío central para la región consiste en transformar la alfabetización tecnológica en innovación científica sustentada en disciplinas STEM, promoviendo una adopción responsable de la inteligencia artificial, el desarrollo sostenible y una participación activa en la producción global de soluciones basadas en IA. Este objetivo requiere una formación continua y especializada, así como la implementación estratégica de arquitecturas subagénticas en los distintos sectores productivos y en los sistemas educativos, de modo que la inteligencia distribuida se consolide como un motor de competitividad, equidad y bienestar social.

REFERENCIAS

Auto-GPT. (s. f.). *Auto-GPT – The next evolution of data driven Chat AI*. https://auto-gpt.ai/

Babaei, G., & Giudici, P. (2025). Explainable artificial intelligence (XAI) in investment decision-making. *Academia AI and Applications*, *1*(2). https://doi.org/10.20935/AcadAI8017

Bustillos Ortega, O., Murillo Gamboa, J., Núñez Peralta, O., & Rodríguez Sibaja, F. (2025). Creando asistentes, agentes y plataformas colaborativas de inteligencia artificial. *Interfases*, (21), 35-57. https://doi.org/10.26439/interfases2025.n021.7801

Carrascosa, C., Terrasa, A., Fabregat, J., & Botti, V. (2005). Gestión de comportamientos en agentes de tiempo real. *Inteligencia Artificial: Revista Iberoamericana de Inteligencia Artificial*, *9*(25), 39-48. https://dialnet.unirioja.es/servlet/articulo?codigo=1212985

Chaimontree, S., Atkinson, K., & Coenen, F. (2010). Multi-agent based clustering: Towards generic multi-agent data mining. En P. Perner (Ed.), *Advances in data mining: Applications and theoretical aspects* (pp. 115-127). Springer. https://doi.org/10.1007/978-3-642-14400-4_9

Chan, A., Ezell, C., Kaufmann, M., Wei, K., Hammond, L., Bradley, H., Bluemke, E., Rajkumar, N., Krueger, D., Kolt, N., Heim, L., & Anderljung, M. (2024). Visibility into AI agents. En *FAccT '24: The 2024 ACM conference on fairness, accountability, and transparency* (pp. 958-973). Association for Computing Machinery. https://doi.org/10.1145/3630106.3658948

Chevrot, A., Vernotte, A., Falleri, J.-R., Blanc, X., Legeard, B., & Cretin, A. (2025). Are autonomous web agents good testers? *Proceedings of the ACM on Software Engineering*, *2*(ISSTA), 206-228. https://doi.org/10.1145/3728879

Claude Code Docs. (2025). *Crear subagentes personalizados*. https://code.claude.com/docs/es/sub-agents

Deng, Z., Guo, Y., Han, C., Ma, W., Xiong, J., Wen, S., & Xiang, Y. (2025). AI agents under threat: A survey of key security challenges and future pathways. *ACM Computing Surveys*, *57*(7), Artículo 182. https://doi.org/10.1145/3716628

Durán, R., Moreno, A., Adasme, S., Rovira, S., Jordán, V., & Poveda, L. (Coords.). (2025). *Índice Latinoamericano de inteligencia artificial (ILIA) 2025*. Comisión Económica para América Latina y el Caribe. https://www.cepal.org/es/publicaciones/82514-indice-latinoamericano-inteligencia-artificial-ilia-2025

Elmahalawy, A. M. (2015). A new approach in learning for intelligent multi agent systems. En *Proceedings 21st European Conference on Modelling and Simulation*. European Council for Modelling and Simulation. https://www.academia.edu/70025062/Intelligent_Agent_Technology_Intelligent_Multi_Agent

Github. (s. f.). *Github Copilot. Domina tu oficio*. https://github.com/features/copilot?locale=es-419

Gonçalves, E. J. T., Cortés, M. I., Campos, G. A. L., Lopes, Y. S., Freire, E. S. S., Da Silva, V. T., De Oliveira, K. S. F., & De Oliveira, M. A. (2015). MAS-ML 2.0: Supporting the modelling of multi-agent systems with different agent architectures. *Journal of Systems and Software*, *108*, 77-109. https://doi.org/10.1016/j.jss.2015.06.008

Ning, L., Liang, Z., Jiang, Z., Qu, H., Ding, Y., Fan, W., Wei, X. Y., Lin, S., Liu, H., Yu, P. S., & Li, Q. (2025). A survey of web agents: Towards next-generation AI agents for web automation with large foundation models. En *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 6140-6150). Association for Computing Machinery. https://doi.org/10.1145/3711896.3736555

Rother, D., Pajarinen, J., Peters, J., & Weisswange, T. H. (2025). Open-ended coordination for multi-agent systems using modular open policies. *Autonomous Agents and Multi-Agent Systems*, *39*, Artículo 40. https://doi.org/10.1007/s10458-025-09723-7

Samdani, G., Viswanathan, G., & Dasu Jegadeesh, A. (2025). Human-AI collaboration: Balancing agentic ai and autonomy in hybrid systems. *International Journal on Cloud Computing: Services and Architecture*, *15*(1). https://doi.org/10.5121/ijccsa.2025.15101

Soto Arriaza, A., & Durán Rojas, R. (Dirs.). (2023). *Índice latinoamericano de inteligencia artificial*. Centro Nacional de Inteligencia Artificial. https://indicelatam.cl/wp-content/uploads/2023/08/ILIA-2023.pdf

White, R. W. (2024). Advancing the search frontier with AI agents. *Communications of the ACM*, *67*(9), 54-65. https://doi.org/10.1145/3655615

Zhu, Z., Tan, Y., Yamashita, N., Lee, Y. C., & Zhang, R. (2025, 26 de abril). The benefits of prosociality towards AI agents: Examining the effects of helping AI agents on human well-being. En *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-18). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713116