

ARQUITECTURAS AGÉNTICAS DE IA: UN ESTUDIO COMPARATIVO DE *WORKFLOWS* (FLUJOS DE TRABAJO) VERSUS A2A (*AGENT TO AGENT*) Y SU APLICACIÓN EN LOS SECTORES DE EDUCACIÓN, INDUSTRIA Y SERVICIOS

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

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

CARLOS DE LA O FONSECA

<https://orcid.org/0009-0002-3413-4990>

cdelaof@uia.ac.cr

Escuela de Ingeniería Informática, Universidad Internacional de las Américas,
Costa Rica

CARLOS AGUILAR MORA

<https://orcid.org/0000-0003-0946-6053>

chaguilarm@uia.ac.cr

Escuela de Ingeniería Informática, Universidad Internacional de las Américas,
Costa Rica

Recibido: 26 de setiembre del 2025 / Aceptado: 23 de octubre del 2025

doi: <https://doi.org/10.26439/interfases2025.n022.8430>

RESUMEN. El desarrollo de arquitecturas agénticas de inteligencia artificial (IA) ha dado lugar a nuevos modelos de interacción y automatización, entre los que destacan los flujos

de trabajo basados en agentes de IA (*agentic workflows*) y la comunicación entre agentes A2A (*agent-to-agent*). Se presenta un estudio comparativo entre ambas arquitecturas, analizando sus principios de diseño, capacidades de coordinación humano-IA, escalabilidad y adaptabilidad en distintos entornos, así como sus obstáculos y desafíos. El análisis evidencia que, si bien los *workflows* agénticos proporcionan eficiencia y control en entornos estables con procesos delimitados, las arquitecturas A2A sobresalen en contextos distribuidos y heterogéneos, al ofrecer mayor flexibilidad y autonomía. Todo ello bajo la premisa de que la tecnología debe fortalecer, y no sustituir, las capacidades humanas, generando un impacto significativo en los ámbitos en los que más se requiere. Se examinan ejemplos representativos en los sectores de educación, salud e industria y se analiza cómo estas arquitecturas transforman procesos clave. Se presentan figuras y tablas comparativas que integran diversos enfoques entre ambas arquitecturas agénticas de IA. Finalmente, se discuten los desafíos técnicos y éticos asociados con su implementación y se plantean líneas futuras de investigación para una adopción responsable y eficaz de estas tecnologías.

PALABRAS CLAVE: agentes / educación / flujos de trabajo / inteligencia artificial / tecnología

AI AGENT ARCHITECTURES: A COMPARATIVE STUDY OF WORKFLOWS VERSUS A2A (AGENT TO AGENT) AND THEIR APPLICATION IN THE EDUCATION, INDUSTRY AND SERVICES SECTORS

ABSTRACT. The development of agentic architectures for artificial intelligence (AI) has led to new models of interaction and automation, most notably AI agent-based workflows and agent-to-agent (A2A) communication. This paper presents a comparative study of both architectures, analyzing their design principles, human-AI coordination capabilities, scalability and adaptability in different environments, as well as obstacles and challenges. The analysis shows that while agentic workflows provide efficiency and control in stable environments with well-defined processes, A2A architectures excel in distributed and heterogeneous contexts, offering greater flexibility and autonomy. Based on the premise that technology should strengthen, not replace, human capabilities, this study aims to generate a significant impact in the areas where it is most needed. Representative examples from the education, healthcare, and industrial sectors are examined, demonstrating how these architectures transform key processes. Comparative figures and tables are presented, integrating various approaches between the two agentic AI architectures. Finally, the technical and ethical challenges associated with their implementation are discussed, and future lines of research are proposed for a responsible and effective adoption of these technologies.

KEYWORDS: agents / artificial intelligence / education / technology / workflows

INTRODUCCIÓN

En el ámbito de las ciencias de la computación, la inteligencia artificial (IA) se puede definir como una disciplina orientada al diseño de sistemas informáticos y modelos algorítmicos capaces de reproducir o emular procesos cognitivos propios de la inteligencia humana. Los sistemas de automatización inteligente, que combinan procesos automatizados e inteligencia artificial para gestionar flujos de trabajo estructurados, constituyen la base de los agentes de IA modernos.

Un agente de IA generativa se puede definir como una aplicación autónoma diseñada para alcanzar un objetivo específico, que consigue al percibir su entorno y contar con herramientas disponibles para tomar decisiones y ejecutar acciones. Puede razonar y actuar sin supervisión humana directa ni instrucciones explícitas (Viswanathan et al., 2025).

Estos agentes de IA requieren autonomía y flexibilidad para operar en entornos dinámicos, impredecibles y sociales, siendo su objetivo tomar decisiones proactivas y apoyar al usuario en tareas, enseñanza y supervisión. La capacidad de aprendizaje, esencial para el comportamiento inteligente, les permite adaptarse a entornos desconocidos y cambiantes mediante técnicas sofisticadas. Se vislumbra un futuro en el que estos sistemas mejoran significativamente sectores como la industria, la salud, la educación y la resolución de problemas complejos, generando un impacto real más allá de la mera automatización (Shoham & Leyton-Brown, 2008).

De acuerdo con Xu y King (2025), la IA se encuentra en un punto de inflexión, pues sus rápidos avances están transformando industrias, economías y sociedades, al tiempo que generan una serie de desafíos complejos que requieren atención tanto por parte de la comunidad investigadora como de los profesionales del sector. Desde una perspectiva técnica, la IA presenta limitaciones significativas en cuanto a generalización, interpretabilidad, robustez y eficiencia energética.

Si bien los sistemas actuales muestran un alto rendimiento en dominios específicos, aún enfrentan dificultades en el aprendizaje por transferencia, el razonamiento bajo incertidumbre y la integración coherente de información multimodal. Esta brecha evidencia la necesidad de innovar en algoritmos, arquitecturas y paradigmas de aprendizaje.

Ramalingam (2025), basado en su experiencia implementando agentes de IA, ha enfrentado diversos obstáculos, entre los que destaca los siguientes:

- La integración con sistemas heredados y el cumplimiento de regulaciones estrictas.
- La resistencia a la reingeniería de procesos dentro de los flujos de trabajo establecidos complica la adopción de soluciones impulsadas por IA.
- Los beneficios potenciales incluyen niveles sin precedentes de eficiencia y escalabilidad.

Bajo la premisa de que la tecnología de IA debe fortalecer, y no sustituir, las capacidades humanas, a fin de generar un impacto significativo en los ámbitos en los que más se requiere, es fundamental que se desarrollen máquinas inteligentes que aprendan de la experiencia, se adapten a nuevas condiciones y ejecuten tareas tradicionalmente asociadas al pensamiento humano. Por ejemplo, el reconocimiento del lenguaje hablado, la toma de decisiones, la traducción automática o el procesamiento visual mediante visión por computadora, entre otros usos.

Se evidencia la necesidad de avanzar hacia una IA agéntica que integre de manera eficaz el procesamiento y razonamiento con la ejecución autónoma, lo cual permitiría transformar la interacción entre humanos y máquinas, optimizando los flujos de procesos y la orquestación colaborativa entre agentes autónomos, además de potenciar la automatización inteligente.

El foco principal del trabajo descrito en este artículo son las arquitecturas agénticas de IA, particularmente los modelos de flujos de trabajo (*AI workflows*) y los protocolos entre agentes A2A: no solo de manera comparativa, sino como herramientas conceptuales y prácticas que sirvan para avanzar teórica y prácticamente en el campo de la IA.

Los términos *AI workflow* y flujo de trabajo de IA se emplean de manera intercambiable en este artículo, aludiendo al mismo significado.

Este artículo se subdivide en los siguientes temas.

- Automatizando con IA: despliegue y limitaciones, agentes de IA, colaboración humano-IA, marco de referencia SPAR y la metáfora del árbol de IA.
- *AI workflows* y protocolos A2A: dimensiones de comparación, operando con flujos de trabajo de IA y agentes A2A, relevancia de los enfoques y tabla comparativa de características clave.
- Ejemplos de automatización en educación, salud, industria y servicios.
- Discusión y resultados: implicaciones para el diseño de sistemas inteligentes (*workflow* y A2A), consideraciones éticas y sociales de la adopción de estas arquitecturas.

Como productos de esta investigación, se ofrecen también tablas, figuras y glosarios.

OBJETIVOS DE LA INVESTIGACIÓN

Examinar las arquitecturas agénticas de inteligencia artificial, los flujos de trabajo automatizados (*AI workflows*) y la interacción de agentes autónomos (A2A), con el fin de comprender su estructura funcional, capacidades operativas, coordinación, escalabilidad y potencial de aplicación en los sectores de salud, educación, industria y servicios.

Entre los objetivos específicos de la investigación, se consideran los siguientes:

1. Caracterizar conceptualmente las arquitecturas *AI workflows* y A2A, sus fundamentos teóricos, modelos de diseño y principios operativos en el marco de la inteligencia artificial distribuida.
2. Analizar las diferencias estructurales y funcionales entre ambas arquitecturas, las dimensiones de autonomía, escalabilidad, adaptabilidad, coordinación y eficiencia.
3. Explorar aplicaciones representativas de cada arquitectura en los sectores educativo, industrial y de servicios, estudios de caso e impactos en los procesos organizacionales.
4. Proponer orientaciones teóricas y metodológicas que contribuyan al desarrollo, implementación y evaluación de arquitecturas agénticas de IA desde una perspectiva crítica, ética y contextualizada.

Se busca responder a la pregunta de investigación: ¿cómo se comparan las arquitecturas agénticas de *workflow* y de agente-a-agente (A2A) en términos de autonomía, escalabilidad, adaptabilidad y eficacia en su aplicación a los sectores de educación, salud, industria y servicios?

METODOLOGÍA

Se aplicará un enfoque metodológico comparativo, exploratorio y documental, utilizando criterios de comparación entre modelos de *workflows* agénticos y protocolos de agente-a-agente (A2A), aplicado a los sectores educación, salud, industria y servicios.

La revisión de literatura (análisis de documentos) se realizó utilizando los criterios de búsqueda en temas sobre *workflows* agénticos y protocolos A2A y sus impactos en diversos sectores.

Se accedió a diferentes bases de datos, tales como Google Académica, ProQuest Digital Dissertation and Theses, IEEE Xplore y Academia.edu.

Los criterios de selección aplicados se definieron utilizando búsquedas con palabras clave sobre IA en *workflows* y A2A, principalmente.

Se excluyen de esta investigación las búsquedas sobre ChatGPT, lenguajes de programación o bases de datos asociadas con el tema de la inteligencia artificial.

AUTOMATIZANDO CON IA

Despliegue y limitaciones de la IA

De acuerdo con Sedat Arslan (2025), la IA está transformando la industria alimentaria al incorporar técnicas de aprendizaje automático, procesamiento de lenguaje natural y redes neuronales para fortalecer tanto la seguridad como la nutrición. Ofrece recomendaciones individualizadas mediante la detección de contaminantes, la optimización del almacenamiento y la trazabilidad con *blockchain*¹, basadas en datos genéticos, metabólicos, del microbioma y del estilo de vida. Tenemos nutricionistas digitales impulsados por IA ofreciendo apoyo dietético en ambientes en los que existe un acceso limitado a profesionales de la salud.

Surge acá una paradoja: muchas organizaciones que buscan implementar automatizaciones se ven restringidas por el hecho de que, pese a contar con sistemas de IA generativa altamente sofisticados en su capacidad de razonamiento, estos carecen en la práctica de la habilidad para ejecutar acciones efectivas. Pueden analizar datos complejos en segundos, crear presentaciones convincentes y ofrecer perspectivas sobre cualquier tema. Sin embargo, aunque la IA ha alcanzado un alto nivel de razonamiento, sigue sin poder actuar de manera autónoma, lo que obliga a que los trabajadores del conocimiento dediquen gran parte de su tiempo (hasta 60 %) a tareas mecánicas, como supervisar y ejecutar manualmente las recomendaciones generadas por estos sistemas. “Tratamos a los humanos como robots y a la IA como creativos y es hora de invertir la ecuación” (Bornet et al., 2025, p. 2).

A manera de ejemplos ilustrativos, se presentan tres casos sobre automatizaciones con IA: empresa de servicios, industria alimentaria y robots de inspección inteligentes.

- Una empresa del sector servicios enfrentaba dificultades para optimizar su atención al cliente. Frente a ello, implementó un chatbot de alta precisión que interpreta y responde consultas, junto con bots de automatización de procesos (RPA)² que ejecuta secuencias dentro de flujos de trabajo predefinidos. No obstante, existía una limitación clave: la falta de un enlace efectivo entre la comprensión proporcionada por el chatbot y la ejecución automática de acciones. El personal de atención al cliente seguía actuando como intermediario, trasladando manualmente las recomendaciones del chatbot y

1 La cadena de bloques (*blockchain*) constituye un registro digital en continua expansión, integrado por múltiples bloques de datos organizados de forma cronológica y enlazados entre sí mediante mecanismos criptográficos que garantizan su integridad y seguridad (Binance Academy, s. f.).

2 La RPA (*robotic process automation*) es un método de automatización de los procesos de negocio. Emplea robots de *software* (bots) para automatizar tareas digitales que suelen realizar los humanos. Trabaja junto con herramientas de IA, lo que incluye la IA generativa (Automation Anywhere, 2025).

activando los flujos de trabajo. La próxima evolución de la IA pretende mejorar la capacidad de razonamiento, con mayor autonomía y capacidad para actuar de forma independiente.

- En la industria alimentaria, la IA estaba revolucionando los procesos mediante la aplicación de aprendizaje automático, procesamiento de lenguaje natural y redes neuronales para generar recomendaciones nutricionales personalizadas y mejorar la trazabilidad a través de tecnología blockchain. A pesar de que ya poseían una alta capacidad de razonamiento, los sistemas de IA generativa carecían de autonomía para ejecutar acciones, lo que obligaba a los trabajadores a realizar tareas repetitivas y supervisar manualmente los resultados.
- Los robots de inspección inteligentes han demostrado sus ventajas en diversos escenarios, como galerías de tuberías, túneles y subestaciones eléctricas, pero su aplicación generalizada aún es insuficiente y su desarrollo diversificado requiere mejoras urgentes. Se han logrado avances significativos en el control y monitoreo remoto, pero persisten ciertas limitaciones, tales como problemas de seguridad, procesamiento, análisis de datos y respuesta en tiempo real, entre otros. Debido a la complejidad e inestabilidad en los sistemas de monitoreo remoto multirobot, se han propuesto arquitecturas multiagente adaptadas a la teleoperación y con mayor autonomía realizando tareas (Li et al., 2025).

Estos tres ejemplos ponen de manifiesto un desfase entre comprensión y acción, junto con la necesidad de desarrollar una inteligencia artificial agéntica que integre de manera coherente el análisis de datos, el razonamiento y la ejecución autónoma. Un avance de este tipo permitiría transformar la interacción entre humanos y máquinas (sectores de servicios e industria alimentaria) o bien optimizar los flujos de procesos y orquestar a múltiples agentes autónomos, como sucede en el caso de los robots inteligentes para inspección remota.

Se evidencia un despliegue a gran escala en soluciones de IA y, de acuerdo con Frank X. Shaw (2025), unos quince millones de desarrolladores utilizan GitHub Copilot para programar, revisar y desplegar código. Cientos de miles de usuarios se sirven de él para hacer investigación y desarrollar soluciones y más de 230 000 organizaciones emplean Copilot Studio para crear agentes de IA y automatizaciones. En suma, no solo están transformando los procesos de desarrollo de *software* de IA, sino también la forma en que los individuos, equipos y organizaciones realizan su trabajo.

En resumen, se expande una visión emergente de la web agentiva abierta, en la que agentes de IA operan de manera autónoma en contextos individuales, organizacionales y empresariales de extremo a extremo, tomando decisiones y ejecutando tareas en representación de los usuarios e instituciones. Sin embargo, aparecen limitaciones con respecto a la habilidad para ejecutar acciones efectivas de manera completamente autónoma.

A pesar de sus limitaciones, la IA se consolida rápidamente como una disciplina central de las ciencias de la computación, orientada a reproducir funciones cognitivas humanas mediante algoritmos capaces de aprender, adaptarse y actuar en entornos complejos.

Agentes de IA

Los agentes de IA son entidades de *software* que emplean técnicas de inteligencia artificial para ejecutar tareas y alcanzar objetivos sin necesidad de entradas de datos explícitos ni resultados predefinidos. Pueden recibir instrucciones, diseñar planes de acción, utilizar herramientas para completar tareas y generar resultados dinámicos.

Las arquitecturas de IA apoyada con agentes están reconfigurando procesos fundamentales: desde la gestión del conocimiento y la personalización del aprendizaje, hasta la automatización industrial y la optimización de los servicios de atención al cliente.

¿Cuándo aparecieron los agentes de IA?

El año 2022 se considera como el nacimiento de los agentes de IA, cuyo campo sigue evolucionando rápidamente y casi a diario surgen nuevas posibilidades, con el objetivo de explicar estas nuevas tecnologías, y también de ofrecer herramientas para las personas y empresas; todo con el propósito de construir un mundo mejor, en el que la computación basada en agentes constituye un dominio científico con amplia posibilidad de difusión. Se busca superar la brecha entre el razonamiento y la acción, por lo que se requiere avanzar hacia una inteligencia artificial de carácter agéntico, capaz de integrar de forma autónoma los procesos de comprensión y ejecución, como una nueva etapa en la relación entre tecnología, conocimiento y práctica humana.

¿Cómo funcionan los agentes de IA?

Existen dos tipos de agentes de IA: la generativa, cuyo enfoque principal es la producción de contenido, y la IAA (inteligencia artificial agéntica), concebida para actuar de forma proactiva en representación de los usuarios, minimizando (o eliminando) la necesidad de contar con una supervisión humana y adquiriendo la capacidad de tomar decisiones y ejecutar acciones de manera autónoma (Empresa Actual, 2025).

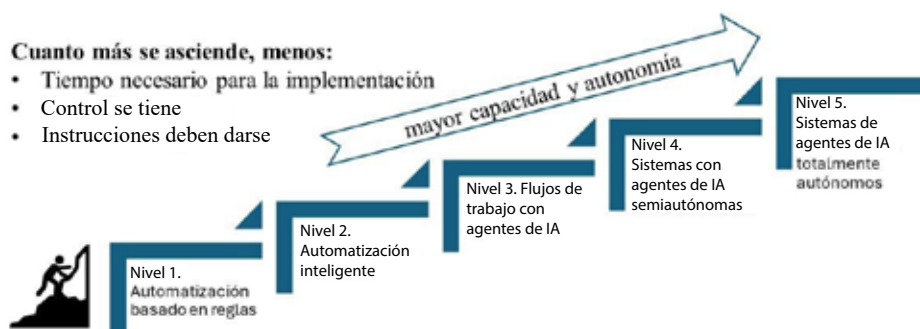
De acuerdo con Coshov (2024), la IAA emplea sensores o diversas fuentes de datos para recopilar información en tiempo real del entorno en el que opera. Posteriormente, el sistema utiliza algoritmos de aprendizaje automático y redes neuronales avanzadas para procesar la información, reconocer patrones y formular predicciones sobre escenarios futuros. La IAA mejora su rendimiento a través de un aprendizaje continuo, incorporando mecanismos de acción que le permiten ejecutar tareas específicas de acuerdo con los objetivos definidos y mantener una interacción efectiva con su entorno.

¿Hay diferentes niveles de automatización con agentes de IA?

Para comprender los niveles de automatización y de autonomía de los agentes de IA, Bornet et al. (2025) proponen cinco niveles para medir la automatización (más que madurez) del uso de la inteligencia artificial dentro de una organización (ver Figura 1).

Figura 1

Niveles de progresión de automatización con la IA



Nota. Adaptado de *Agentic Artificial Intelligence: Harnessing AI Agents to reinvent Business, Work and Life* (p. 78, Figura 1.3. The Agentic AI Progression Framework), por P. Bornet, J. Wirtz, T. H. Davenport, D. De Cremer, B. Evergreen, P. Fersht, R. Gohel y S. Khiyara, 2025, World Scientific Publishing Company.

De acuerdo con este marco de ascenso o progresión de la automatización con IA, el mejor nivel no necesariamente es el último (nivel 5), sino que depende sobre todo de las necesidades y circunstancias específicas de la organización y del nivel de impacto deseado. Esto es similar a contar con un catálogo de diferentes tipos de agentes, cada uno entrenado y adecuado para las necesidades y contextos específicos en los que cada uno se ubicaría en diferente nivel.

Por ejemplo, los automóviles modernos que vienen equipados con asistentes para la conducción automatizada: si bien es posible técnicamente activar la conducción totalmente autónoma en una autopista, muchos conductores prefieren ser más precavidos y activan solo el control de cruce básico (por ejemplo, los niveles 1 o 2).

Es posible diseñar catálogos complejos con múltiples funciones para los automóviles, el hogar y la industria, pero el humano es quien decide, finalmente, el nivel de automatización de IA requerido para cada caso específico. El marco de progresión anterior viene a servir de guía para ubicar el nivel esperado al realizar una propuesta de automatización.

COLABORACIÓN HUMANO-IA

En las últimas décadas, los avances en investigación sobre interacción humano-computadora (HCI) e inteligencia artificial (IA) han transformado las formas de colaboración

entre las personas y los agentes de *software* inteligentes. El desarrollo de un marco teórico sólido que explique la relación entre el comportamiento humano y el computacional sigue en investigación. A medida que las tecnologías de IA se materializan en sistemas y establecen interacciones directas con los humanos, estos conceptos se vuelven fundamentales. Comprender la autonomía de la IA es crucial para el control y la influencia colaborativa, las relaciones de confianza entre humanos e IA, y la gestión de los flujos de trabajo (Samdani et al., 2025).

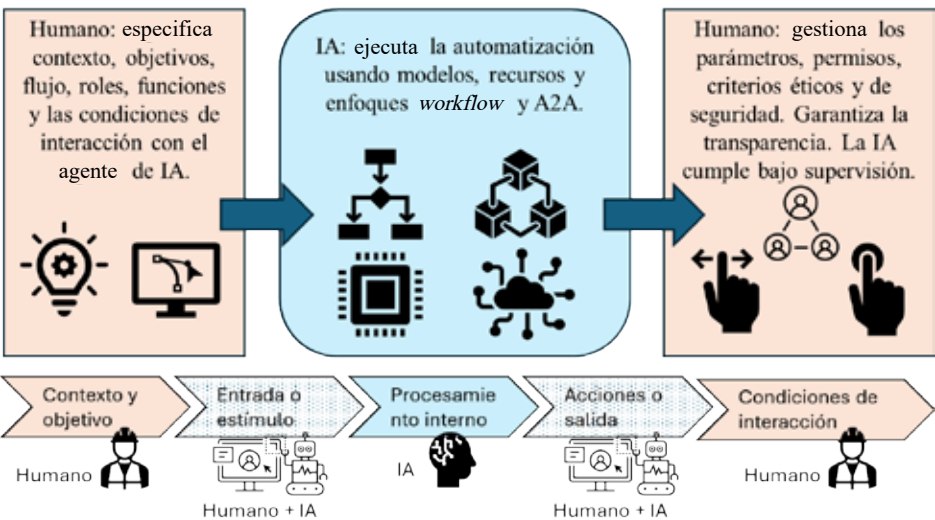
La interacción humano-IA puede ubicarse en un espectro de niveles de autonomía, que va desde la actuación independiente hasta la supervisada, en la que la intervención humana plantea diversos desafíos éticos y legales (entre los cuales se encuentran la falta de confianza, la descoordinación y la disparidad en el control de las decisiones).

La colaboración humano-IA, por otro lado, se fundamenta en la interacción entre usuarios y sistemas de IA, que abarcan una amplia variedad de contextos de actividad.

Por ello, resulta esencial garantizar la transparencia en el diseño de las soluciones de IA y definir con claridad las responsabilidades en los procesos de toma de decisiones.

Al respecto, se propone un marco de referencia para la colaboración humano-IA, el cual se estructura en cinco etapas que describen la interacción entre una persona humana y un entorno compuesto por agentes de IA. La Figura 2 muestra esta secuencia, conformada por los siguientes pasos: contexto y objetivo, entrada o estímulo, procesamiento interno, acciones o salidas, y condiciones de interacción.

Figura 2
Colaboración humano-IA



A continuación, se describen en detalle los pasos y sus elementos.

- Paso 1 (contexto y objetivo). El humano delimita con claridad el propósito de la interacción con el agente, el contexto, objetivos, alcance, flujo, roles, funciones y las condiciones del dominio de aplicación y las tareas que se espera que ejecute el agente o el workflow.
- Paso 2 (entrada y estímulo). El humano provee y estructura los datos que la IA interpreta. Debe especificar la naturaleza y origen de los datos que el agente de IA recibirá, así como el formato y la estructura de la información (texto, voz, imágenes, sensores), además de incluir las condiciones iniciales que orientan su procesamiento.
- Paso 3 (procesamiento interno). Enfoques de A2A o workflow, a través de los cuales la IA ejecuta de acuerdo con el modelo de razonamiento preestablecido (basado en reglas, aprendizaje automático o generación), con recursos de conocimiento, métodos de adaptación o aprendizaje continuo.
- Paso 4 (acciones o salida). La IA, bajo supervisión humana, incluye el modo en que el agente generará respuestas (textuales, visuales o a través de la ejecución de acciones) e indicará el nivel de autonomía con el que actúa y los mecanismos de retroalimentación previstos. La IA genera las salidas correspondientes en respuestas textuales, visuales o acciones ejecutables.
- Paso 5 (condiciones de interacción). El humano preestablece lo que la IA cumple dentro de los parámetros definidos. El humano gestiona los roles y permisos que delimitan la acción del agente, los criterios éticos y de seguridad que regulan la protección de datos, la mitigación de sesgos y el cumplimiento normativo para garantizar la transparencia en la interacción.

Vemos entonces que la interacción con un agente de IA requiere una distribución clara de responsabilidades entre el usuario humano y el propio sistema. En esta colaboración, la IA asume el procesamiento interno mediante modelos de razonamiento, recursos de conocimiento y mecanismos de adaptación. Las acciones se realizan bajo la supervisión del usuario humano, quien conserva el rol de garante ético y regulador, al definir los criterios de seguridad, privacidad, transparencia y explicabilidad que guían la operación del agente.

A continuación, se analizan dos casos de colaboración humano-IA:

- Un estudio de Casini et al (2023) con modelos de segmentación semántica como medio de colaboración entre humanos con la IA, permitieron detectar sitios arqueológicos en las llanuras de Mesopotamia. Al usar modelos de aprendizaje profundo preentrenados y apoyados en imágenes satelitales, fue posible detectar sitios arqueológicos con una precisión cercana al 80 %. La

colaboración entre expertos en arqueología y científicos de datos fue crucial para la creación y evaluación del conjunto de datos con el que se contaba. Se mostró que la integración de la experiencia, conocida como humanos en el bucle (*human in the loop*), fue esencial para mejorar la precisión de un modelo de IA y refinar el conjunto de datos con nuevas anotaciones basadas en las predicciones del modelo. Para optimizar la detección de futuros sitios arqueológicos, se propone modelar un flujo de trabajo de IA agéntico que combine las predicciones del modelo junto con la experiencia humana.

- Otra experiencia consistió en integrar agentes de IA a través de enfoques empresariales (*Enterprise AI - EAI*), en combinación con sistemas de intercambio electrónico de datos (*Electronic Data Interchange - EDI*). Se buscó potenciar la eficiencia operativa y optimizar los procesos de toma de decisiones, estimando mejoras de hasta un 30 % en el desempeño global. Para la interacción con los clientes, se incorporó un chatbot de servicio como interfaz, respaldado por un agente orquestador de IA que asignaba tareas a agentes especializados. Finalmente, la capa de gobernanza cierra el ciclo, asegurando la integridad del sistema y el cumplimiento de las normativas vigentes (Ramalingam, 2025).

De acuerdo con Viswanathan et al. (2025), se espera que el futuro de los agentes de inteligencia artificial experimente avances significativos en cuanto a autonomía, adaptabilidad y colaboración humano-IA. Estos agentes autónomos y automejorables emplearán técnicas como el aprendizaje por refuerzo, el aprendizaje continuo y el meta-aprendizaje para optimizar la toma de decisiones en entornos dinámicos.

De acuerdo con Balic (2005), los sistemas autónomos de IA multiagente están llamados a transformar diversos sectores, en particular la industria del desarrollo de *software* y las actividades basadas en el conocimiento. A pesar del rápido avance en sus capacidades técnicas, existe una brecha significativa en la comprensión de cómo los profesionales perciben estos sistemas, incluyendo sus capacidades, limitaciones, implicaciones éticas y el posible impacto en los empleos actuales. Estas percepciones trascienden el ámbito teórico, ya que influyen de manera directa en las tasas de adopción, las decisiones de inversión, y las estrategias de formación y preparación de la fuerza laboral.

En resumen, la interacción humana-IA se concibe como un proceso colaborativo en el que las capacidades cognitivas y normativas del humano enmarcan y orientan la autonomía de la IA.

De la Figura 1 se desprende que el ser humano es quien establece los objetivos, los datos de entrada y los criterios éticos, mientras que la IA (*workflow* o agentes A2A) se encarga del procesamiento y la generación de respuestas. De esta manera,

las capacidades normativas y de supervisión humanas delimitan y condicionan la autonomía técnica del agente, asegurando que su desempeño se enfoque en la eficiencia y la responsabilidad. Para más detalle, ver el Glosario (sección Colaboración humano-IA).

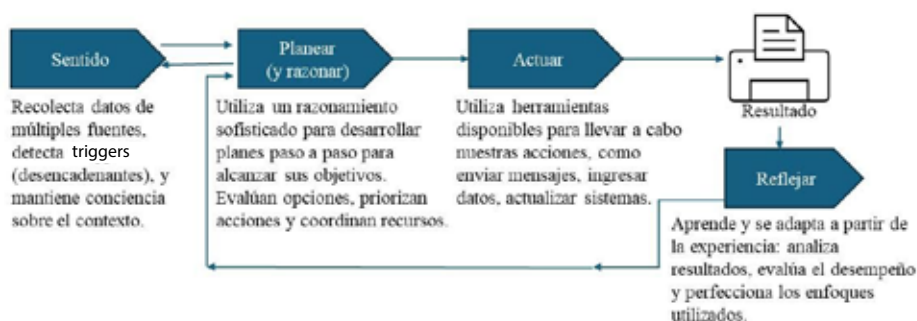
Marco de referencia SPAR

¿Qué sucede al darle a un agente de IA esta instrucción?: “genere el resultado que necesito”.

Para responder a esta pregunta, Pascal Bornet et al. (2025) proponen utilizar el marco integral SPAR (acrónimo de los términos *sense, plan, act* y *reflect*), que ayuda a comprender las capacidades de los agentes de IA. Para analizar el ciclo de comportamiento de un agente de IA, podemos utilizar el marco de referencia SPAR mostrando la secuencia de pasos a seguir.

Figura 2

El marco de referencia SPAR -cómo actúa un agente de IA



Nota. De *Agentic Artificial Intelligence: Harnessing AI Agents to reinvent Business, Work and Life* (p. 23, Figura 1.2. How an AI agent takes action: The SPAR Framework), por P. Bornet, J. Wirtz, T. H. Davenport, D. De Cremer, B. Evergreen, P. Fersht, R. Gohel y S. Khiyara, 2025, World Scientific Publishing Company.

El modelo SPAR inicia con el seguimiento básico de reglas hasta lograr resultados, luego iterar y regresa a planear y actuar de nuevo. Este ciclo de retroalimentación se realiza con suficiente autonomía, lo cual permite al agente desenvolverse ante un panorama complejo, apoyado en decisiones informadas sobre las soluciones más adecuadas para sus necesidades.

A partir de la Figura 2, se describen paso a paso cada una de las acciones de los agentes de IA:

- Sentido (*sensing*):** son los ojos y los oídos de los agentes de IA. Integra datos de diversas fuentes, detecta condiciones desencadenantes y sostiene una

conciencia contextual dinámica del entorno operativo. A manera de ejemplo, un automóvil autónomo necesita comprender su entorno de manera integral. De igual manera, los agentes de IA deben ser capaces de percibir su espacio de trabajo digital. Cuando se introduce una ruta destino en la pantalla del auto, se le está estableciendo el objetivo o destino hacia el cual debe dirigirse.

- b) Planear y razonar (*plan and reason*): emplea procesos de razonamiento avanzado para formular planes estructurados orientados al cumplimiento de objetivos específicos. El agente de IA analiza alternativas, establece prioridades y gestiona los recursos disponibles.
- c) Actuar (*act*): emplea herramientas tecnológicas para ejecutar acciones operativas, como la transmisión de mensajes, el ingreso de datos y la actualización de sistemas. Facilita la implementación de decisiones previamente planificadas y la interacción del entorno digital con otros componentes del sistema.
- d) Reflejar (*reflect*): adquiere conocimiento con cada iteración y mejora su desempeño mediante un proceso de aprendizaje basado en la experiencia. Se incluyen el análisis de resultados, la evaluación del rendimiento y el perfeccionamiento continuo de los enfoques aplicados.

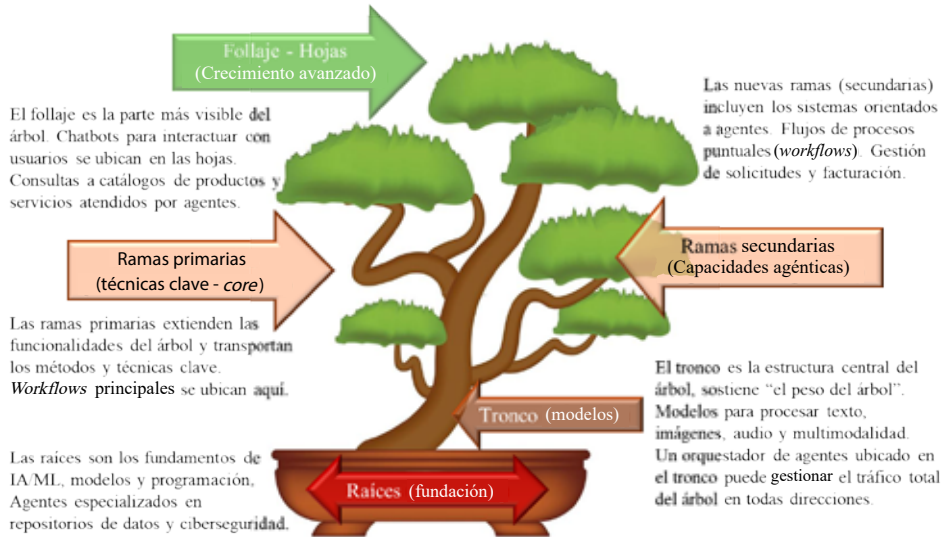
En resumen, el modelo SPAR facilita que los agentes de IA operen mediante un ciclo iterativo compuesto por cuatro fases fundamentales: En la fase de percepción (*sensing*) se recolectan los datos y objetivos. En la fase de planificar y razonar (*plan and reason*), se aplican técnicas de razonamiento para priorizar los pasos a seguir. En la fase de actuación (*act*) se llevan a cabo las acciones. Finalmente, en la fase de reflexión (*reflect*) se logra el aprendizaje con base en la evaluación del desempeño y el análisis de resultados, para luego iterar hacia la fase de planeación. Este ciclo integral SPAR se convierte en una guía de referencia útil para el desarrollo de agentes autónomos de IA capaces de operar eficazmente en entornos dinámicos y complejos.

Metáfora del árbol de IA

Una manera de conceptualizar los componentes, modelos de agentes y flujos de trabajo con IA es por medio de una analogía que denominamos la metáfora del árbol de IA, en la que cada elemento tiene su funcionalidad y ubicación estratégica, simulando una organización.

Figura 4

Metáfora del árbol de IA



¿En cuáles áreas se están desarrollando proyectos de IA? Ante esta pregunta, una manera rápida y simple de responder consiste en apoyarse en la metáfora del árbol de IA. En el árbol podemos ubicar las diferentes iniciativas de automatización con IA:

- **Raíces:** modelos y herramientas de IA a seguir. Agentes de IA especializados en el acceso a los repositorios de datos y esquemas de ciberseguridad.
- **Tronco:** "soporte" del árbol. Modelos para procesar todo tipo y formato de datos. Orquestador de agentes de IA pueden gestionar todo el tráfico dentro del árbol.
- **Ramas primarias:** métodos y técnicas clave. *Workflows* de uso general permiten trasladar información de la raíz hasta las hojas. Por ejemplo, la cadena de abastecimiento (supply chain), gestión de bodegas y almacenes e inventarios.
- **Ramas secundarias:** sistemas y *workflows* puntuales usando agentes de IA. Aplicaciones para gestionar funcionalidades de la organización. Por ejemplo, facturación y solicitudes.
- **Follaje – hojas:** es la parte "visible" del árbol. Los usuarios interactúan por este medio con agentes, gestionando catálogos de bienes y servicios en línea. Si usan "carro de compras", este puede ser otro agente de IA especializado el cual es llamado para gestionar el pago.

La elección final de la ubicación dependerá de una decisión estratégica y de las particularidades de cada caso. Alternativamente, se puede considerar la ubicación de

un agente orquestador en una rama primaria, organizando el flujo específico de *workflows* de IA al resto del árbol.

Un *workflow* para la cadena de abasto puede interactuar con varios agentes y flujos de trabajo, siempre orquestados por un agente central.

En cualquier caso, es fundamental incorporar principios de autonomía y escalabilidad en todos los componentes del sistema, para de esa manera maximizar su eficacia y adaptabilidad.

Esta metáfora del árbol de IA ofrece un marco conceptual, simple y práctico, para gestionar proyectos de IA en empresas, instituciones académicas o sectores industriales. Para más detalle, véase el Glosario (Sección Metáfora del árbol de IA).

Desafíos y ética con la IA

La adopción de la IA enfrenta diversas dificultades, entre las que destacan la privacidad y la seguridad de los datos, especialmente en sectores sensibles. La complejidad inherente a la integración de la IA con sistemas existentes puede extender los plazos de los proyectos. La necesidad de una capacitación adecuada de la fuerza laboral, esencial para garantizar una adopción exitosa, es otro desafío. Las restricciones en costos y recursos, por su parte, pueden ser mitigadas mediante el uso de plataformas en la nube y modelos preentrenados.

De acuerdo con Xu & King (2025), existen otros desafíos con el desarrollo impulsado por la IA:

- como frontera emergente
- alineada con valores humanos
- aprendizaje en contextos de recursos limitados,
- en la investigación científica

Estos desafíos deben ir acompañados de una reflexión ética constante respecto de los sesgos, la transparencia y la rendición de cuentas, a la vez que evaluar, a nivel mundial, la demanda creciente de marcos regulatorios adecuados. En este contexto, lograr un equilibrio entre innovación y gobernanza será crucial para mantener la confianza en general.

Por otro lado, se hace imprescindible instaurar marcos normativos y auditables, como el Reglamento General de Protección de Datos de la Unión Europea (General Data Protection Regulation, GDPR) y la Ley de Portabilidad y Responsabilidad del Seguro Médico (Health Insurance Portability and Accountability Act, HIPAA), ley que regula la confidencialidad y seguridad de la información médica y de salud, además de establecer mecanismos de gobernanza que incluyan la detección de sesgos y una evaluación ética previa a su implementación.

En este contexto, la convergencia entre tecnología y gestión pública transforma constantemente los paradigmas tradicionales e impulsa la apertura de nuevos escenarios de oportunidad, lo cual permite optimizar procesos y aumentar la eficiencia, pero también habilitar espacios para la innovación, el análisis y el aprovechamiento de grandes volúmenes de datos (Ramalingam, 2025).

Dentro del ámbito académico, Corvalán & Sánchez Caparrós (2025) señalan que uno de los desafíos centrales consiste en la formación de profesionales con la capacidad de adaptarse y gestionar, con base en criterios éticos y de eficiencia, el diseño e implementación de sistemas emergentes basados en componentes de inteligencia artificial.

Según Hauser (2019), la reflexión ética sobre la tecnología, y en particular sobre la IA, ha tendido a aplicarse directamente desde teorías en principio concebidas para los seres humanos hacia los sistemas de IA. Un ejemplo de esta tendencia es el enfoque que considera a la IA como un instrumento peligroso: se enfatizan las diferencias entre los sistemas de IA y los humanos bajo el argumento de que dichos sistemas no deben ser considerados agentes morales, pues las teorías de agencia ética fueron desarrolladas únicamente para sujetos humanos.

Desde la academia, se propone abordar la ética desde una perspectiva más amplia, centrada en los sistemas de información, considerando a la IA como un subconjunto dentro de este marco. Esta aproximación facilita la incorporación de los estudios de la información en el análisis de los desafíos éticos planteados por las tecnologías de IA.

Asimismo, resulta fundamental comprender las percepciones de los profesionales en estos ámbitos para anticipar los retos relacionados con la adopción de una IA segura, las implicaciones éticas y la evolución del mercado laboral.

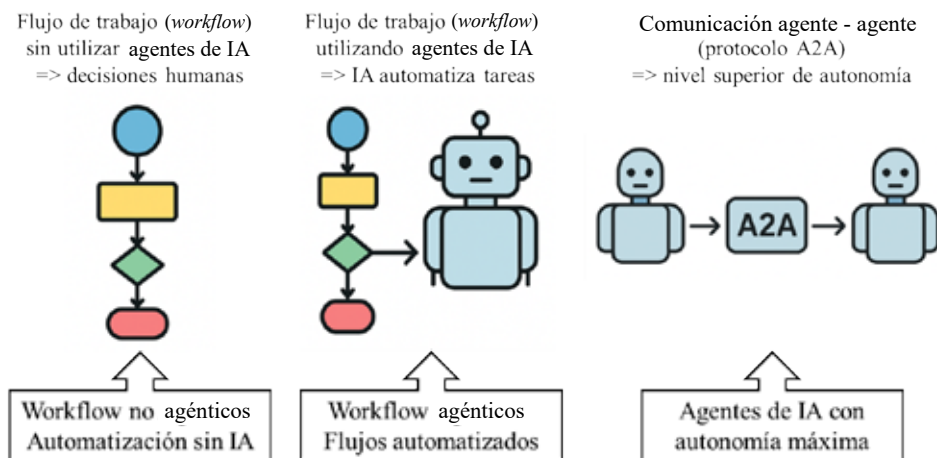
AI WORKFLOWS Y PROTOCOLOS A2A

Categorizando *workflows* de IA y agentes A2A

Michael Lieben (2025) hace una propuesta para categorizar *workflows* y agentes de IA, utilizando tres tipos de agentes de IA. Estos se muestran en la Figura 5.

Figura 5

Tipos de workflows y agentes de IA



Nota. Adaptado del flujograma *3 Types of AI Workflows*, por M. Lieben, 2025. LinkedIn. <https://www.linkedin.com/posts/the-best-of-ai-ai-agents-are-replacing-human-employees-activity-7364786657318838273-SH-/>

Las principales diferencias se analizan a continuación.

- Workflows no agénticos:** se distinguen porque la responsabilidad de las decisiones recae enteramente en el ser humano, mientras que la inteligencia artificial se limita a actuar como herramienta de apoyo. En este esquema, el usuario dirige cada acción. Por ejemplo, cuando se consulta un sistema de gestión y luego se solicita a un modelo de lenguaje la redacción de un correo, que finalmente es enviado por el humano de manera manual. La IA, en este caso, cumple una función instrumental, subordinada al control humano.
- Workflows agénticos:** la inteligencia artificial se integra dentro de un sistema regido por reglas y disparadores que permiten automatizar determinadas tareas. Así, por ejemplo, el propio sistema puede detectar la necesidad de programar una reunión, verificar la disponibilidad en la agenda y generar automáticamente un correo de invitación. Este tipo de flujos incrementa la eficiencia operativa, aunque su capacidad de respuesta es limitada ante situaciones no previstas.
- Agentes de IA:** nivel superior de autonomía. Actúan en función de metas establecidas, como la satisfacción del cliente. Estos agentes son capaces de interpretar datos contextuales, monitorear interacciones previas, generar comunicaciones personalizadas, dar seguimiento a la respuesta del interlocutor y coordinar actividades posteriores, como la elaboración de una agenda. A diferencia de los flujos anteriores, los agentes no solo ejecutan tareas, sino que articulan acciones estratégicas orientadas a objetivos.

De acuerdo con Corvalán & Sánchez Caparrós (2025), los flujos de trabajo automatizados que integran modelos de lenguaje combinan herramientas y modelos de IA bajo rutas de código establecidas. Los agentes de IA, en contraste, poseen mayor grado de autonomía al no basarse en rutas preestablecidas y siendo más adaptables a situaciones de decisión.

La progresión de los sistemas de IA puede comprenderse en tres niveles:

- En primer lugar, los flujos de trabajo sin agentes, en los que la IA cumple un papel estrictamente instrumental bajo control humano.
- En segundo término, los flujos agénticos (*workflows*), caracterizados por la automatización de tareas mediante reglas y disparadores, lo que incrementa la eficiencia, pero con limitada adaptabilidad frente a contingencias.
- Finalmente, los agentes A2A introducen un grado superior de autonomía al operar con metas definidas, interpretar información contextual y coordinar acciones orientadas a objetivos estratégicos. De este modo, se consolidan como un avance sustantivo en la transición de la automatización hacia la toma de decisiones inteligente (The Best of AI, 2025).

Comparando *workflows* de IA y protocolos A2A

Para llevar a cabo el análisis comparativo, utilizaremos los criterios de autonomía, escalabilidad, flexibilidad y coordinación, bajo los cuales ubicaremos las características de *workflows* y A2A.

Autonomía y adaptabilidad

Workflows

- Los AI *workflows* generalmente siguen una secuencia de pasos predefinidos, aunque pueden incorporar componentes de IA (como modelos predictivos o generativos) que permitan cierta adaptabilidad.
- Su autonomía depende del diseño, que suele requerir supervisión humana.
- La adaptabilidad está limitada por las reglas y modelos integrados, y responde a cambios dentro de lo previsto, pero no a contextos totalmente nuevos.

A2A

- Los protocolos A2A poseen autonomía inherente, toman decisiones en función de su entorno y objetivos.
- Son altamente adaptativos y pueden modificar su comportamiento al interactuar con otros agentes y aprender de la dinámica del sistema.
- Permiten mayor descentralización y autogestión que los flujos tradicionales.

- En cuanto a autonomía y adaptabilidad, la ventaja la tiene el protocolo A2A, por su naturaleza autoorganizativa y adaptativa.

Escalabilidad y eficiencia

Workflow

- Los AI workflows escalan de manera lineal. Añadir más procesos o nodos implica extender el flujo, lo que a su vez puede aumentar la complejidad.
- La eficiencia depende de la optimización del pipeline y del poder de cómputo.
- Funcionan bien en entornos centralizados (por ejemplo, en el procesamiento de grandes volúmenes de datos).
- Los AI workflows son mejores en sistemas centralizados de alto volumen.

A2A

- Los protocolos A2A escalan de manera distribuida: cada agente gestiona sus tareas y la carga se reparte entre múltiples entidades.
- La eficiencia se puede ver afectada por la sobrecarga de comunicación entre agentes, especialmente en sistemas muy grandes.
- Se desempeñan mejor en escenarios descentralizados o dinámicos (IoT, redes inteligentes).
- Los protocolos A2A son mejores en entornos distribuidos dinámicos.

Flexibilidad y robustez

Workflow

- Los AI workflows son flexibles en cuanto a la integración de distintos módulos o servicios (por ejemplo, IA para visión, procesamiento del lenguaje natural - NLP, análisis de datos).
- Poco robustos ante fallos inesperados fuera del diseño del flujo. Por tanto, una interrupción puede afectar todo el proceso.
- La resiliencia depende de redundancias planificadas.

A2A

- Los protocolos A2A son altamente flexibles; los agentes pueden reorganizarse, asumir roles distintos y redistribuir tareas.
- Robustos frente a fallos locales: si un agente cae, otros pueden compensar o reorganizarse.
- Adecuados para entornos inciertos o con cambios frecuentes.
- A2A es más ventajosa, por su descentralización y resiliencia.

Coordinación y comunicación

Workflows

- Los AI workflows poseen una coordinación jerárquica y secuencial, en la que cada etapa depende de la anterior.
- Comunicación limitada, generalmente de tipo unidireccional (paso de datos de un módulo a otro).
- Adecuados para procesos determinísticos o en cadena.

A2A

- Los protocolos A2A coordinan de manera colaborativa y negociada; los agentes intercambian información, establecen acuerdos y se adaptan en tiempo real.
- Comunicación rica, bidireccional y en ocasiones con protocolos complejos (cooperación, competencia y negociación).
- Ideales para entornos heterogéneos y con múltiples actores.
- A2A es más ventajosa, dada su mayor comunicación y cooperación dinámica.

Resumen de los criterios anteriores

- En los criterios de autonomía y adaptabilidad, la A2A es más ventajosa.
- En escalabilidad y eficiencia, los workflows de IA son mejores en sistemas centralizados de alto volumen y los A2A lo son en entornos distribuidos dinámicos.
- En cuanto a flexibilidad y robustez y a coordinación y comunicación, A2A es más ventajosa.

Usos prácticos de *workflows* y agentes de IA

Utilizando los criterios anteriores (autonomía, escalabilidad, flexibilidad y coordinación) se analizarán algunos ejemplos de la operación con IA *workflows*.

- Autonomía y adaptabilidad. Los flujos de trabajo de IA en el ámbito bancario, como la evaluación de solicitudes de crédito, siguen pasos predefinidos (verificación documental, análisis de scoring y decisión final). Su capacidad de adaptación es limitada y depende de reconfiguraciones humanas cuando se enfrentan a nuevos formatos o condiciones no previstas. Si aparece un documento en un formato nuevo, el flujo podría fallar o requerir intervención humana para actualizar la lógica.
- Escalabilidad y eficiencia. En plataformas de comercio electrónico, los workflows permiten procesar transacciones masivas de manera centralizada,

aunque escalar requiere incrementar infraestructura (por ejemplo, más servidores). La eficiencia se sostiene en entornos estables y controlados y escalar implica añadir más servidores que repliquen el flujo.

- Flexibilidad y robustez. En el ámbito hospitalario, un workflow que integra IA para diagnóstico puede verse interrumpido si uno de sus módulos falla (por ejemplo, el de procesamiento de lenguaje NLP para historiales clínicos); se compromete la totalidad del proceso y el flujo entero puede interrumpirse.
- Coordinación y comunicación. En procesos de reclutamiento automatizado, la coordinación se estructura de forma jerárquica y secuencial (filtrado de currículos, ranking, entrevistas automatizadas), con comunicación secuencial y del tipo unidireccional entre las etapas.

Un caso específico es el de la industria 4.0, en la que la convergencia entre tecnologías digitales y físicas está transformando el *workflow* de los procesos de manufactura. Entre estos, la integración de inteligencia artificial (IA) con la manufactura aditiva (AM) destacan como ejes fundamentales de innovación, permitiendo un control preciso de cada capa en la fabricación de productos: desde la recopilación de datos en tiempo real mediante sensores hasta la optimización de parámetros de diseño, geometría y desempeño estructural. Esta relación sinérgica entre IA y AM no solo resulta esencial para alcanzar la visión de la industria 4.0, sino que también se proyecta hacia la industria 5.0, caracterizada por la sostenibilidad, la colaboración humano-máquina y la hiper personalización de los procesos productivos (Bassey et al., 2025).

Algunos ejemplos de la operación utilizando protocolos A2A son los siguientes:

- Autonomía y adaptabilidad: en entornos de negociación financiera, los agentes A2A muestran alta autonomía al ajustar sus decisiones de compra y venta en tiempo real, coordinando con otros agentes y adaptándose a la aparición de nuevos activos sin necesidad de rediseñar todo el sistema.
- Escalabilidad y eficiencia: en sistemas de redes eléctricas inteligentes (smart grids), cada agente gestiona su consumo y producción local, logrando escalabilidad distribuida sin depender de un nodo central. La eficiencia se mantiene en contextos dinámicos mediante la optimización local y, al aumentar el número de dispositivos, la coordinación sigue siendo posible sin necesidad de un controlador central.
- Flexibilidad y robustez: en escenarios en los que se utilicen robots para rescate, los agentes cooperan y redistribuyen las tareas cuando alguno falla, garantizando así la continuidad de la misión y demostrando resiliencia ante fallos locales. Ante un robot que falla, son capaces de reorganizar la búsqueda y de compensar la pérdida, manteniendo la misión activa.

- **Coordinación y comunicación:** en flotas de drones de reparto, la comunicación es bidireccional y negociada; los agentes ajustan rutas en tiempo real para evitar colisiones y optimizar entregas colectivamente. La coordinación es bidireccional y si un dron detecta tráfico aéreo congestionado, negocia con otros para replanificar la ruta en tiempo real.

A continuación, y siempre utilizando los mismos criterios, se presenta un análisis comparativo de los ejemplos analizados de flujos de trabajo y protocolos A2A.

Tabla 1

Criterios y análisis comparativo de workflow y A2A

Criterio	Análisis de <i>AI workflow</i>	Análisis de protocolos A2A
Autonomía y adaptabilidad	<ul style="list-style-type: none"> - Los <i>workflows</i> mantienen autonomía restringida de acuerdo con sus reglas predefinidas. - Automatización de procesos productivos acorde con la industria 4.0. - La aparición de un documento en un formato nuevo puede interrumpir el flujo y requerir intervención humana para su adaptación. 	<ul style="list-style-type: none"> - Los protocolos A2A exhiben mayor capacidad adaptativa y autoorganización frente a escenarios imprevistos. - Muestran alta autonomía al coordinar con otros agentes y adaptarse a los cambios.
Escalabilidad y eficiencia	<ul style="list-style-type: none"> - Los <i>workflows</i> escalan en arquitecturas centralizadas y homogéneas. - La eficiencia se mantiene en entornos estables; para escalar, requiere replicar el flujo utilizando más servidores. 	<ul style="list-style-type: none"> - Los A2A son más apropiados para sistemas distribuidos y heterogéneos. - Al aumentar el número de dispositivos, la coordinación sigue siendo posible sin necesidad de un controlador central.
Flexibilidad y robustez	<ul style="list-style-type: none"> - Los <i>workflows</i> ofrecen integración modular, pero son más vulnerables a fallos críticos. - El fallo de un módulo puede interrumpir todo el flujo del proceso. 	<ul style="list-style-type: none"> - Los A2A muestran mayor robustez y flexibilidad gracias a su naturaleza descentralizada. - Si un robot falla, el sistema puede reorganizarse para compensar la pérdida y mantener la misión.
Coordinación y comunicación	<ul style="list-style-type: none"> - Los <i>workflows</i> se sustentan en coordinación lineal y control centralizado. - La coordinación bidireccional permite a los drones replanificar rutas en tiempo real ante congestión aérea. 	<ul style="list-style-type: none"> - Los A2A promueven cooperación dinámica y una comunicación distribuida. - Los drones se coordinan bidireccionalmente para replanificar rutas en tiempo real.

De acuerdo con este análisis, los *AI workflows* son buenos para procesos estandarizados, repetitivos y centralizados, mientras que los protocolos A2A son ideales para entornos distribuidos, dinámicos y con alta interacción entre entidades autónomas.

DISCUSIÓN DE HALLAZGOS

La revisión de los aportes recientes en torno a la colaboración humano-IA y el desarrollo de agentes inteligentes evidencia una convergencia hacia modelos más autónomos, adaptativos y escalables. Se propone una categorización fundamental de los flujos de trabajo y agentes de IA —desde los no agénticos hasta los plenamente autónomos— que permita comprender el grado de delegación cognitiva y operativa entre humanos y sistemas inteligentes.

De manera complementaria, se sugiere la necesidad de replantear la relación entre humanos e IA a través del marco de referencia SPAR —*sensing, planning, acting, reflecting*— que orienta los niveles de automatización esperados y subraya la importancia de la percepción contextual, el razonamiento estructurado, la acción coordinada y el aprendizaje iterativo. Utilizar la metáfora del árbol nos ofrece una visión estructural del ecosistema de la IA generativa, pues ayuda a identificar las capas de fundamentos, modelos, técnicas y capacidades que sustentan el desarrollo y ubicación de agentes inteligentes y *workflows* (raíces, tronco, ramas y hojas), además de detallar los componentes esenciales para su implementación: desde marcos agénticos y sistemas de memoria, hasta mecanismos de gobernanza y despliegue.

Los sistemas de flujos de trabajo basados en agentes de IA (*agentic workflows*) y la comunicación entre agentes A2A (*agent-to-agent*), destacan por su autonomía y adaptabilidad para enfrentar desafíos complejos. Se recomienda su implementación apoyándose en los modelos, referencias y guías analizados en este artículo, con el objetivo de diseñar novedosas propuestas de automatización, aplicables en sectores laborales y académicos en los que se proyecta que, hacia 2027, aproximadamente el 82 % de las organizaciones habrá incorporado estas prácticas, ya sea en la modalidad de flujos de trabajo de IA o con protocolos A2A.

El análisis comparativo evidencia que los *AI workflows* son mejores para procesos estandarizados, repetitivos y centralizados, mientras que los protocolos A2A son ideales para entornos distribuidos, dinámicos y con alta interacción entre entidades autónomas. Otro hallazgo es que los *AI workflows* resultan más adecuados en contextos centralizados, estables y de alta carga transaccional, en los que la predictibilidad y la eficiencia dependen de procesos secuenciales bien definidos. Sin embargo, su autonomía y resiliencia son limitadas, pues requieren mayor supervisión humana para adaptarse a cambios imprevistos y presentan vulnerabilidad frente a fallos críticos en sus componentes.

Por el contrario, los agentes de IA trabajando en conjunto y comunicándose vía protocolos A2A destacan en entornos distribuidos, dinámicos y heterogéneos, gracias a su capacidad de autoorganización, comunicación bidireccional y resiliencia ante la pérdida de nodos individuales. Su flexibilidad los hace idóneos para escenarios en los

que la cooperación, la negociación y la adaptación en tiempo real son esenciales, así como en sistemas de energía inteligente, robótica colaborativa o logística distribuida.

CONCLUSIONES

La IA aún enfrenta retos significativos en materia de transparencia, responsabilidad y equidad, lo cual puede amplificar sesgos y socavar la confianza social. Para mitigar estos riesgos, resulta esencial el cumplimiento de marcos normativos auditables, como el GDPR y la HIPAA, junto con la implementación de mecanismos de gobernanza que contemplen la detección de sesgos y una evaluación ética previa a su despliegue.

El desafío actual de la IA trasciende la mera creación de agentes y se centra en su despliegue confiable a gran escala. El éxito en su adopción dependerá de concebir sistemas que, desde el inicio, tengan capacidad de escalabilidad, integrando de forma coherente marcos de referencia, memoria, gobernanza y mecanismos efectivos de implementación.

Las arquitecturas agénticas de *AI workflows* y A2A no deben entenderse como enfoques excluyentes, sino más bien complementarios. Mientras los primeros optimizan procesos repetitivos y de gran volumen en marcos controlados, los segundos permiten gestionar la complejidad e incertidumbre en redes autónomas. La convergencia de ambos enfoques puede abrir nuevas posibilidades para el diseño de sistemas híbridos, capaces de combinar eficiencia estructurada con adaptabilidad distribuida.

La construcción de un agente de IA constituye solo la etapa inicial; el verdadero desafío se encuentra en su escalamiento hacia entornos de producción. En este proceso, la fiabilidad se erige como el elemento diferenciador entre un prototipo y un producto consolidado, ya que sin ella los sistemas de IA permanecen limitados a la experimentación, sin generar un impacto real. En ese sentido, sería deseable, por ejemplo, enfocarse en el desarrollo de sistemas robustos de memoria y gobernanza como clave para el éxito de la implementación de *workflows* y agentes de IA.

Bajo este marco, llevar a cabo el análisis de una arquitectura por capas y de los componentes esenciales evidencia la necesidad de una planificación estratégica orientada a aplicaciones concretas y flujos de trabajo. Esto resulta especialmente relevante en un campo dinámico y en constante evolución como la inteligencia artificial.

En conclusión, la integración de tecnologías de IA ha transformado profundamente a las organizaciones, orientándolas hacia niveles crecientes de autonomía y optimización operativa. Sin embargo, este avance plantea la necesidad de desarrollar arquitecturas de agentes impulsadas por IA que sean explicables, auditables y que respalden procesos de toma de decisiones confiables y transparentes. Ello exige mantener una reflexión ética permanente sobre aspectos como los sesgos algorítmicos, la transparencia y la rendición de cuentas, en consonancia con la creciente demanda global de marcos regulatorios robustos y responsables.

En síntesis, las investigaciones revisadas consolidan una visión de la inteligencia artificial como un ecosistema en continua evolución, en el cual la colaboración humano-IA progresa hacia esquemas de autonomía compartida, aprendizaje adaptativo y gobernanza ética. Este enfoque integrador proyecta una nueva etapa en la ciencia computacional, caracterizada por la convergencia entre el diseño algorítmico, la interacción humana y la sostenibilidad tecnológica, sentando las bases para el desarrollo de sistemas inteligentes más confiables, transparentes y orientados al bien común.

Finalmente, se recomienda utilizar las figuras y tablas presentadas como recursos prácticos y concisos, a fin de facilitar la planificación, programación y ejecución de flujos de trabajo y de iniciativas de colaboración basadas en protocolos de agentes de IA.

GLOSARIOS

Colaboración humano-IA

Elemento de automatización	Descripción de aspectos a especificar	Responsable principal
Paso 1. Contexto y objetivo	<ul style="list-style-type: none">- Propósito de la interacción- Dominio de aplicación- Alcance de las tareas	Humano
Paso 2. Entrada o estímulo	<ul style="list-style-type: none">- Tipo de datos (texto, voz, imágenes, sensores, etcétera)- Formato y estructura- Condiciones iniciales	Humano provee y estructura datos; la IA interpreta
Paso 3. Procesamiento interno	<ul style="list-style-type: none">- Modelo de razonamiento (reglas, ML, generativo)- Recursos de conocimiento- Capacidad de adaptación	IA
Paso 4. Acciones o salida	<ul style="list-style-type: none">- Modo de respuesta (texto, visualización, acción en sistema)- Nivel de autonomía- Retroalimentación	IA bajo supervisión humana
Paso 5. Condiciones de interacción	<ul style="list-style-type: none">- Roles y permisos- Ética y seguridad (privacidad, sesgos, normativa)- Usabilidad y explicabilidad	Humano establece roles y permisos que delimitan la acción. La IA interpreta y opera dentro de las reglas y parámetros definidos.

Metáfora del árbol de IA

Nivel (metáfora)	Componentes	Ejemplos de productos de IA
Raíces (fundación)	Fundamentos de IA/ML, programación y matemáticas aplicadas	PyTorch, TensorFlow, Keras; Python (Jupyter, VS Code); Pandas, NumPy, SciPy, Matplotlib.

(continúa)

(continuación)

Nivel (metáfora)	Componentes	Ejemplos de productos de IA
Tronco (modelos generativos)	Modelos centrales para texto, imágenes, audio y multimodalidad	GPT, Claude, LLaMA; Stable Diffusion, MidJourney, Adobe Firefly; Runway, Descript, ElevenLabs; GPT-4o, Gemini, LLaVA, Kosmos-1.
Ramas (técnicas core)	Métodos clave para optimizar y personalizar modelos	Prompt Engineering (LangChain, DSPy, FlowGPT); Fine-Tuning (LoRA, QLoRA, PEFT); RAG (Pinecone, ChromaDB, Weaviate).
Nuevas ramas (capacidades agénticas)	Sistemas orientados a agentes y orquestación de flujos	AutoGPT, CrewAI, LangGraph, Microsoft Autogen; Airflow, n8n, Zapier, Make.com.
Nuevas hojas (crecimiento avanzado)	Implementación, escalamiento y casos de uso especializados	Docker, Kubernetes, AWS Bedrock, GCP Vertex AI; Healthcare AI, FinTech AI, Creative AI, Enterprise Automation.

Evolución de la automatización con agentes de IA

Niveles de automatización	Descripción	Década	Agentes de IA
Primer nivel: agentes conversacionales iniciales	Primeros bots de IA usando language natural, con capacidades limitadas	1960	Introducción de Eliza, el primer agente conversacional
Segundo nivel: la IA generativa	Mejoró la versatilidad y accesibilidad de la IA	2011	Lanzamiento de agentes modernos: Siri y Alexa.
		2022	El auge de la IA generativa con el lanzamiento de ChatGPT.
Tercer nivel: avance multimodal	La IA se volvió multimodal, al integrar múltiples formatos de entrada y salida.	2023	Lanzamiento de Google Gemini para texto, imágenes y audio. Microsoft Copilot para desarrollo de software, sugiriendo líneas de código y ayudando a resolver problemas de programación.
Cuarto nivel: mejora del razonamiento	La IA mejoró el razonamiento: maneja tareas complejas con menos intervención humana.	2024 inicios	Avance de la IA multimodal. Lanzamiento de OpenAI Sora, modelo multimodal para generar contenido audiovisual. Claude AI para documentos extensos, análisis de datos, seguridad, ética y personalización.
Quinto nivel: mayor autonomía	La IA con más autonomía y adaptabilidad, con herramientas, memoria y acceso a internet	2024 finales	Evolución de los sistemas de IA siendo más autónomos y adaptables. Claude AI crea chatbots y asistentes virtuales que proporcionan respuestas seguras y responsables.

(continúa)

(continuación)

Niveles de automatización	Descripción	Década	Agentes de IA
Sexta generación: agentes especializados y de carácter generalista	Agentes de IA altamente especializados, demostrando alta autonomía en tareas específicas.	2025 inicios	Introducción de: <ul style="list-style-type: none">• Agentes altamente autónomos: Gemini 1.5, Deep Research y Computer Use.• Agentes de carácter generalista: Anthropic's Computer Use, Google's Project Mariner y OpenAI's Operator.

Modelos de lenguaje

Modelo	Descripción general	Fortalezas	Aplicaciones relevantes
GPT-5 / GPT-4o (OpenAI)	Modelo de propósito general, ampliamente reconocido por su versatilidad en tareas textuales y computacionales.	Equilibrio entre redacción, programación y razonamiento; capacidades multimodales en GPT-4o.	Redacción académica y creativa, generación de código, asistentes conversacionales, análisis de datos.
Claude 4 (Anthropic)	Modelo orientado a la seguridad y el razonamiento avanzado, diseñado para contextos extensos.	Manejo de largos contextos, razonamiento profundo, énfasis en alineación ética.	Investigación académica, análisis de documentos extensos, apoyo en tareas de reflexión compleja.
Gemini 2.5 (Google DeepMind)	Modelo multimodal de última generación con énfasis en integración de texto, imágenes y otras modalidades.	Conocimiento actualizado, razonamiento lógico, capacidades multimodales robustas.	Sistemas interactivos, análisis multimedia, aplicaciones educativas y científicas.
Grok (xAI)	Modelo desarrollado por xAI con un enfoque en interacción directa y expresiva.	Inmediatez informativa, estilo comunicativo más dinámico y expresivo.	Plataformas de comunicación, asistentes personales, generación de contenido en tiempo real.
Mistral 7B / Mixtral	Modelo de código abierto, optimizado para eficiencia y personalización.	Ligereza, velocidad, bajo costo computacional, flexibilidad en entornos de innovación.	Laboratorios de investigación aplicada, startups, despliegues en entornos con recursos limitados.
LLaMA 4 (Meta)	Modelo abierto y ampliamente adoptado por la comunidad investigadora y de desarrolladores.	Transparencia, soporte comunitario, independencia de APIs propietarias.	Investigación académica, desarrollo de aplicaciones personalizadas, proyectos que requieren soberanía tecnológica.

REFERENCIAS

Arslan, S. (2025). Artificial intelligence in food safety and nutrition practices: opportunities and risks. *Academia Nutrition and Dietetics*, 2(3). <https://doi.org/10.20935/AcadNutr7904>

- Automation Anywhere. (2025). *¿Qué es la automatización robótica de procesos (RPA)? Una guía empresarial*. <https://www.automationanywhere.com/la/rpa/robotic-process-automation>
- Balic, N. (2025). *Will agents replace us? Perceptions of autonomous multi-agent AI*. arXiv. <https://doi.org/10.48550/arXiv.2506.02055>
- Bassey, M., Akpan, J., Ikpe, A., David, V., & Kehinde, T. O. (2025, 26 de agosto). The convergence of additive manufacturing and artificial intelligence with LLMs for smart product design. *Academia Materials Science*, 2(3). <https://doi.org/10.20935/AcadMatSci7868>
- Binance Academy. (s. f.). *Blockchain*. <https://academy.binance.com/en/glossary/blockchain>
- Bornet, P., Wirtz, J., & Davenport, T. H., De Cremer, D., Evergreen, B. Fersht, P. Gohel, R. & Khiyara, S. (2025). *Agentic artificial intelligence. Harnessing AI agents to reinvent business, work and life*. World Scientific Publishing Company.
- Casini, L., Marchetti, N., Montanucci, A., Orrù, V., & Rocchetti, M. (2023). A human–AI collaboration workflow for archaeological sites detection. *Scientific Reports*, 13(1), 8699. <https://doi.org/10.1038/s41598-023-36015-5>
- Corvalán, J. G., & Sánchez Caparrós, M. (2025). *Agentes de inteligencia artificial y workflows agénticos: la nueva frontera de la automatización*. Laboratorio de Inteligencia Artificial de la Facultad de Derecho UBA. <https://www.web.ialab.com.ar/webia/wp-content/uploads/2025/libros/Agentes-de-inteligencia-artificial-y-workflows-agenticos/Agentes-de-inteligencia-artificial-y-workflows-agenticos.pdf>
- Coshow, T. (2024, 1 de octubre). Los agentes inteligentes de IA pueden operar de forma autónoma. *Gartner*. <https://www.gartner.es/es/articulos/agente-inteligente-de-ia>
- Empresa Actual. (2025, 11 de febrero). ¿Qué son los agentes de IA o la inteligencia artificial agéntica (IAA)? *EmpresaActual.com* <https://www.empresaactual.com/que-son-los-agentes-de-ia-o-la-inteligencia-artificial-agentica-iaa/>
- Hauser, E. (2019, 9-13 de noviembre). *AI systems as ethical agents* [Presentación dentro del taller Good Systems: Ethical AI]. 22nd ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW), Austin, Texas, Estados Unidos.
- Li, Y. H., Chang, Z. H., Shan Ou, Y., Chen, Y., Yu, L. Y., Tan, J. Q., Liu, Y., & Chang, Y. Z. (2025). Exploration of the development and technical features of intelligent inspection robots. *International Journal of Advanced Engineering Research and Science*, 12(10), 61-70. <https://doi.org/10.22161/ijaers.1210.9>

- Lieben, M. (2025). *3 types of AI workflows* [flujograma]. LinkedIn. https://www.linkedin.com/posts/the-best-of-ai_ai-agents-are-replacing-human-employees-activity-7364786657318838273-SH--/
- Ramalingam, B. C. (2025). Efficient implementation of AI agents in enterprise application integration (EAI) and electronic data interchange (EDI). *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 11(2), 150-170. <https://doi.org/10.32628/cseit251112397>
- 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), 1-15. <https://doi.org/10.5121/ijccsa.2025.15101>
- Shaw, F. X. (2025, 19 de mayo). Microsoft Build 2025: The age of AI agents and building the open agentic web. *Official Microsoft Blog*. <https://blogs.microsoft.com/blog/2025/05/19/microsoft-build-2025-the-age-of-ai-agents-and-building-the-open-agentic-web/>
- Shoham, Y., & Leyton-Brown, K. (2008). Multiagent systems. Algorithmic, game-theoretic, and logical foundations. *Cambridge University Press*.
- The Best of AI. (2025). *AI agents are replacing human employees. Build yours* [Publicación]. LinkedIn. https://www.linkedin.com/posts/the-best-of-ai_ai-agents-are-replacing-human-employees-activity-7364786657318838273-SH--/
- Viswanathan, G., Samdani, G., & Dixit, Y. (2025). AI Agents. *International Journal of Advanced Information Technology*, 15(1/2), 9-17. <https://doi.org/10.5121/ijait.2025.15202>
- Xu, Z., & King, I. (2025, 2 de septiembre). Introducing *Academia AI and Applications*: A new platform for responsible and interdisciplinary AI research. *Academia AI and Applications*. <https://doi.org/10.20935/AcadAI7885>