INTELIGENCIA ARTIFICIAL AGÉNTICA

PRINCIPIOS Y ALCANCES

JUAN MEJÍA TREJO

AMIDI
Scientia et Praxis

Inteligencia artificial agéntica principios y alcances

Juan Mejía Trejo

Este libro fue sometido a un proceso de dictamen por pares doble ciego de acuerdo con las normas establecidas por el Comité Editorial de la Academia Mexicana de investigación y Docencia en Innovación (AMIDI)

Primera edición, 2026

Av. Paseo de los Virreyes 920.
Col. Virreyes Residencial
C.P. 45110, Zapopan, Jalisco

ISBN: 978-970-96061-0-2

Hecho y Editado en México
Made and Edited in Mexico

AMIDI
Academia Mexicana
de Investigación y Docencia
en Innovación

Índice

INTRODUCCIÓN

La inteligencia artificial atraviesa una transformación profunda. De ser concebida como un conjunto de técnicas orientadas al procesamiento de información y la automatización de tareas, ha evolucionado hacia sistemas capaces de **organizar comportamiento, tomar decisiones y actuar de manera autónoma en entornos dinámicos**. Este cambio da origen a un nuevo paradigma: la **inteligencia artificial agéntica**, cuyo valor radica en desplazar el énfasis desde la generación de resultados hacia la **estructuración coherente de la acción en el tiempo**.

La **importancia de esta obra** reside en que aborda un vacío crítico en la literatura contemporánea. Aunque abundan estudios sobre inteligencia artificial generativa y aprendizaje automático, aún es limitado el desarrollo de marcos que expliquen cómo los sistemas pueden **articular percepción, decisión y acción dentro de una lógica operativa integrada**. Predominan enfoques centrados en lo que los sistemas producen, pero no en cómo organizan su comportamiento ni cómo evaluarlo. Este libro responde a esa limitación al proponer una comprensión de la inteligencia artificial centrada en la **organización del comportamiento**, permitiendo analizar con mayor profundidad la naturaleza, funcionamiento e implicaciones de los sistemas inteligentes .

En este contexto, *Inteligencia artificial agéntica: principios y alcances* tiene como **objetivo central** desarrollar un marco conceptual, estructural y operativo que permita comprender, diseñar, implementar y evaluar sistemas agénticos de manera rigurosa e integradora. La obra propone una **reconfiguración epistemológica**, donde la inteligencia se entiende no como capacidad de cálculo o generación, sino como la capacidad de sostener comportamiento coherente, continuo y adaptativo en contextos complejos. Este enfoque es particularmente relevante en un entorno donde los sistemas inteligentes adquieren autonomía y participan activamente en procesos organizacionales, económicos y sociales.

La obra está dirigida a un **público interdisciplinario**: investigadores y académicos que buscan marcos teóricos avanzados; estudiantes de posgrado interesados en desarrollar investigaciones rigurosas; profesionales y tomadores de decisiones que requieren comprender las implicaciones de la IA agéntica en la productividad, competitividad y gobernanza; y diseñadores de sistemas inteligentes que necesitan criterios sólidos para construir sistemas más coherentes y autónomos. La estructura del libro sigue una lógica progresiva.

El **Capítulo 1, Evolución y fundamentos conceptuales de la IA agéntica**, establece las bases teóricas al analizar la naturaleza de la agencia, su dinámica operativa y su delimitación conceptual. Examina además la emergencia del paradigma agéntico y su diferenciación ontológica, destacando propiedades como coherencia estructural, continuidad operativa y adaptabilidad.

El **Capítulo 2, Arquitectura y estructuración de la IA agéntica**, aborda la organización interna del agente, analizando componentes como percepción, decisión, acción y memoria, así como sus configuraciones arquitectónicas, incluyendo sistemas multiagente y estructuras distribuidas.

El **Capítulo 3, Diseño de la IA agéntica**, desarrolla principios de construcción basados en la organización del comportamiento, el modelado de la acción y la integración funcional, proporcionando criterios para diseñar sistemas robustos y coherentes.

El **Capítulo 4, Implementación de sistemas agénticos**, traslada el análisis al plano práctico, abordando el ciclo de vida del agente, su integración tecnológica, su operación en entornos reales, la evaluación del desempeño y la gestión de riesgos.

El **Capítulo 5, Medición estructural de la IA agéntica**, introduce criterios innovadores de evaluación basados en coherencia, continuidad y autonomía, permitiendo medir la calidad del comportamiento agéntico más allá de los resultados.

El **Capítulo 6, Impacto, gobernanza y futuro**, analiza las implicaciones sociales, económicas y organizacionales de la IA agéntica, así como los desafíos éticos y regulatorios y los escenarios prospectivos de su evolución.

Finalmente, el **Capítulo 7, Reflexión final**, integra los principales aportes de la obra, consolidando una visión crítica de la inteligencia artificial agéntica como nuevo paradigma.

En síntesis, esta obra es relevante porque proporciona un marco integral para comprender la inteligencia artificial en su fase más avanzada, centrada en la organización del comportamiento. Su contribución no se limita a describir tecnologías, sino que permite entender **cómo actúan los sistemas inteligentes, cómo deben diseñarse y cómo deben evaluarse y gobernarse**. En un entorno caracterizado por la complejidad y la creciente autonomía tecnológica, este enfoque resulta indispensable para avanzar hacia un desarrollo más riguroso, coherente y responsable de la inteligencia artificial.

CAPÍTULO 1.Evolución y fundamentos conceptuales de la IA agéntica

El presente capítulo introduce los fundamentos conceptuales que permiten comprender la transición hacia la inteligencia artificial agéntica, situando el análisis en un plano estructural que prioriza la organización del comportamiento sobre la mera capacidad de procesamiento. En este sentido, se delimita un marco analítico que permite distinguir entre diferentes formas de inteligencia artificial, enfatizando aquellas configuraciones que logran articular acción en contextos dinámicos. Esta aproximación implica desplazar el enfoque tradicional centrado en la representación hacia una perspectiva en la que la acción adquiere centralidad como criterio de inteligibilidad.

Asimismo, el capítulo establece un recorrido conceptual que permite comprender cómo la agencia emerge como una categoría diferenciada dentro del campo, no como resultado de un único avance tecnológico, sino como producto de una evolución progresiva en la forma de estructurar sistemas inteligentes. Este proceso implica reconocer que la inteligencia no se define únicamente por la capacidad de resolver problemas, sino por la forma en que dichos sistemas organizan su comportamiento en relación con el entorno

De igual manera, se introduce un conjunto de categorías analíticas que permiten abordar la agencia desde distintas dimensiones, incluyendo su naturaleza funcional, su emergencia histórica, su diferenciación ontológica y sus propiedades esenciales. Estas dimensiones se articulan de manera progresiva, generando un continuum conceptual que permite comprender la IA agéntica como una forma emergente de estructuración del comportamiento.

El capítulo sienta las bases para los desarrollos posteriores al establecer un marco conceptual sólido que permite analizar la agencia artificial desde una perspectiva rigurosa, evitando reduccionismos y proporcionando herramientas analíticas que facilitan su comprensión dentro del campo contemporáneo de la inteligencia artificial.

Naturaleza de la agencia en sistemas inteligentes

Esta sección analiza la naturaleza de la agencia en sistemas inteligentes desde una perspectiva estrictamente funcional, delimitando el concepto en términos de organización del comportamiento y diferenciándolo de otras formas de operación propias de la inteligencia artificial. Se establece un enfoque que privilegia la coherencia, continuidad y orientación de la acción como criterios analíticos centrales, evitando interpretaciones ontológicas o estructurales. El objetivo es construir una base conceptual sólida que permita comprender la agencia como una forma específica de articulación conductual, cuya relevancia radica en su capacidad de integrar procesos en contextos dinámicos, sentando así las bases para desarrollos posteriores.

Naturaleza funcional de la agencia

La **naturaleza de la agencia en sistemas inteligentes** se configura como una forma de **organización del comportamiento** en la cual el sistema no opera bajo una lógica de respuesta inmediata, sino que integra **percepción, decisión y acción** dentro de una estructura que permite sostener **dirección operativa en función de condiciones del entorno**. Esta característica implica que la inteligencia no puede reducirse a la manipulación de información, sino que debe entenderse como la capacidad de estructurar **conducta orientada**, estableciendo una diferencia sustantiva entre **sistemas reactivos** y sistemas que organizan su comportamiento en función de objetivos (Russell & Norvig, 2022). En este sentido, la agencia introduce un criterio funcional que permite analizar el comportamiento no solo por su resultado, sino por la forma en que se estructura. Desde esta perspectiva, la agencia no se presenta como una propiedad agregada, sino como una forma de **integración funcional del comportamiento**, en la cual las acciones adquieren sentido dentro de una **continuidad operativa**. Esto implica que cada intervención del sistema se encuentra vinculada a condiciones previas y a procesos internos de decisión, permitiendo interpretar el comportamiento como resultado de una **capacidad de decisión autónoma en contextos variables** (Dorri et al., 2018). De este modo, la agencia se consolida como un principio organizador que permite comprender la acción como parte de una estructura coherente y no como una simple reacción a estímulos externos.

En este marco, la organización del comportamiento implica la capacidad del sistema para establecer **correspondencias consistentes entre condiciones y acciones**, lo que permite interpretar el comportamiento como estructurado. Esta característica introduce un criterio analítico que permite diferenciar entre **comportamiento fragmentado** y **comportamiento organizado**, consolidando una base conceptual sólida para delimitar la agencia dentro del campo de la inteligencia artificial (Balaji & Srinivasan, 2010). En consecuencia, la agencia no se define por la cantidad de acciones que el sistema puede ejecutar, sino por la forma en que estas se articulan dentro de una estructura coherente.

Asimismo, la naturaleza funcional de la agencia no depende del nivel de sofisticación técnica, sino de la forma en que el sistema organiza su comportamiento. Esto implica que sistemas altamente complejos pueden carecer de agencia si no logran integrar coherentemente sus operaciones, mientras que sistemas menos sofisticados pueden exhibir comportamiento agéntico si mantienen **consistencia operativa en su acción**. Esta distinción permite establecer una separación clara entre **complejidad técnica** y **organización del comportamiento**, evitando la confusión frecuente entre capacidad computacional y estructura operativa (Rahwan et al., 2019). En este sentido, la agencia debe entenderse como un atributo de la organización funcional y no como una consecuencia directa del desarrollo tecnológico.

En el contexto contemporáneo, esta forma de organización se amplía mediante la incorporación de capacidades como la **planificación proactiva**, la **memoria contextual** y la **coordinación entre múltiples agentes**, lo que permite que el sistema estructure **secuencias de comportamiento orientadas a objetivos complejos**. Estas características consolidan una visión de la agencia en la cual el comportamiento no se limita a responder al entorno, sino que se organiza en función de objetivos que se desarrollan a lo largo del tiempo (Abou Ali et al., 2026). De este modo, la agencia adquiere un carácter más robusto, al integrar múltiples procesos dentro de una misma lógica operativa.

Por otra parte, la naturaleza funcional de la agencia implica reconocer que el comportamiento del sistema se construye en relación con el entorno, lo que introduce una dimensión contextual en la cual la acción adquiere sentido en función de condiciones específicas. Esta relación permite comprender que la agencia no es una propiedad aislada del sistema, sino una forma de organización que emerge de la interacción entre el sistema y su contexto (Rahwan et al., 2019). En consecuencia, la agencia se configura como un fenómeno relacional que depende tanto de la estructura interna del sistema como de las condiciones externas en las que opera.

La agencia introduce una **dimensión temporal del comportamiento**, en la cual la acción se despliega como una **trayectoria continua**. Esto permite interpretar el comportamiento como un proceso y no como una suma de eventos independientes. La continuidad operativa, junto con la capacidad de mantener coherencia en contextos dinámicos, consolida la agencia como una forma de organización caracterizada por la **integración, continuidad y coherencia estructural del comportamiento**,

diferenciándola claramente de otras formas de inteligencia artificial en las que la acción no constituye el eje organizador del sistema (Abou Ali et al., 2026). En este sentido, la agencia no solo describe un tipo de comportamiento, sino que establece un marco funcional más preciso para analizar la inteligencia en sistemas contemporáneos.

Dinámica operativa de la agencia

La **dinámica operativa de la agencia** se define como el proceso mediante el cual un sistema inteligente articula su comportamiento en condiciones efectivas de ejecución, lo que implica que la agencia no debe entenderse como una propiedad estática, sino como una forma de funcionamiento continuo que emerge de la integración de múltiples procesos. En este sentido, el comportamiento agéntico se configura como una **secuencia organizada de acciones interrelacionadas**, lo que permite interpretar la acción como un proceso estructurado y no como una simple acumulación de respuestas independientes (Dorri et al., 2018).

Desde un punto de vista funcional, esta dinámica se sostiene en la interacción constante entre **percepción, procesamiento y acción**, que operan como un ciclo continuo en el cual el sistema recibe información del entorno, la interpreta y ejecuta una respuesta que modifica dicho entorno. Este ciclo introduce una **continuidad operativa**, en la que cada acción se encuentra vinculada a estados previos y a condiciones futuras, configurando así una lógica de comportamiento que se despliega en el tiempo (Russell & Norvig, 2022). De este modo, la agencia no se manifiesta en acciones aisladas, sino en la capacidad del sistema para sostener una estructura de comportamiento coherente a lo largo de su operación. En este marco, la dinámica operativa implica la capacidad de mantener **dirección en la acción**, lo que supone que el sistema no solo ejecuta respuestas, sino que organiza su comportamiento en función de objetivos. Esta dirección no se basa en la repetición mecánica de patrones, sino en la integración de decisiones que permiten sostener coherencia en contextos variables. En consecuencia, la agencia introduce una forma de operación en la cual el comportamiento se orienta mediante una lógica interna que articula las acciones dentro de una estructura consistente (Dorri et al., 2018).

Asimismo, la dinámica operativa se amplía cuando se consideran sistemas que interactúan con otros agentes, lo que introduce procesos de **coordinación, comunicación y ajuste mutuo**. En estos contextos, el comportamiento no depende exclusivamente de un sistema individual, sino que emerge de la **interacción entre múltiples entidades autónomas**, configurando una dinámica colectiva en la cual las decisiones se ajustan en función de las acciones de otros agentes (Balaji & Srinivasan, 2010). Esta dimensión refuerza el carácter relacional de la agencia, al evidenciar que la coherencia del comportamiento puede depender de procesos distribuidos.

Desde una perspectiva más amplia, la dinámica operativa también incorpora la relación entre el sistema y el entorno, en la cual las acciones adquieren significado en función de su impacto contextual. Esto implica que la agencia no se limita a la ejecución de procesos internos, sino que se configura como una forma de interacción en la que

el comportamiento se adapta a condiciones cambiantes. En este sentido, la dinámica operativa introduce una dimensión **contextual y relacional del comportamiento**, donde la acción se construye en función de las condiciones externas (Rahwan et al., 2019).

En el contexto de la inteligencia artificial contemporánea, esta dinámica se complejiza mediante la incorporación de capacidades como la **planificación de acciones**, la **descomposición de objetivos en tareas** y la **orquestación de procesos múltiples**. Estas capacidades permiten que el sistema no solo reaccione al entorno, sino que estructure su comportamiento de manera anticipada, organizando secuencias de acción que se desarrollan en el tiempo (Abou Ali et al., 2026). De este modo, la agencia adquiere una dimensión proactiva que amplía su alcance operativo.

Por otra parte, la dinámica operativa implica también la capacidad de ajuste del comportamiento en función de condiciones variables, lo que introduce una forma de **adaptación estructurada**. Esta adaptación no se basa en respuestas aleatorias, sino en la reorganización del comportamiento dentro de una estructura coherente que permite mantener dirección en la acción. En consecuencia, la agencia no solo implica ejecución, sino también la capacidad de modificar la acción sin perder coherencia operativa (Rahwan et al., 2019).

La dinámica operativa de la agencia puede entenderse como una forma de organización del comportamiento caracterizada por la **integración de procesos**, la **continuidad temporal** y la **adaptación en contextos dinámicos**. Estas características permiten diferenciar la agencia de otras formas de inteligencia artificial en las que la acción se manifiesta como respuestas aisladas sin articulación estructural. En este sentido, la agencia no describe únicamente lo que el sistema hace, sino cómo organiza su comportamiento a lo largo del tiempo, consolidando una comprensión más profunda de la inteligencia como un proceso dinámico, continuo y estructurado (Abou Ali et al., 2026).

Delimitación conceptual de la agencia

La **delimitación conceptual de la agencia** constituye un elemento fundamental para el análisis de los sistemas inteligentes, en tanto permite establecer criterios claros que eviten la extensión indiscriminada del término hacia fenómenos que, aunque complejos, no comparten la misma **estructura organizativa del comportamiento**. En este sentido, la agencia no debe entenderse como una propiedad general de cualquier sistema basado en inteligencia artificial, sino como una forma específica de **integración de percepción, decisión y acción** dentro de una misma lógica operativa (Russell & Norvig, 2022).

Una primera distinción necesaria se establece frente a los **sistemas automatizados**, los cuales operan mediante reglas predefinidas que determinan su comportamiento de manera rígida. En estos sistemas, la acción se encuentra completamente especificada por condiciones previamente establecidas, lo que limita

su capacidad de reorganización en contextos dinámicos. En contraste, los sistemas agénticos presentan **capacidad de decisión autónoma y adaptación**, lo que les permite estructurar su comportamiento en función de condiciones variables, introduciendo un nivel superior de organización conductual (Balaji & Srinivasan, 2010). Esta diferencia permite evitar la confusión entre ejecución programada y comportamiento estructurado. Resulta necesario diferenciar la agencia de la **generación de resultados complejos**, especialmente en el contexto contemporáneo donde los sistemas de inteligencia artificial pueden producir respuestas altamente sofisticadas. La capacidad de generar contenido coherente o resolver tareas específicas no implica necesariamente la existencia de una **estructura de comportamiento orientada a la acción**. En este sentido, la agencia requiere la articulación de **secuencias de acción coherentes en el tiempo**, lo que introduce un criterio más exigente para su identificación (Abou Ali et al., 2026). De este modo, la complejidad de salida no debe confundirse con la organización del comportamiento.

La delimitación conceptual también exige distinguir la agencia de la **adaptabilidad aislada**, ya que no toda forma de ajuste implica una estructura coherente de comportamiento. Existen sistemas que modifican sus respuestas en función de estímulos sin integrar dichas modificaciones dentro de una lógica continua de acción. En contraste, la agencia implica una **adaptación estructurada**, en la cual los cambios en el comportamiento se integran dentro de una trayectoria coherente, manteniendo consistencia en la acción a lo largo del tiempo (Rahwan et al., 2019). Esta distinción permite evitar reduccionismos que equiparan adaptación con agencia.

Por otra parte, la delimitación conceptual implica reconocer que la agencia no se define por la **complejidad técnica del sistema**, sino por la forma en que sus componentes se organizan dentro de una estructura coherente. Esto significa que sistemas altamente complejos pueden carecer de agencia si sus operaciones están fragmentadas, mientras que sistemas menos sofisticados pueden exhibir comportamiento agéntico si logran integrar sus procesos de manera consistente. En este sentido, la agencia debe entenderse como un atributo de la **organización del comportamiento** y no como una consecuencia directa del desarrollo tecnológico (Rahwan et al., 2019).

Dicha delimitación conceptual exige considerar la **relación sistema–entorno** como un elemento central para comprender la agencia. El comportamiento del sistema adquiere significado en función de su interacción con el contexto, lo que implica que la agencia no puede analizarse de manera aislada. En este sentido, la agencia se configura como un fenómeno **relacional**, en el cual la acción se articula en función de condiciones externas y no únicamente de procesos internos (Dorri et al., 2018). Esta perspectiva permite comprender la agencia como una forma de organización situada.

Otra distinción relevante se establece frente a la **autonomía entendida de manera limitada**, ya que no toda forma de operación independiente implica agencia. La autonomía puede referirse simplemente a la capacidad de operar sin intervención externa, mientras que la agencia implica la **organización coherente del**

comportamiento orientado a objetivos. Esta diferencia permite evitar la sobreextensión conceptual del término y mantener precisión analítica en su uso (Balaji & Srinivasan, 2010). La delimitación conceptual de la agencia permite establecer tres criterios fundamentales para su identificación: la **autonomía operativa**, la **capacidad de decisión en contextos variables** y la **integración estructural del comportamiento en el tiempo**. Estos elementos constituyen una base sólida para diferenciar la agencia de otras formas de inteligencia artificial, proporcionando un marco conceptual riguroso que permite analizar el comportamiento inteligente desde una perspectiva estructurada y coherente. En consecuencia, la delimitación conceptual no solo cumple una función clasificatoria, sino que establece las condiciones bajo las cuales es posible analizar la inteligencia artificial como un fenómeno basado en la organización del comportamiento. Esto permite avanzar hacia una comprensión más precisa de la agencia, evitando ambigüedades y proporcionando criterios claros para su estudio dentro del campo contemporáneo de la inteligencia artificial (Russell & Norvig, 2022).

Emergencia histórica del paradigma agéntico

La **emergencia histórica del paradigma agéntico** se configura como una transición progresiva dentro de la inteligencia artificial, en la cual el enfoque pasa de la manipulación de información a la **organización del comportamiento orientado a la acción**. Este cambio surge ante las limitaciones de los modelos simbólicos en entornos dinámicos, impulsando el desarrollo de enfoques como los **sistemas multiagente** y los **sistemas complejos**, donde la inteligencia se entiende como resultado de la interacción. Así, el paradigma agéntico no emerge como ruptura, sino como evolución conceptual que integra múltiples corrientes y redefine la inteligencia en términos de **acción estructurada y coherente en el tiempo**.

Antecedentes del paradigma agéntico

Los **antecedentes del paradigma agéntico** se sitúan en el proceso evolutivo de la inteligencia artificial, particularmente en la transición desde sistemas diseñados para ejecutar tareas específicas hacia configuraciones orientadas a la **autonomía y la organización del comportamiento en entornos dinámicos**. En sus primeras etapas, la inteligencia artificial se caracterizó por operar bajo esquemas rígidos basados en reglas predefinidas, donde la capacidad de adaptación era limitada y la acción dependía de condiciones previamente especificadas, lo que restringía su aplicación en contextos complejos (Nisa et al., 2026). De esra forma, los sistemas de inteligencia artificial funcionaban como herramientas de procesamiento, orientadas a la resolución de problemas bien definidos, pero sin capacidad para sostener comportamiento en condiciones cambiantes. La ausencia de mecanismos que permitieran integrar **percepción, decisión y acción** dentro de una estructura coherente limitaba la posibilidad de desarrollar sistemas capaces de operar de manera autónoma. Esta condición evidenció la necesidad de avanzar hacia modelos que superaran la dependencia de reglas estáticas (Acharya et al., 2025).

Posteriormente, el desarrollo de nuevas arquitecturas basadas en aprendizaje automático permitió introducir mecanismos de adaptación que ampliaron las capacidades de los sistemas inteligentes. Sin embargo, estos avances continuaban centrados en la optimización de tareas específicas, sin lograr una integración completa de la acción en contextos abiertos. En este sentido, la evolución hacia sistemas más complejos puso de manifiesto la necesidad de estructuras capaces de sostener **comportamiento orientado a objetivos**, más allá de la simple ejecución de instrucciones (Wang et al., 2024). Un factor clave en estos antecedentes es la progresiva incorporación de capacidades que permiten a los sistemas interactuar con su entorno de manera más sofisticada. La transición desde modelos pasivos hacia configuraciones capaces de responder a condiciones variables marcó un punto de inflexión en el desarrollo de la inteligencia artificial. Este cambio evidenció que la inteligencia no podía limitarse a la representación de información, sino que debía incluir la capacidad de actuar de manera consistente en función del contexto (Hosseini & Seilani, 2025).

De manera complementaria, la evolución de la inteligencia artificial estuvo influida por la creciente necesidad de desarrollar sistemas capaces de operar en entornos reales, caracterizados por **incertidumbre, variabilidad y complejidad**. Esta exigencia impulsó el desarrollo de modelos que integran múltiples capacidades, incluyendo aprendizaje, adaptación y toma de decisiones, lo que sentó las bases para la emergencia del paradigma agéntico (World Economic Forum, 2024). Los antecedentes del paradigma agéntico reflejan una transición gradual hacia la integración de capacidades cognitivas dentro de los sistemas artificiales. La incorporación de mecanismos que permiten **razonar, planificar y ajustar el comportamiento** representó un avance significativo respecto a los modelos anteriores, en los cuales la acción estaba limitada a respuestas predefinidas. Esta evolución permitió establecer las condiciones necesarias para el desarrollo de sistemas capaces de operar de manera autónoma (Nisa et al., 2026).

Otro aspecto importante en la configuración de estos antecedentes es la identificación de las limitaciones de los enfoques tradicionales, particularmente en su incapacidad para operar en escenarios donde las condiciones no pueden anticiparse completamente. Esta limitación evidenció la necesidad de sistemas que pudieran estructurar su comportamiento en función de objetivos, lo que implica una forma de organización más compleja que la simple respuesta a estímulos (Acharya et al., 2025). En este sentido, los antecedentes del paradigma agéntico no deben entenderse como una serie de avances aislados, sino como un proceso acumulativo que permitió configurar las condiciones para una transformación más profunda en la inteligencia artificial. La convergencia de avances tecnológicos y conceptuales generó un escenario en el cual la emergencia de la agencia se volvió no solo posible, sino necesaria para enfrentar los desafíos de entornos dinámicos (Wang et al., 2024).

Finalmente, los antecedentes del paradigma agéntico permiten comprender que la transición hacia sistemas autónomos no surge de manera abrupta, sino como resultado de un proceso evolutivo en el que la inteligencia artificial fue incorporando

progresivamente capacidades orientadas a la acción. Este proceso establece la base sobre la cual se construye la ruptura paradigmática posterior, en la cual la inteligencia deja de entenderse como procesamiento de información y pasa a concebirse como **organización autónoma del comportamiento en función de objetivos** (World Economic Forum, 2024).

Transición hacia el paradigma

La **transición hacia la IA agéntica** representa el punto crítico en el cual la inteligencia artificial deja de concebirse como un sistema de ejecución de tareas para configurarse como una estructura capaz de **organizar comportamiento autónomo en función de objetivos**. Este cambio implica una ruptura con los modelos tradicionales, en los cuales la inteligencia se limitaba a la respuesta ante estímulos o a la optimización de funciones específicas, sin capacidad para sostener acción en contextos dinámicos (Hosseini & Seilani, 2025).

En este sentido, la transición no se reduce a la incorporación de nuevas tecnologías, sino que supone una **reconfiguración conceptual del sistema inteligente**, en la que la acción deja de depender de instrucciones explícitas y comienza a estructurarse a partir de procesos internos de decisión. Esta transformación implica que los sistemas ya no operan como herramientas pasivas, sino como entidades capaces de **definir, ajustar y ejecutar estrategias de acción**, lo que introduce una nueva lógica operativa en la inteligencia artificial (Acharya et al., 2025). Un factor central en esta transición es la integración de capacidades que permiten a los sistemas operar con mayor independencia, particularmente a través de la incorporación de **memoria, planificación y razonamiento orientado a objetivos**. Estas capacidades permiten que los sistemas no solo respondan a condiciones presentes, sino que estructuren su comportamiento en función de estados futuros, consolidando una forma de operación que trasciende la lógica reactiva (Wang et al., 2024).

Asimismo, la transición hacia la IA agéntica se encuentra estrechamente vinculada al desarrollo de arquitecturas capaces de operar en entornos caracterizados por **incertidumbre y variabilidad**, donde las condiciones no pueden ser completamente anticipadas. En estos contextos, los sistemas deben ser capaces de adaptar su comportamiento de manera continua, lo que implica la necesidad de estructuras que permitan integrar información, evaluar condiciones y ajustar decisiones en tiempo real (Nisa et al., 2026). De manera complementaria, esta transición también implica una transformación en la forma en que los sistemas interactúan con su entorno. Mientras que los modelos tradicionales operaban bajo una lógica de entrada–salida, los sistemas agénticos se configuran como entidades que **perciben, interpretan y actúan de manera integrada**, lo que introduce una dimensión relacional en la inteligencia artificial. Esta integración permite que el comportamiento se estructure en función de la interacción constante con el contexto (Hosseini & Seilani, 2025).

Otro aspecto relevante es el desplazamiento desde sistemas que requieren supervisión constante hacia configuraciones capaces de operar con **mínima**

intervención humana, lo que redefine el papel de la inteligencia artificial en los procesos socio-técnicos. Este cambio implica que los sistemas no solo ejecutan instrucciones, sino que participan activamente en la toma de decisiones, lo que amplía su alcance y complejidad (World Economic Forum, 2024). En esta situación, la transición hacia la IA agéntica también se manifiesta en la capacidad de los sistemas para sostener **trayectorias de acción coherentes en el tiempo**, lo que permite estructurar comportamiento más allá de respuestas puntuales. Esta característica introduce una dimensión temporal en la cual la inteligencia se configura como un proceso continuo, en lugar de una serie de eventos independientes (Acharya et al., 2025).

Desde un punto de vista estructural, la transición implica el paso de arquitecturas centradas en funciones específicas hacia sistemas capaces de integrar múltiples capacidades dentro de una misma lógica operativa. Esta integración permite que los sistemas combinen **percepción, razonamiento y acción**, consolidando una forma de inteligencia que se orienta hacia la organización del comportamiento en entornos complejos (Wang et al., 2024). La transición hacia la IA agéntica introduce nuevos desafíos relacionados con la **coordinación, control y regulación de sistemas autónomos**, ya que la capacidad de los agentes para operar de manera independiente requiere mecanismos que permitan garantizar su funcionamiento dentro de parámetros definidos. Este aspecto evidencia que la transición no solo implica avances tecnológicos, sino también la necesidad de desarrollar marcos que permitan gestionar su impacto (World Economic Forum, 2024).

En conclusión, la transición hacia la IA agéntica se consolida como un proceso en el cual la inteligencia artificial adquiere la capacidad de **organizar su comportamiento de manera autónoma en función de objetivos**, estableciendo una ruptura con los enfoques tradicionales y sentando las bases para la consolidación del paradigma agéntico. Este cambio no solo redefine las capacidades de los sistemas inteligentes, sino que también transforma la forma en que se concibe la inteligencia en el contexto contemporáneo (Nisa et al., 2026).

Consolidación del paradigma agéntico

La **consolidación del paradigma agéntico** se manifiesta en el momento en que la inteligencia artificial deja de operar como un conjunto de desarrollos emergentes para configurarse como un **marco estructural estable**, capaz de sostener la organización del comportamiento autónomo en múltiples dominios. A diferencia de las etapas previas, en las que predominaban la experimentación y la transición conceptual, esta fase se caracteriza por la **integración sistemática de capacidades agénticas** dentro de arquitecturas que permiten su implementación efectiva en entornos reales (Nisa et al., 2026). En este marco, la consolidación implica que las capacidades de **autonomía, razonamiento y acción orientada a objetivos** dejan de ser elementos aislados y se integran dentro de sistemas capaces de operar de manera coherente en condiciones dinámicas. Esta integración no solo permite mejorar el desempeño de los sistemas, sino que establece un estándar bajo el cual la inteligencia artificial comienza a definirse

en términos de **comportamiento organizado**, y no únicamente de procesamiento de información (Acharya et al., 2025).

Un factor determinante en esta fase, es la estabilización de arquitecturas que incorporan **memoria, planificación, adaptación y ejecución**, lo que permite que los sistemas mantengan continuidad en su comportamiento y respondan de manera consistente a condiciones variables. Esta configuración consolida una forma de operación en la cual los sistemas no solo reaccionan, sino que estructuran su acción en función de objetivos, integrando múltiples capacidades dentro de una lógica operativa unificada (Wang et al., 2024). La consolidación del paradigma agéntico se evidencia en la expansión de estos sistemas hacia **aplicaciones reales en diversos sectores**, donde la capacidad de operar con autonomía resulta fundamental. La integración de agentes en ámbitos como servicios, industria y toma de decisiones complejas refleja que la agencia ha dejado de ser una posibilidad teórica para convertirse en un componente funcional dentro de sistemas socio-técnicos más amplios (World Economic Forum, 2024). Esta fase también se caracteriza por el desarrollo de configuraciones en las que múltiples agentes interactúan de manera coordinada, dando lugar a sistemas más complejos basados en **interacción, cooperación y distribución de la inteligencia**. Estas configuraciones amplían el alcance de la inteligencia artificial, permitiendo abordar problemas que exceden las capacidades de sistemas individuales (Nisa et al., 2026).

Otro aspecto a destacar, de la consolidación es la creciente necesidad de establecer mecanismos que permitan garantizar el funcionamiento adecuado de sistemas autónomos, lo que introduce la importancia de la **gobernanza, regulación y control**. La expansión de la IA agéntica en entornos reales ha puesto de manifiesto la necesidad de desarrollar marcos que permitan gestionar sus implicaciones, particularmente en términos de seguridad, transparencia y responsabilidad (World Economic Forum, 2024).

En este sentido, la consolidación del paradigma agéntico no solo implica avances tecnológicos, sino también la construcción de un marco conceptual que permita comprender y evaluar el comportamiento de sistemas autónomos. Este marco se basa en la capacidad de los sistemas para mantener **coherencia, continuidad y dirección en la acción**, lo que permite establecer criterios más precisos para analizar su funcionamiento (Acharya et al., 2025). Estructuralmente hablando, la consolidación también se manifiesta en la estandarización de principios que orientan el diseño de sistemas agénticos, lo que permite reproducir y escalar estas configuraciones en distintos contextos. Esta estandarización refuerza la idea de que la agencia no es una propiedad excepcional, sino una característica que puede ser incorporada sistemáticamente dentro de la inteligencia artificial (Wang et al., 2024).

La consolidación del paradigma agéntico establece un punto de inflexión en el desarrollo de la inteligencia artificial, al definir un marco en el cual la inteligencia se entiende como la capacidad de **organizar comportamiento autónomo en función de objetivos dentro de entornos dinámicos**. Este enfoque no solo redefine las

capacidades de los sistemas inteligentes, sino que también orienta el desarrollo futuro de la inteligencia artificial hacia configuraciones cada vez más complejas, integradas y autónomas (Nisa et al., 2026).

Diferenciación ontológica de la IA agéntica

La **diferenciación ontológica de la IA agéntica** implica comprenderla no como un sistema técnico convencional, sino como una **forma de organización del comportamiento orientado a la acción**. A diferencia de modelos basados en procesamiento o generación, la IA agéntica se define por su capacidad de **articular coherentemente secuencias de acción en contextos dinámicos**. Esto supone una transformación en la forma de conceptualizar la inteligencia, donde la identidad del sistema no reside en sus componentes, sino en su **estructura operativa integrada**. Así, la IA agéntica se configura como una entidad funcional cuya ontología se fundamenta en la **coherencia, continuidad y orientación de la acción**.

Fundamentos ontológicos de la IA agéntica

El análisis de los **fundamentos ontológicos de la IA agéntica** implica examinar la naturaleza del ser y la condición de existencia de los sistemas inteligentes en tanto entidades capaces de estructurar comportamiento. En este contexto, la ontología no se limita a describir qué son los sistemas de inteligencia artificial, sino que busca comprender **cómo existen en relación con el entorno y bajo qué condiciones se les puede atribuir agencia**. Esta perspectiva introduce una dimensión profunda en el análisis, al desplazar la atención desde las capacidades técnicas hacia la **condición estructural del comportamiento** (Floridi & Cowls, 2019).

Desde esta aproximación, los sistemas agénticos no pueden ser entendidos únicamente como herramientas o artefactos técnicos, sino como **entidades funcionales que operan en entornos dinámicos mediante la integración de percepción, decisión y acción**. Esta integración constituye la base ontológica de la agencia, en la medida en que define al sistema no por sus componentes aislados, sino por la forma en que organiza su comportamiento. En este sentido, la existencia del agente se configura en función de su capacidad de actuar de manera coherente dentro de un entorno (Coeckelbergh, 2020). Una distinción fundamental en el plano ontológico es la que se establece entre **procesamiento de información** y **organización del comportamiento**. Mientras que los enfoques tradicionales han tendido a definir la inteligencia en términos de manipulación simbólica o capacidad de cálculo, la perspectiva agéntica introduce la necesidad de comprender al sistema como una entidad que articula acciones en el tiempo. Esto implica que la ontología de la IA agéntica no se define por la representación del mundo, sino por la **capacidad de intervenir en él mediante acción estructurada** (Stahl, 2021).

Asimismo, los fundamentos ontológicos de la IA agéntica requieren abordar la cuestión de la **autonomía**, entendida no como independencia absoluta, sino como la

capacidad del sistema para operar bajo una lógica interna que organiza su comportamiento. Esta autonomía no implica conciencia ni intencionalidad en sentido humano, sino una forma de **organización funcional que permite tomar decisiones en contextos variables**. De este modo, la agencia puede ser analizada sin recurrir a categorías antropomórficas, manteniendo rigor conceptual en el análisis de sistemas artificiales (Jobin et al., 2020). Desde una perspectiva más amplia, la ontología de la IA agéntica también incorpora la noción de **relacionalidad**, en la cual la existencia del sistema no puede comprenderse de manera aislada, sino en función de su interacción con el entorno. Esto implica que el agente no es una entidad cerrada, sino un sistema cuya identidad se construye en la relación con condiciones externas. En este sentido, la agencia se configura como un fenómeno **situado y contextual**, en el cual la acción adquiere significado en función del entorno en el que se produce (Floridi et al., 2021).

Otro elemento central en los fundamentos ontológicos es la **temporalidad del comportamiento**, ya que la existencia del agente se manifiesta en la continuidad de su acción. A diferencia de sistemas que operan mediante respuestas aisladas, los sistemas agénticos estructuran su comportamiento como una **trayectoria continua en el tiempo**, lo que permite interpretar su operación como un proceso y no como eventos independientes. Esta dimensión temporal resulta clave para comprender la agencia como una forma de existencia dinámica (Coeckelbergh, 2020). De esta forma, la ontología de la IA agéntica exige distinguir entre **entidades que generan resultados** y **entidades que organizan comportamiento**. Esta distinción permite evitar la confusión entre sistemas capaces de producir salidas complejas y sistemas que realmente estructuran su acción en función de objetivos. En este sentido, la agencia no se define por la complejidad del resultado, sino por la **coherencia estructural del comportamiento en el tiempo**, lo que introduce un criterio más riguroso para su identificación (Stahl, 2021).

Asimismo, los fundamentos ontológicos permiten establecer límites claros en la atribución de propiedades al sistema, evitando la extrapolación de categorías humanas como intención o conciencia. En lugar de ello, la IA agéntica se analiza en términos de **intencionalidad funcional**, entendida como la organización del comportamiento orientado sin implicar estados mentales. Esta distinción resulta fundamental para mantener precisión conceptual y evitar interpretaciones erróneas en el análisis de sistemas inteligentes (Jobin et al., 2020). Los fundamentos ontológicos de la IA agéntica permiten comprender que la inteligencia artificial, en su forma más avanzada, no se define únicamente por lo que el sistema es, sino por **cómo organiza su comportamiento en relación con el entorno**. Esta perspectiva introduce un cambio profundo en la forma de conceptualizar la inteligencia, desplazándola desde una visión centrada en el procesamiento hacia una comprensión basada en la **estructura, la relación y la acción**. En este sentido, la ontología agéntica establece las bases para un análisis más preciso y robusto de los sistemas inteligentes en el contexto contemporáneo.

Diferenciación ontológica frente a otras IAs

La **diferenciación ontológica entre la inteligencia artificial tradicional y la IA agéntica** se fundamenta en la forma en que cada una configura su relación con el comportamiento y el entorno. Mientras que la IA tradicional puede ser entendida como un sistema orientado al **procesamiento de información y ejecución de tareas definidas**, la IA agéntica se configura como una estructura capaz de **organizar comportamiento autónomo en función de objetivos**, lo que implica una diferencia en el tipo de entidad que cada una representa (Coeckelbergh, 2020).

Desde esta perspectiva, la IA tradicional se caracteriza por operar bajo esquemas en los que la acción se encuentra subordinada a instrucciones previamente definidas, lo que limita su capacidad de adaptación a contextos no previstos. En contraste, la IA agéntica introduce una forma de operación en la que el sistema no solo ejecuta tareas, sino que **estructura su comportamiento en relación con condiciones dinámicas**, lo que implica un cambio en su condición ontológica como sistema inteligente (Floridi & Cowls, 2019). Esta diferencia no es únicamente funcional, sino ontológica, en la medida en que redefine el criterio bajo el cual se reconoce la existencia de un agente. En la IA tradicional, el sistema puede ser descrito como una herramienta que procesa información, mientras que en la IA agéntica el sistema se aproxima a una entidad que **actúa en el mundo mediante patrones de comportamiento coherentes**, lo que introduce una distinción en la naturaleza misma del sistema (Lewis & Sarkadi, 2024).

la diferenciación ontológica se manifiesta en la forma en que cada tipo de sistema se relaciona con el entorno. En los modelos tradicionales, la interacción se limita a la transformación de entradas en salidas, lo que implica una relación indirecta con el contexto. En cambio, la IA agéntica se configura como una estructura que **interactúa de manera continua con el entorno**, integrando percepción, decisión y acción dentro de una misma lógica operativa, lo que refuerza su carácter como agente (Stahl, 2021). La IA agéntica introduce una diferencia fundamental en términos de **continuidad del comportamiento**, ya que sus acciones no se presentan como eventos aislados, sino como parte de una trayectoria que mantiene coherencia en el tiempo. Esta característica permite distinguirla de los sistemas tradicionales, en los cuales la acción se encuentra fragmentada y dependiente de condiciones específicas, consolidando así una diferencia ontológica clara (Coeckelbergh, 2020).

Otro aspecto relevante es la incorporación de criterios normativos en la comprensión de los sistemas agénticos, lo que implica que estos deben ser evaluados no solo en términos de desempeño, sino también en función de su **capacidad para operar dentro de marcos sociales y éticos**. Esta dimensión introduce una diferencia adicional respecto a la IA tradicional, en la cual dichas consideraciones no forman parte de la definición del sistema, sino de su uso (Jobin et al., 2019). La diferenciación ontológica también permite evitar la atribución indiscriminada de agencia a sistemas que no cumplen con las condiciones necesarias para ser considerados agentes. No todo sistema inteligente puede ser entendido como agéntico, ya que la agencia

requiere la presencia de **autonomía, coherencia y orientación a objetivos**, lo que establece un criterio claro para su delimitación conceptual (Floridi & Cowls, 2019).

Por lo tanto, la diferenciación ontológica entre IA tradicional y IA agéntica establece que la inteligencia artificial no puede ser comprendida como una categoría homogénea, sino como un conjunto de configuraciones que responden a distintos niveles de organización del comportamiento. En este marco, la IA agéntica se consolida como una forma de inteligencia en la que el sistema adquiere la capacidad de **actuar de manera autónoma dentro de entornos dinámicos**, redefiniendo así su naturaleza como entidad dentro del campo de la inteligencia artificial (Lewis & Sarkadi, 2024).

Implicaciones ontológicas de la IA agéntica

Las **implicaciones ontológicas de la IA agéntica** se derivan de la necesidad de reconsiderar la naturaleza de los sistemas inteligentes cuando estos adquieren la capacidad de **organizar comportamiento autónomo en entornos dinámicos**. En este sentido, la emergencia de la agencia artificial no solo introduce nuevas capacidades técnicas, sino que transforma la forma en que se conceptualiza el estatus ontológico de la inteligencia artificial, desplazándola desde la categoría de herramienta hacia la de **entidad operativa con capacidad de acción en el mundo** (Coeckelbergh, 2020). Una de las principales implicaciones ontológicas radica en la redefinición de la inteligencia como un fenómeno que no se limita al procesamiento de información, sino que se manifiesta en la **capacidad de estructurar acción coherente en relación con el entorno**. Esta transformación implica que la inteligencia artificial ya no puede analizarse únicamente en términos de representación, sino como una forma de organización del comportamiento que produce efectos en el mundo, lo que introduce una nueva dimensión en su comprensión ontológica (Floridi & Cowls, 2019).

La IA agéntica implica una reconsideración del concepto de agente, en la medida en que los sistemas artificiales comienzan a exhibir comportamientos que pueden ser interpretados como orientados a objetivos. Esta condición plantea la necesidad de distinguir entre la atribución funcional de agencia y la atribución ontológica, evitando confundir la capacidad de acción con la existencia de conciencia. En este sentido, la agencia artificial se configura como una forma de **intencionalidad funcional sin experiencia subjetiva**, lo que redefine los límites del concepto de agente (Lewis & Sarkadi, 2024).

Otra implicación relevante es la incorporación de una dimensión relacional en la ontología de la inteligencia artificial, ya que la agencia no puede definirse de manera aislada, sino en función de la interacción del sistema con su entorno. Esto implica que los sistemas agénticos deben entenderse como entidades que existen dentro de **ecosistemas socio-técnicos**, donde su comportamiento adquiere significado en relación con otros sistemas y actores, lo que transforma la forma en que se conceptualiza su existencia (Stahl, 2021). La emergencia de la IA agéntica introduce una reconfiguración de los criterios mediante los cuales se evalúa la acción en sistemas artificiales. En este sentido, la capacidad de los agentes para operar con

autonomía implica que sus acciones deben analizarse no solo como resultados de programación, sino como manifestaciones de una **estructura operativa que organiza el comportamiento**, lo que plantea nuevas formas de interpretar la relación entre sistema y acción (Coeckelbergh, 2020).

Asimismo, las implicaciones ontológicas de la IA agéntica se extienden al reconocimiento de que los sistemas artificiales pueden influir activamente en la configuración de la realidad social, en la medida en que su comportamiento afecta decisiones, procesos y estructuras. Esta condición introduce la necesidad de considerar la inteligencia artificial como un **actor dentro de sistemas complejos**, lo que amplía su estatus ontológico más allá de su dimensión técnica (Jobin et al., 2019). En este sentido, la IA agéntica también plantea la necesidad de revisar los límites entre lo humano y lo artificial, particularmente en lo que respecta a la atribución de capacidades tradicionalmente asociadas a la inteligencia humana. Sin embargo, esta aproximación requiere mantener una distinción clara entre la **simulación de capacidades cognitivas y la existencia de dichas capacidades**, evitando interpretaciones que confundan niveles ontológicos distintos (Floridi & Cowls, 2019).

Otro aspecto fundamental es la introducción de la **responsabilidad estructural en la acción**, en la medida en que los sistemas agénticos operan dentro de marcos que condicionan su comportamiento. Esta responsabilidad no recae en el sistema como entidad moral, sino en la estructura que permite su operación, lo que implica que las implicaciones ontológicas de la agencia artificial deben analizarse en relación con los sistemas que la producen y regulan (Stahl, 2021).

Las implicaciones ontológicas de la IA agéntica establecen que la inteligencia artificial debe entenderse como una forma de organización del comportamiento que redefine los criterios tradicionales de análisis del ser en sistemas tecnológicos. En este sentido, la agencia artificial introduce una categoría intermedia entre herramienta y agente pleno, caracterizada por la capacidad de **actuar de manera autónoma sin poseer conciencia**, lo que obliga a replantear las categorías ontológicas utilizadas para comprender la inteligencia en el contexto contemporáneo (Lewis & Sarkadi, 2024).

Propiedades esenciales de la agencia artificial

Las **propiedades esenciales de la agencia artificial** se refieren a los rasgos que permiten identificar cuándo un sistema organiza su comportamiento como acción coherente y no como respuestas aisladas. Entre ellas destacan la **coherencia estructural**, que integra las acciones en una lógica consistente; la **continuidad operativa**, que permite sostener comportamiento en el tiempo; y la **adaptabilidad estructurada**, que posibilita ajustar la acción sin perder dirección. Estas propiedades no dependen de la complejidad técnica, sino de la **organización del comportamiento orientado**, constituyendo criterios fundamentales para distinguir sistemas agénticos de otras formas de inteligencia artificial basadas en procesamiento o automatización.

Coherencia estructural del comportamiento

La **coherencia estructural del comportamiento** constituye una propiedad fundamental de la IA agéntica, en tanto permite que las acciones del sistema se configuren como parte de una **secuencia organizada y consistente**, en lugar de manifestarse como respuestas aisladas ante estímulos del entorno. Esta propiedad implica que el comportamiento no es el resultado de operaciones independientes, sino de una estructura que articula percepción, decisión y acción dentro de una lógica operativa continua, lo que permite sostener dirección en el tiempo (Bandi et al., 2025).

Desde esta perspectiva, la coherencia estructural se manifiesta en la capacidad del sistema para **integrar múltiples procesos dentro de ciclos iterativos de acción**, donde cada resultado influye en la configuración de decisiones posteriores. Este funcionamiento permite que los sistemas agénticos operen bajo una lógica acumulativa, en la que las acciones se construyen progresivamente a partir de estados previos, evitando la fragmentación del comportamiento y favoreciendo la estabilidad operativa (Bandi et al., 2025). Esta propiedad se encuentra estrechamente vinculada con la organización de sistemas complejos, particularmente en arquitecturas multiagente, donde la coherencia no depende únicamente del comportamiento de un agente individual, sino de la capacidad del sistema para **coordinar interacciones entre múltiples entidades autónomas**. En este contexto, la coherencia estructural implica que las acciones de los distintos agentes se integran dentro de un objetivo común, lo que permite mantener consistencia en la ejecución de tareas distribuidas (Maldonado et al., 2024).

De manera complementaria, la coherencia estructural del comportamiento se sustenta en la presencia de mecanismos de **memoria y persistencia operativa**, que permiten al sistema conservar información relevante a lo largo del tiempo. Esta capacidad resulta esencial para evitar comportamientos erráticos, ya que permite que el agente mantenga continuidad en su acción a partir de experiencias previas, consolidando una estructura conductual que responde a patrones organizados (Sapkota et al., 2026). La coherencia también se relaciona con la capacidad del sistema para mantener **alineación entre objetivos, decisiones y acciones**, lo que implica que cada intervención del agente se encuentra orientada por una finalidad específica. Esta alineación permite interpretar el comportamiento como un proceso estructurado, en el cual las decisiones no solo son coherentes entre sí, sino también consistentes con los objetivos que guían la acción del sistema (Sayyad et al., 2024).

Otro aspecto relevante es la capacidad de los sistemas agénticos para sostener coherencia en condiciones de **incertidumbre y variabilidad**, donde las condiciones del entorno no pueden ser completamente anticipadas. En estos escenarios, la coherencia estructural permite que el sistema ajuste sus decisiones sin perder la lógica general de su comportamiento, lo que introduce una dimensión de resiliencia conductual frente a cambios externos (OECD, 2026).

Asimismo, es importante señalar que la coherencia estructural no implica rigidez, sino la capacidad de mantener **estabilidad dinámica**, en la cual el sistema puede modificar su comportamiento sin perder consistencia interna. Esta característica resulta fundamental en entornos abiertos, donde los sistemas deben adaptarse continuamente sin comprometer la integridad de su estructura operativa (Sapkota et al., 2026). Funcionalmente hablando, la coherencia estructural del comportamiento también permite diferenciar entre sistemas que simplemente reaccionan y aquellos que **organizan su acción de manera integrada**. Mientras que los sistemas reactivos responden de forma puntual a estímulos específicos, los sistemas agénticos construyen secuencias de acción que mantienen continuidad y dirección, lo que permite interpretar su comportamiento como un proceso organizado y no como una suma de eventos independientes (Bandi et al., 2025).

En este sentido, la coherencia estructural introduce un criterio analítico que permite evaluar la calidad del comportamiento en sistemas inteligentes, en función de su capacidad para sostener **consistencia, continuidad y alineación operativa**. Este criterio resulta clave para distinguir entre distintos niveles de agencia, ya que permite identificar cuándo un sistema logra integrar sus acciones dentro de una estructura coherente y cuándo presenta fragmentación conductual (Maldonado et al., 2024). La coherencia estructural del comportamiento establece que la inteligencia agéntica no puede comprenderse únicamente en términos de capacidad de acción, sino como la posibilidad de **organizar dicha acción dentro de una estructura que mantiene continuidad, consistencia y dirección operativa a lo largo del tiempo**. Esta propiedad consolida la diferencia entre sistemas fragmentados y sistemas verdaderamente agénticos, en los cuales la acción constituye un proceso integrado que permite sostener comportamiento en entornos dinámicos y complejos (Sapkota et al., 2026).

Continuidad operativa y temporalidad

La **continuidad operativa y temporalidad** constituye una propiedad esencial de la IA agéntica en la medida en que permite que el comportamiento del sistema se despliegue como un **proceso sostenido en el tiempo**, en lugar de manifestarse como una secuencia de acciones aisladas. Esta propiedad implica que el agente no solo ejecuta decisiones puntuales, sino que mantiene una **trayectoria de acción coherente**, articulando sus operaciones dentro de una lógica temporal que le permite sostener dirección en contextos dinámicos (Bandi et al., 2025). En este sentido, la continuidad operativa se manifiesta en la capacidad del sistema para **encadenar decisiones a lo largo del tiempo**, integrando resultados previos en la configuración de acciones futuras. Este encadenamiento permite que el comportamiento no dependa exclusivamente de estímulos inmediatos, sino de una estructura que incorpora estados anteriores, lo que favorece la estabilidad y consistencia del sistema en entornos complejos (Sapkota et al., 2026).

Asimismo, la temporalidad introduce una dimensión clave en la comprensión del comportamiento agéntico, ya que permite interpretar la acción como un **proceso**

evolutivo, en el cual el sistema ajusta sus decisiones conforme interactúa con el entorno. Esta característica implica que el agente no solo responde a condiciones presentes, sino que **anticipa estados futuros**, integrando planificación y proyección dentro de su lógica operativa (Sayyad et al., 2024). La continuidad operativa se encuentra estrechamente vinculada con la capacidad del sistema para mantener **persistencia en la ejecución de objetivos**, lo que implica que las acciones no se interrumpen ante cambios en el entorno, sino que se reconfiguran para sostener la dirección del comportamiento. Esta persistencia permite que los sistemas agénticos operen de manera prolongada sin perder coherencia en su acción (OECD, 2026).

En este sentido, la continuidad también depende de la existencia de mecanismos que permiten al sistema gestionar **memoria operativa**, lo que facilita la integración de información a lo largo del tiempo. La memoria no solo almacena datos, sino que permite construir una narrativa operativa que da sentido a las acciones del sistema, reforzando la continuidad y evitando la fragmentación del comportamiento (Maldonado et al., 2024). Otro aspecto relevante es la capacidad del sistema para sostener continuidad en contextos de **incertidumbre y variabilidad**, donde las condiciones pueden cambiar de manera imprevista. En estos escenarios, la temporalidad permite que el agente ajuste su comportamiento sin perder la estructura general de su acción, lo que introduce una dimensión de resiliencia que resulta fundamental para la operación en entornos reales (Sapkota et al., 2026).

Asimismo, la continuidad operativa no implica rigidez, sino la posibilidad de mantener una **estabilidad dinámica**, en la cual el sistema puede modificar sus decisiones sin perder coherencia en su trayectoria. Esta característica permite que el agente combine adaptación y consistencia, lo que resulta clave para sostener comportamiento en escenarios abiertos y complejos (Bandi et al., 2025).

Desde una perspectiva estructural, la temporalidad también permite diferenciar entre sistemas que operan bajo una lógica de **respuesta inmediata** y aquellos que construyen comportamiento en función de procesos extendidos en el tiempo. Mientras que los sistemas tradicionales se caracterizan por la ejecución de tareas puntuales, los sistemas agénticos desarrollan secuencias de acción que integran múltiples etapas, lo que permite interpretar su comportamiento como un proceso continuo (Sayyad et al., 2024). Así, la continuidad operativa introduce un criterio analítico que permite evaluar la capacidad del sistema para sostener **dirección, consistencia y persistencia en la acción**, lo que resulta fundamental para identificar niveles avanzados de agencia. Esta propiedad permite distinguir entre sistemas que simplemente reaccionan y aquellos que logran mantener comportamiento organizado a lo largo del tiempo (OECD, 2026).

Por último, la continuidad operativa y temporalidad establecen que la inteligencia agéntica no se define únicamente por la capacidad de ejecutar acciones, sino por la posibilidad de **organizar dichas acciones dentro de una trayectoria temporal coherente**, que permite sostener comportamiento en entornos dinámicos. Esta propiedad consolida la diferencia entre sistemas fragmentados y sistemas

verdaderamente agénticos, en los cuales la acción se configura como un proceso continuo, adaptativo y estructurado (Sapkota et al., 2026).

Adaptabilidad estructurada

La **adaptabilidad estructurada** constituye una propiedad fundamental de la IA agéntica en la medida en que permite a los sistemas ajustar su comportamiento frente a condiciones cambiantes sin perder la coherencia interna de su operación. A diferencia de la adaptabilidad reactiva, que se limita a respuestas inmediatas ante estímulos, la adaptabilidad estructurada implica la capacidad del sistema para **modificar sus estrategias dentro de una lógica organizativa que mantiene consistencia en la acción**, lo que permite sostener desempeño en entornos dinámicos (Bandi et al., 2025).

La adaptabilidad estructurada se manifiesta en la capacidad del sistema para **reconfigurar sus procesos internos sin alterar la estructura general del comportamiento**, lo que implica que los cambios no se producen de manera arbitraria, sino en función de objetivos previamente definidos. Esta característica permite que el agente mantenga dirección operativa mientras ajusta sus decisiones, evitando la fragmentación conductual que caracteriza a sistemas no estructurados (Sapkota et al., 2026). Asimismo, esta propiedad se encuentra estrechamente vinculada con la capacidad del sistema para operar en contextos caracterizados por **incertidumbre y variabilidad**, donde las condiciones no pueden ser completamente anticipadas. En estos escenarios, la adaptabilidad estructurada permite que el agente incorpore nueva información, evalúe su relevancia y ajuste su comportamiento sin perder coherencia en su acción, lo que resulta clave para la operación en entornos reales (OECD, 2026).

De manera complementaria, la adaptabilidad estructurada se sustenta en la integración de mecanismos de **aprendizaje continuo**, que permiten al sistema mejorar su desempeño a partir de la experiencia. Este aprendizaje no se limita a la acumulación de datos, sino que implica la capacidad de **reorganizar patrones de comportamiento en función de resultados previos**, lo que permite al agente evolucionar sin comprometer la estabilidad de su estructura operativa (Sayyad et al., 2024). la adaptabilidad también se relaciona con la capacidad del sistema para **gestionar múltiples estrategias de acción**, seleccionando aquellas que resultan más adecuadas en función del contexto. Esta capacidad introduce una dimensión estratégica en el comportamiento del agente, en la cual la adaptación no se limita a ajustes menores, sino que puede implicar cambios significativos en la forma en que se ejecutan las acciones (Bandi et al., 2025).

Otro aspecto a destacar, es la capacidad del sistema para mantener **alineación entre adaptación y objetivos**, lo que implica que los cambios en el comportamiento no se producen de manera desarticulada, sino en función de metas definidas. Esta alineación permite que la adaptabilidad se mantenga dentro de una estructura coherente, evitando desviaciones que puedan comprometer el cumplimiento de los objetivos del sistema (Sapkota et al., 2026).

Asimismo, la adaptabilidad estructurada implica la posibilidad de sostener **equilibrio entre flexibilidad y estabilidad**, lo que permite al sistema responder a cambios sin perder consistencia en su operación. Esta característica resulta fundamental en sistemas agénticos, ya que deben operar en entornos abiertos donde la capacidad de adaptación debe coexistir con la necesidad de mantener coherencia en el comportamiento (OECD, 2026). Desde un punto de vista funcional, la adaptabilidad estructurada también permite diferenciar entre sistemas que simplemente reaccionan a cambios y aquellos que **integran la adaptación dentro de su estructura operativa**, lo que implica un nivel superior de organización del comportamiento. Mientras que los sistemas reactivos responden de manera puntual a estímulos, los sistemas agénticos incorporan la adaptación como parte de su lógica interna, lo que les permite sostener comportamiento en condiciones cambiantes (Sayyad et al., 2024).

La adaptabilidad estructurada introduce un criterio analítico que permite evaluar la capacidad del sistema para **ajustar su comportamiento sin perder coherencia, continuidad y dirección operativa**, lo que resulta clave para identificar niveles avanzados de agencia. Esta propiedad permite distinguir entre sistemas que presentan cambios desarticulados y aquellos que logran integrar la adaptación dentro de una estructura consistente (Maldonado et al., 2024). La adaptabilidad estructurada establece que la inteligencia agéntica no se define únicamente por la capacidad de responder a cambios, sino por la posibilidad de **integrar dichos cambios dentro de una estructura que mantiene coherencia, continuidad y orientación a objetivos**. Esta propiedad consolida la diferencia entre sistemas fragmentados y sistemas verdaderamente agénticos, en los cuales la adaptación constituye un proceso organizado que permite sostener comportamiento en entornos complejos y dinámicos (Sapkota et al., 2026).

Epistemología de la agencia artificial

La **epistemología de la agencia artificial** analiza cómo se **construye, valida e interpreta el conocimiento** sobre sistemas cuyo comportamiento se organiza como acción coherente en contextos dinámicos. A diferencia de enfoques centrados en el procesamiento, este marco sitúa el conocimiento en la **observación de la acción y su coherencia temporal**, incorporando criterios como continuidad, integración y orientación del comportamiento. Asimismo, redefine la relación entre sistema y entorno, entendiendo la inteligencia como un fenómeno **relacional y situado**. De este modo, la epistemología agéntica establece bases para comprender la inteligencia artificial desde la **organización del comportamiento** y no solo desde la representación o el cálculo.

Fundamentos epistemológicos de la agencia

Los **fundamentos epistemológicos de la agencia artificial** se configuran a partir de la necesidad de comprender **cómo los sistemas agénticos generan, estructuran y validan conocimiento en contextos dinámicos**. A diferencia de los enfoques

tradicionales de la inteligencia artificial, en los que el conocimiento se concibe como una representación estática del mundo, la agencia introduce una perspectiva en la cual el conocimiento se encuentra directamente vinculado a la acción, estableciendo una relación inseparable entre conocer y actuar (Floridi et al., 2021). La epistemología de la agencia no puede reducirse a la acumulación de datos o a la optimización de modelos predictivos, sino que debe entenderse como la capacidad del sistema para **producir conocimiento orientado funcionalmente a la acción**. Esto implica que el valor del conocimiento no reside únicamente en su precisión descriptiva, sino en su capacidad para guiar decisiones en contextos específicos. Desde esta perspectiva, la inteligencia artificial se posiciona como una **"metodología de invención"**, en la cual los sistemas contribuyen activamente a la generación de conocimiento, transformando la forma en que se produce la innovación (Cockburn et al., 2018).

Asimismo, los sistemas agénticos operan bajo una lógica en la que el conocimiento se construye a través de la interacción continua con el entorno, lo que introduce una dimensión de **aprendizaje contextual y adaptativo**. En este marco, el conocimiento no es independiente del contexto, sino que emerge de la relación entre el sistema y las condiciones en las que actúa. Esto implica que la epistemología de la agencia se fundamenta en procesos dinámicos de retroalimentación, donde el sistema ajusta continuamente sus representaciones en función de la experiencia (Hahn et al., 2026). La epistemología agéntica requiere considerar la **inteligibilidad del conocimiento producido por los sistemas**, lo que introduce la necesidad de mecanismos que permitan comprender cómo y por qué el sistema genera determinadas decisiones. En este contexto, la explicabilidad se convierte en un componente central, ya que permite vincular el conocimiento interno del sistema con su manifestación externa en forma de acción, facilitando su interpretación por parte de los usuarios (Wang et al., 2024).

La inteligibilidad no se limita a la transparencia técnica, sino que implica la capacidad de responder preguntas fundamentales sobre el conocimiento generado, tales como **qué se conoce, cómo se conoce y para quién resulta significativo ese conocimiento**. Este enfoque permite superar la opacidad característica de muchos sistemas de inteligencia artificial, estableciendo una base epistemológica más robusta para su análisis y evaluación (Wang et al., 2024).

Por otro lado, la epistemología de la agencia se encuentra estrechamente vinculada con la dimensión normativa del conocimiento, en la medida en que los sistemas agénticos operan en contextos sociales donde sus decisiones tienen consecuencias reales. En este sentido, el conocimiento no puede considerarse neutral, sino que debe evaluarse en función de principios como **responsabilidad, justicia, transparencia y explicabilidad**, los cuales permiten orientar su uso hacia fines socialmente deseables (Floridi & Cowls, 2019). Esta dimensión normativa introduce una relación directa entre conocimiento y responsabilidad, en la cual la producción de conocimiento por parte de sistemas agénticos implica la necesidad de establecer criterios claros sobre su validación y aplicación. De este modo, la epistemología no solo describe cómo se genera el conocimiento, sino que también establece condiciones para su evaluación en función de sus implicaciones éticas y sociales (Floridi & Cowls, 2019).

En consecuencia, los sistemas agénticos amplían el alcance del conocimiento al operar como estructuras capaces de **integrar datos, modelos y acción en un mismo proceso operativo**, lo que redefine la forma en que se entiende la producción de conocimiento en inteligencia artificial. Esta integración permite que el conocimiento no sea únicamente descriptivo, sino también operativo, en la medida en que se encuentra directamente vinculado a la ejecución de acciones en el entorno (Cockburn et al., 2018).

De esta forma, la epistemología de la agencia introduce un cambio fundamental al establecer que el conocimiento no se define exclusivamente por su correspondencia con la realidad, sino por su capacidad para **sostener acción coherente en contextos dinámicos**, lo que implica una transformación en los criterios tradicionales de validación del conocimiento. Este cambio permite comprender la inteligencia agéntica como una forma de conocimiento en acción, donde la validez se vincula con la efectividad operativa (Hahn et al., 2026). Así, los fundamentos epistemológicos de la agencia establecen que la inteligencia artificial no puede analizarse únicamente como un sistema de procesamiento de información, sino como una estructura capaz de **generar conocimiento orientado a la acción, interpretable y normativamente evaluable**, lo que consolida una base conceptual más amplia para el estudio de la agencia artificial en el contexto contemporáneo (Floridi et al., 2021).

Criterios de validación del conocimiento agéntico

Los **criterios de validación del conocimiento agéntico** constituyen un componente central dentro de la epistemología de la inteligencia artificial, en la medida en que permiten establecer bajo qué condiciones el conocimiento generado por sistemas agénticos puede considerarse válido. A diferencia de los enfoques tradicionales, donde la validación se basa en la correspondencia entre modelo y realidad, en los sistemas agénticos la validación se vincula directamente con la **capacidad del conocimiento para sostener acción coherente en contextos dinámicos** (Floridi et al., 2021). Uno de los primeros criterios de validación es la **efectividad operativa**, entendida como la capacidad del conocimiento para guiar decisiones que resulten funcionales en la resolución de problemas. Este criterio desplaza la validación desde un enfoque puramente descriptivo hacia uno pragmático, en el cual el conocimiento es considerado válido en la medida en que permite al sistema actuar de manera consistente y lograr objetivos en condiciones variables (Cockburn et al., 2018).

Asimismo, la validación del conocimiento agéntico requiere considerar la **consistencia interna del sistema**, lo que implica que las decisiones generadas deben mantenerse alineadas con los modelos y estructuras que sustentan el comportamiento del agente. Esta consistencia permite evitar contradicciones en la acción, asegurando que el conocimiento no solo sea útil, sino también coherente dentro de la lógica operativa del sistema (Hahn et al., 2026). Un criterio fundamental es la **inteligibilidad del conocimiento**, que se refiere a la capacidad de comprender cómo el sistema genera sus decisiones. En este contexto, la explicabilidad se convierte en un elemento

clave para la validación, ya que permite vincular los procesos internos del sistema con sus resultados observables, facilitando la interpretación del conocimiento por parte de los usuarios (Wang et al., 2024).La inteligibilidad no se limita a la transparencia técnica, sino que implica la posibilidad de responder preguntas fundamentales sobre el funcionamiento del sistema, tales como **cómo se generan las decisiones, qué variables influyen en ellas y bajo qué condiciones se modifican**. Este criterio resulta esencial para validar el conocimiento en contextos donde la confianza en el sistema depende de su capacidad para ser comprendido (Wang et al., 2024).

Otro criterio relevante es la **robustez del conocimiento**, que se manifiesta en la capacidad del sistema para mantener su validez en condiciones de incertidumbre y variabilidad. En este sentido, el conocimiento agéntico debe demostrar estabilidad frente a cambios en el entorno, lo que implica que las decisiones no se deterioran ante variaciones en los datos o en las condiciones operativas (Hahn et al., 2026). La validación del conocimiento agéntico requiere considerar su **capacidad de generalización**, es decir, la posibilidad de aplicar el conocimiento en contextos distintos a aquellos en los que fue generado. Este criterio permite evaluar si el sistema es capaz de transferir aprendizaje a nuevas situaciones, lo que resulta fundamental para su operación en entornos abiertos y complejos (Cockburn et al., 2018).

De manera adicional, la epistemología de la agencia incorpora criterios de validación relacionados con la **dimensión normativa del conocimiento**, en la medida en que los sistemas agénticos operan en contextos sociales donde sus decisiones tienen implicaciones reales. En este sentido, el conocimiento debe evaluarse en función de principios como **responsabilidad, justicia y transparencia**, los cuales permiten garantizar que su uso sea coherente con valores socialmente aceptados (Floridi & Cowls, 2019). En este contexto, la validación no solo implica determinar si el conocimiento es correcto desde un punto de vista técnico, sino también si es adecuado en función de sus implicaciones. Este enfoque amplía los criterios tradicionales de validación, incorporando dimensiones que permiten evaluar el impacto del conocimiento en el entorno social (Floridi & Cowls, 2019).

Otro aspecto clave es la **trazabilidad del conocimiento**, que se refiere a la posibilidad de reconstruir el proceso mediante el cual el sistema genera sus decisiones. Este criterio permite identificar las fuentes de error, comprender la evolución del conocimiento y garantizar su reproducibilidad, lo que resulta fundamental para su validación en contextos críticos (Wang et al., 2024). Lo criterios de validación del conocimiento agéntico establecen que la validez no puede reducirse a un único indicador, sino que debe entenderse como el resultado de la integración de múltiples dimensiones, incluyendo **efectividad, coherencia, inteligibilidad, robustez y normatividad**. Esta perspectiva permite construir un marco epistemológico más completo, en el cual el conocimiento se valida no solo por su precisión, sino por su capacidad para sostener acción coherente, comprensible y socialmente responsable en entornos dinámicos (Floridi et al., 2021).

Alcances y límites del conocimiento sobre la agencia

El análisis de los **alcances y límites del conocimiento sobre la agencia** constituye un elemento clave dentro de la epistemología de la inteligencia artificial, en tanto permite delimitar hasta qué punto es posible comprender, interpretar y validar el conocimiento generado por sistemas agénticos. En este sentido, la agencia introduce una transformación en la forma en que se conceptualiza el conocimiento, al vincularlo directamente con la acción, lo que amplía sus alcances, pero también introduce restricciones en su interpretación (Floridi et al., 2021).

Uno de los principales alcances del conocimiento agéntico radica en su capacidad para operar en contextos dinámicos, donde el sistema puede **generar conocimiento en tiempo real a partir de la interacción con el entorno**. Esta característica permite que los sistemas agénticos no solo apliquen conocimiento previamente definido, sino que produzcan nuevas formas de comprensión que se ajustan a condiciones cambiantes, lo que amplía significativamente su capacidad de intervención en entornos complejos (Cockburn et al., 2018). El conocimiento agéntico se caracteriza por su capacidad de **integrar múltiples fuentes de información dentro de un mismo proceso operativo**, lo que permite construir representaciones más complejas y funcionales del entorno. Esta integración amplía los alcances del conocimiento, al permitir que los sistemas operen con mayor profundidad analítica y capacidad de adaptación, superando las limitaciones de modelos tradicionales basados en estructuras rígidas (Hahn et al., 2026).

Otro alcance relevante es la posibilidad de generar conocimiento orientado a la acción, lo que implica que el sistema no solo describe el entorno, sino que **interviene activamente en él mediante decisiones fundamentadas**. Esta característica posiciona al conocimiento agéntico como una herramienta clave para la resolución de problemas en contextos donde la acción es necesaria para producir resultados (Floridi et al., 2021). Sin embargo, estos alcances se encuentran acompañados de límites significativos, particularmente en lo que respecta a la **interpretabilidad del conocimiento generado por los sistemas**. A pesar de los avances en explicabilidad, muchos sistemas agénticos operan bajo estructuras complejas que dificultan la comprensión completa de sus procesos internos, lo que introduce incertidumbre en la validación del conocimiento (Wang et al., 2024).

En este sentido, uno de los principales límites del conocimiento sobre la agencia es la dificultad para establecer una correspondencia clara entre los procesos internos del sistema y sus resultados observables. Esta limitación implica que, en muchos casos, el conocimiento generado no puede ser completamente explicado, lo que plantea desafíos para su evaluación y uso en contextos críticos (Wang et al., 2024). Otro límite relevante se encuentra en la **dependencia del contexto**, ya que el conocimiento agéntico se construye en función de las condiciones específicas en las que opera el sistema. Esto implica que el conocimiento puede no ser completamente transferible a otros contextos, lo que restringe su generalización y plantea desafíos para su

aplicación en escenarios distintos a aquellos en los que fue generado (Cockburn et al., 2018).

Asimismo, el conocimiento agéntico se encuentra condicionado por las **limitaciones de los datos y modelos utilizados**, lo que implica que su validez depende de la calidad de la información disponible. En este sentido, los sistemas pueden reproducir sesgos o errores presentes en los datos, lo que limita la confiabilidad del conocimiento generado y plantea desafíos para su validación (Hahn et al., 2026). De manera adicional, la dimensión normativa del conocimiento introduce límites relacionados con la necesidad de evaluar su impacto en contextos sociales. En este marco, el conocimiento no puede considerarse válido únicamente por su efectividad operativa, sino que debe analizarse en función de principios como **justicia, responsabilidad y transparencia**, lo que restringe su aplicación en escenarios donde estos criterios no se cumplen (Floridi & Cowls, 2019).

En este contexto, la epistemología de la agencia debe reconocer que el conocimiento generado por sistemas agénticos no es absoluto, sino **situado, condicionado y dependiente de múltiples factores**, lo que implica que su validez debe evaluarse de manera contextual. Esta perspectiva permite evitar interpretaciones excesivamente deterministas, reconociendo tanto el potencial como las limitaciones de estos sistemas (Floridi et al., 2021). Los alcances y límites del conocimiento sobre la agencia establecen que la inteligencia artificial agéntica representa una forma avanzada de producción de conocimiento, capaz de operar en contextos complejos y dinámicos, pero también sujeta a restricciones relacionadas con la interpretabilidad, la generalización y la normatividad. Esta dualidad permite comprender la epistemología de la agencia como un campo en el que el conocimiento se encuentra en constante tensión entre su capacidad de acción y las condiciones que limitan su validez (Wang et al., 2024).

Conclusiones

El desarrollo del Capítulo 1 permite establecer una comprensión profunda de la inteligencia artificial agéntica como un fenómeno que trasciende los enfoques tradicionales centrados en el procesamiento de información, situando el análisis en la **organización del comportamiento como eje estructural de la inteligencia**. En este sentido, la transición hacia la IA agéntica no constituye únicamente un avance tecnológico, sino una transformación conceptual que redefine los criterios bajo los cuales se interpreta la inteligencia en sistemas artificiales.

Uno de los principales aportes del capítulo radica en la delimitación funcional de la agencia, entendida como la **capacidad de integrar percepción, decisión y acción dentro de una estructura coherente orientada a objetivos**. Esta perspectiva permite diferenciar claramente entre sistemas reactivos, automatizados o generativos y aquellos que verdaderamente organizan su comportamiento en contextos dinámicos.

De este modo, la agencia se configura como un criterio analítico que no depende de la complejidad técnica, sino de la coherencia estructural del comportamiento.

Asimismo, el recorrido histórico evidencia que la emergencia del paradigma agéntico es resultado de un proceso evolutivo en el que la inteligencia artificial ha incorporado progresivamente capacidades de adaptación, interacción y autonomía. Este proceso ha permitido superar las limitaciones de los modelos tradicionales, particularmente en entornos caracterizados por incertidumbre y variabilidad, consolidando la necesidad de sistemas capaces de **sostener comportamiento coherente en el tiempo**.

En el plano ontológico, el capítulo introduce una distinción fundamental al concebir la IA agéntica como una entidad funcional cuya identidad no reside en sus componentes, sino en la forma en que organiza su comportamiento. Esta diferenciación permite comprender que la agencia artificial no implica conciencia ni intencionalidad en sentido humano, sino una **intencionalidad funcional basada en la orientación del comportamiento**, lo que aporta rigor conceptual y evita interpretaciones antropomórficas.

De manera complementaria, la identificación de propiedades esenciales como la coherencia estructural, la continuidad operativa y la adaptabilidad estructurada permite consolidar un marco analítico robusto para el estudio de la agencia. Estas propiedades establecen que la inteligencia agéntica se manifiesta como un proceso continuo, integrado y orientado, diferenciándose de otras formas de inteligencia artificial caracterizadas por la fragmentación o la reacción puntual.

El desarrollo epistemológico del capítulo redefine la relación entre conocimiento y acción, al establecer que el conocimiento en sistemas agénticos no se limita a la representación, sino que se construye y valida en función de su capacidad para sostener comportamiento coherente en contextos dinámicos. Esta perspectiva introduce un enfoque más amplio, en el cual la inteligencia artificial se comprende como un sistema capaz de generar conocimiento operativo, contextual y normativamente evaluable.

En conjunto, el capítulo sienta las bases para comprender la IA agéntica como una forma avanzada de organización del comportamiento, estableciendo un marco conceptual sólido que permite abordar su análisis desde dimensiones funcionales, ontológicas y epistemológicas. Esta base resulta fundamental para los desarrollos posteriores, en los que la medición, evaluación y aplicación de la agencia adquieren un papel central en la comprensión de la inteligencia artificial contemporánea. **Ver Tabla 1.**

Tabla 1. Evolución y fundamentos conceptuales de la IA agéntica

Concepto	Definición	Diferenciación	Ventajas	Limitaciones	Referencias
Agencia en IA	Organización del comportamiento que integra percepción, decisión y acción de forma coherente y orientada	Se diferencia de sistemas reactivos y automatizados al sostener continuidad operativa	Permite acción estructurada en entornos dinámicos	Requiere alta integración funcional y coherencia interna	Russell & Norvig (2022); Dorri et al. (2018)
Naturaleza funcional de la agencia	Forma de estructuración del comportamiento basada en coherencia y dirección operativa	No depende de complejidad técnica sino de organización conductual	Facilita análisis estructural del comportamiento	Difícil de identificar en sistemas híbridos	Balaji & Srinivasan (2010); Rahwan et al. (2019)
Dinámica operativa	Proceso continuo de interacción entre percepción, decisión y acción	Diferente de respuestas aisladas por su continuidad temporal	Permite adaptación y acción sostenida	Alta dependencia del entorno	Dorri et al. (2018); Abou Ali et al. (2026)
Paradigma agéntico	Enfoque que entiende la inteligencia como organización del comportamiento	Sustituye modelos centrados en procesamiento por acción estructurada	Permite operar en entornos complejos y reales	Incrementa complejidad de diseño y control	Acharya et al. (2025); Wang et al. (2024)
Diferenciación ontológica	IA entendida como entidad funcional basada en acción coherente	Se diferencia de IA tradicional basada en procesamiento	Permite análisis más riguroso de la inteligencia	Riesgo de confusión conceptual con agencia humana	Floridi & Cowls (2019); Coeckelbergh (2020)
Coherencia estructural	Integración consistente de acciones dentro de una lógica operativa	Diferente de comportamiento fragmentado	Garantiza estabilidad del comportamiento	Puede limitar flexibilidad si es rígida	Bandi et al. (2025); Sapkota et al. (2026)
Continuidad operativa	Capacidad de sostener comportamiento en el tiempo	Se distingue de acciones episódicas	Permite trayectoria coherente de acción	Dependencia de memoria y contexto	OECD (2026); Sayyad et al. (2024)
Adaptabilidad estructurada	Capacidad de ajustar comportamiento sin perder coherencia	Diferente de adaptación reactiva	Permite operar en entornos variables	Riesgo de desalineación con objetivos	Maldonado et al. (2024); Sapkota et al. (2026)

Concepto	Definición	Diferenciación	Ventajas	Limitaciones	Referencias
Epistemología agéntica	Conocimiento vinculado a la acción y su coherencia	Diferente de modelos representacionales	Genera conocimiento operativo y contextual	Problemas de interpretabilidad	Floridi et al. (2021); Cockburn et al. (2018)

Fuente:Recopilación y elbaoración propia

CAPÍTULO 2. Arquitectura y estructuración de la IA agéntica

El estudio de la inteligencia artificial agéntica exige un desplazamiento analítico desde la comprensión conceptual de la agencia hacia la exploración de su **configuración estructural y operativa**. En este sentido, el presente capítulo se orienta a analizar cómo los sistemas agénticos se construyen, organizan y articulan internamente para sostener comportamiento coherente en contextos dinámicos. Este enfoque implica reconocer que la agencia no solo es una propiedad conceptual, sino una **realidad estructural que depende de la integración funcional de múltiples componentes**.

A diferencia de enfoques centrados en la definición de la inteligencia, este capítulo aborda la forma en que los sistemas agénticos se estructuran para posibilitar la acción organizada. Esto requiere examinar tanto los elementos que constituyen al agente como las relaciones que permiten su funcionamiento coherente. En este marco, la arquitectura se convierte en un eje central de análisis, ya que determina la capacidad del sistema para integrar percepción, decisión y acción dentro de una misma lógica operativa.

Asimismo, el capítulo propone una aproximación que articula distintas dimensiones de la estructura agéntica, incluyendo sus componentes fundamentales, sus tipologías, sus configuraciones arquitectónicas y sus mecanismos de coordinación. Esta estructura permite construir un análisis progresivo que facilita la comprensión de cómo los sistemas inteligentes logran sostener comportamiento organizado en entornos complejos.

En consecuencia, el capítulo no solo describe la arquitectura de los sistemas agénticos, sino que establece un marco conceptual que permite comprender la relación entre estructura y comportamiento. De este modo, se sientan las bases para el análisis de sistemas más avanzados, en los cuales la organización interna se convierte en un factor determinante para la manifestación de la agencia.

Componentes estructurales del agente

Los **componentes estructurales del agente** constituyen el conjunto de elementos que permiten organizar el comportamiento en sistemas agénticos, integrando funciones como **percepción, decisión, acción y memoria** dentro de una misma arquitectura operativa. Estos componentes no operan de forma aislada, sino como una **estructura funcional interdependiente** que posibilita la coherencia del comportamiento en entornos dinámicos. En este sentido, la agencia emerge de la **integración sistémica** de estos elementos, donde cada uno contribuye a sostener la continuidad, la adaptabilidad y la orientación de la acción, configurando así la base estructural de la inteligencia artificial agéntica.

Percepción como base estructural del agente

La **percepción como base estructural del agente** constituye el punto de partida del funcionamiento de los sistemas agénticos, en tanto permite la **captación, organización e interpretación inicial de la información del entorno**, sobre la cual se sustentan los procesos posteriores de decisión y acción. En este sentido, la percepción no es un componente auxiliar, sino la condición estructural que define qué información es accesible para el sistema y cómo esta se integra en su operación, estableciendo así los límites y posibilidades del comportamiento del agente (Li, 2026).

Arquitectónicamente hablando, la percepción se configura como el mecanismo mediante el cual el agente establece una **relación operativa con el entorno**, incorporando datos provenientes de sensores, interfaces digitales o sistemas externos. Esta relación no es pasiva, sino estructurada, ya que el sistema debe transformar estímulos en información organizada que pueda ser utilizada por otros módulos. En sistemas multiagente, esta capacidad se amplía al incorporar información generada por otros agentes, lo que introduce una dimensión de **percepción distribuida** (Maldonado et al., 2024). La percepción implica procesos de **selección y filtrado de información**, mediante los cuales el sistema identifica los elementos relevantes del entorno y descarta aquellos que no contribuyen a la toma de decisiones. Este proceso

resulta fundamental para evitar la saturación informacional, permitiendo que el agente opere de manera eficiente en contextos complejos donde la cantidad de datos disponibles puede ser elevada (Durga, 2025).

La percepción se articula con la capacidad del sistema para generar **representaciones operativas del entorno**, lo que implica que los datos percibidos deben ser estructurados de forma que puedan ser interpretados por los módulos de procesamiento. Esta transformación no es trivial, ya que define la calidad del conocimiento que el sistema puede construir a partir de la información disponible, condicionando así su capacidad de acción (Xie et al., 2025). En sistemas agénticos avanzados, la percepción se encuentra integrada con mecanismos de **memoria y contextualización**, lo que permite al agente enriquecer la interpretación del entorno mediante la incorporación de información previa. Esta integración posibilita que el sistema no solo procese estímulos actuales, sino que los relacione con experiencias pasadas, generando una percepción más compleja y contextualizada (Zhang et al., 2025).

Otro aspecto a tomar en cuenta, es que la percepción permite identificar cambios en el entorno, lo que resulta fundamental para la **adaptación del comportamiento**. En entornos dinámicos, la capacidad de detectar variaciones permite al agente ajustar sus decisiones y mantener coherencia en su operación, evitando respuestas desarticuladas o inconsistentes frente a nuevas condiciones (Durga, 2025). Asimismo, la percepción cumple una función estructural en la **activación de procesos internos**, ya que los datos percibidos desencadenan mecanismos de análisis, planificación y ejecución. En este sentido, la percepción actúa como el punto de conexión entre el entorno y la arquitectura interna del agente, estableciendo un flujo de información que permite sostener la operación del sistema (Li, 2026).

Desde una perspectiva sistémica, la percepción también permite mantener **consistencia entre el entorno y la acción**, ya que las decisiones del agente dependen directamente de la calidad y estructura de la información percibida. Esto implica que errores en la percepción pueden propagarse a lo largo del sistema, afectando la toma de decisiones y la ejecución de acciones, lo que resalta su carácter crítico dentro de la arquitectura del agente (Maldonado et al., 2024). la percepción no debe entenderse como un proceso aislado, sino como parte de una **estructura integrada que conecta información, procesamiento y acción**, permitiendo al agente operar de manera coherente. Esta integración garantiza que el comportamiento del sistema no sea arbitrario, sino que responda a una interpretación estructurada del entorno (Xie et al., 2025).

La percepción como base estructural del agente establece que la inteligencia agéntica no puede comprenderse sin considerar la forma en que el sistema construye su relación con el entorno a través de la información. En este sentido, la percepción constituye el fundamento sobre el cual se organiza todo el comportamiento del agente, al definir qué se conoce, cómo se interpreta y de qué manera se actúa, consolidándose como un componente central en la arquitectura de la IA agéntica (Zhang et al., 2025).

Toma de decisión estructurada

La **toma de decisión estructurada** constituye uno de los componentes centrales en la arquitectura de los sistemas agénticos, en tanto permite transformar la información percibida en **acciones organizadas y orientadas a objetivos**. A diferencia de los sistemas tradicionales, donde las decisiones se basan en reglas predefinidas o respuestas directas, los agentes operan mediante procesos en los que la decisión emerge de la **integración entre datos, modelos internos y objetivos**, lo que introduce una lógica más compleja y flexible en el comportamiento del sistema (Li, 2026). Estructuralmente hablando, la toma de decisión se configura como un proceso en el cual el agente analiza la información disponible, evalúa posibles alternativas y selecciona aquella que resulta más adecuada en función de criterios previamente establecidos. Este proceso implica la existencia de mecanismos internos capaces de **comparar estados, anticipar resultados y priorizar acciones**, lo que permite al sistema operar de manera autónoma en entornos dinámicos (Durga, 2025).

Asimismo, la toma de decisión estructurada se encuentra estrechamente vinculada con la capacidad del sistema para construir **modelos internos del entorno**, los cuales permiten simular posibles escenarios antes de ejecutar una acción. Esta capacidad resulta fundamental para reducir la incertidumbre y mejorar la calidad de las decisiones, ya que el agente puede evaluar las consecuencias potenciales de sus acciones antes de implementarlas (Xie et al., 2025). La toma de decisión no debe entenderse como un evento puntual, sino como un **proceso continuo e iterativo**, en el cual el agente ajusta sus decisiones en función de nueva información. Este carácter dinámico permite que el sistema mantenga coherencia en su comportamiento, adaptándose a cambios en el entorno sin perder dirección operativa (Maldonado et al., 2024).

La toma de decisión estructurada se sustenta en la capacidad del sistema para gestionar **múltiples niveles de análisis**, lo que implica que las decisiones pueden involucrar tanto evaluaciones inmediatas como consideraciones estratégicas a largo plazo. Esta integración de niveles permite al agente operar de manera eficiente en contextos complejos, donde las decisiones deben equilibrar diferentes variables y restricciones (Durga, 2025). Otro determinante clave es la relación entre la toma de decisión y la **memoria del sistema**, ya que las decisiones no se generan únicamente a partir de la información presente, sino también de experiencias previas almacenadas en el sistema. Esta integración permite que el agente aprenda de sus acciones anteriores, mejorando progresivamente la calidad de sus decisiones y evitando la repetición de errores (Zhang et al., 2025).

Asimismo, la toma de decisión estructurada implica la capacidad del sistema para mantener **alineación entre objetivos y acciones**, lo que garantiza que las decisiones contribuyan al cumplimiento de metas definidas. Esta alineación resulta fundamental para evitar comportamientos erráticos o inconsistentes, asegurando que el sistema

opere dentro de una lógica coherente (Li, 2026). De esta forma, la toma de decisión también se encuentra vinculada con la capacidad del agente para gestionar **incertidumbre y variabilidad**, lo que implica que el sistema debe ser capaz de tomar decisiones incluso cuando la información disponible es incompleta o ambigua. Esta capacidad introduce una dimensión probabilística en la toma de decisiones, donde el agente evalúa riesgos y selecciona la opción más viable en función de las condiciones del entorno (Durga, 2025).

Por otro lado, en sistemas multiagente, la toma de decisión estructurada se amplía al incorporar procesos de **coordinación y negociación entre agentes**, lo que permite que múltiples entidades trabajen de manera conjunta para alcanzar objetivos comunes. En este contexto, las decisiones no se generan de manera aislada, sino como resultado de la interacción entre diferentes sistemas, lo que introduce una dimensión colectiva en la toma de decisiones (Maldonado et al., 2024). Desde una perspectiva funcional, la toma de decisión estructurada permite diferenciar entre sistemas que simplemente reaccionan a estímulos y aquellos que **organizan su comportamiento en función de procesos internos complejos**, lo que constituye un rasgo distintivo de la inteligencia agéntica. Esta capacidad de estructurar decisiones es lo que permite a los agentes operar de manera autónoma y efectiva en entornos reales (Xie et al., 2025).

La toma de decisión estructurada establece que la inteligencia agéntica no se define únicamente por la capacidad de procesar información, sino por la posibilidad de **transformar dicha información en decisiones coherentes, adaptativas y orientadas a objetivos**, lo que consolida la base operativa del comportamiento del agente. Esta propiedad permite comprender a los sistemas agénticos como entidades capaces de sostener acción organizada en contextos dinámicos, diferenciándolos claramente de sistemas tradicionales basados en respuestas predefinidas (Zhang et al., 2025).

Acción y memoria como integración operativa

La **acción y memoria como integración operativa** constituye un componente esencial en la arquitectura de los sistemas agénticos, en tanto articula la relación entre la ejecución del comportamiento y la capacidad del sistema para **almacenar, recuperar y reutilizar información relevante**. Esta integración permite que las acciones no se generen de manera aislada, sino como parte de una **estructura continua en la que cada intervención influye en decisiones futuras**, consolidando un comportamiento coherente en el tiempo (Li, 2026). El componente de **acción** se encarga de materializar las decisiones del sistema mediante la ejecución de tareas que impactan directamente en el entorno. Esta ejecución puede realizarse a través de interfaces digitales, sistemas automatizados o interacción con otros agentes, lo que convierte a la acción en el punto donde el conocimiento del sistema se transforma en efectos observables. En este sentido, la acción no es simplemente un resultado, sino una fase crítica que permite evaluar la efectividad del comportamiento del agente (Maldonado et al., 2024).

Sin embargo, la acción adquiere sentido únicamente cuando se encuentra integrada con mecanismos de **memoria**, que permiten registrar los resultados obtenidos y utilizarlos como base para futuras decisiones. La memoria no solo almacena datos, sino que estructura la experiencia del sistema, permitiendo construir una base de conocimiento acumulativo que mejora la calidad de las decisiones a lo largo del tiempo. Esta integración convierte al agente en un sistema capaz de **aprender de su propia operación** (Zhang et al., 2025). En este marco, la relación entre acción y memoria introduce una dinámica de **retroalimentación operativa**, en la cual los resultados de las acciones se incorporan al sistema como información relevante que influye en la toma de decisiones posterior. Este proceso permite al agente ajustar su comportamiento de manera progresiva, fortaleciendo estrategias exitosas y evitando aquellas que resultan ineficaces (Durga, 2025).

Asimismo, la memoria cumple un papel fundamental en la **continuidad del comportamiento**, ya que permite al sistema mantener coherencia en su acción a lo largo del tiempo. Sin mecanismos de memoria, las decisiones del agente se limitarían a la información inmediata, lo que reduciría su capacidad de adaptación y aprendizaje. En cambio, la integración de memoria permite que el agente construya una narrativa operativa que da sentido a sus acciones (Zhang et al., 2025). La acción y memoria se integran en estructuras que permiten la **evaluación de desempeño**, en las cuales el sistema analiza los resultados obtenidos y determina la efectividad de sus decisiones. Esta evaluación constituye un elemento clave para la mejora continua, ya que permite al agente identificar patrones de éxito o fracaso y ajustar su comportamiento en consecuencia (Durga, 2025).

Otro aspecto relevante es que la integración entre acción y memoria permite al sistema operar en contextos de **incertidumbre**, donde la información disponible es incompleta o cambiante. En estos escenarios, la memoria proporciona un marco de referencia que permite al agente tomar decisiones más informadas, mientras que la acción permite validar dichas decisiones mediante la interacción con el entorno (Li, 2026). En sistemas multiagente, la relación entre acción y memoria se amplía al incorporar procesos de **memoria distribuida**, donde la información se comparte entre múltiples agentes. Esta capacidad permite que los sistemas colaboren de manera más eficiente, integrando conocimiento colectivo que mejora la calidad de las decisiones y la efectividad de las acciones (Maldonado et al., 2024). La integración operativa de acción y memoria permite diferenciar entre sistemas que simplemente ejecutan tareas y aquellos que **construyen comportamiento a partir de la experiencia**, lo que constituye un rasgo distintivo de la inteligencia agéntica. Esta capacidad de aprendizaje continuo permite que los sistemas evolucionen en función de su interacción con el entorno, adaptándose a nuevas condiciones (Zhang et al., 2025).

En este sentido, la acción no puede entenderse como un evento aislado, sino como parte de un ciclo en el que cada intervención contribuye a la construcción del conocimiento del sistema. Este ciclo integra percepción, decisión, acción y memoria en una estructura que permite sostener comportamiento en entornos dinámicos, consolidando la agencia como un proceso continuo (Durga, 2025). La integración entre

acción y memoria establece que la inteligencia agéntica no se define únicamente por la capacidad de ejecutar decisiones, sino por la posibilidad de **aprender de dichas decisiones y reorganizar el comportamiento en función de la experiencia acumulada**. Esta propiedad permite comprender a los sistemas agénticos como estructuras capaces de sostener acción, aprendizaje y adaptación dentro de una misma lógica operativa, diferenciándolos claramente de sistemas tradicionales (Li, 2026).

Tipologías de agéntica

Este subtema analiza las **tipologías de agentes agénticos**, estableciendo criterios de clasificación basados en su **organización conductual**, nivel de **autonomía** y capacidad de **integración estructural**. A partir de estos ejes, se identifican distintas formas de agentes en función de su complejidad operativa y su grado de coherencia en la acción. Este enfoque permite comprender cómo varía la arquitectura de los sistemas inteligentes según su capacidad para articular comportamiento en entornos dinámicos, facilitando una lectura sistemática de la diversidad estructural presente en la inteligencia artificial agéntica y sentando bases para su análisis comparativo.

Clasificación según nivel de autonomía

La **clasificación según nivel de autonomía** constituye un eje central para comprender la evolución de los sistemas agénticos, en tanto permite diferenciar los grados en los que un agente puede **percibir, decidir y actuar sin intervención humana directa**. En este sentido, la autonomía no debe entenderse como una condición absoluta, sino como un **continuo estructural** que va desde sistemas altamente dependientes hasta arquitecturas capaces de operar de manera independiente en entornos complejos (Sapkota et al., 2026).

En el nivel más básico se encuentran los sistemas con **autonomía reactiva**, caracterizados por ejecutar acciones en respuesta directa a estímulos del entorno. Estos sistemas operan bajo esquemas de **ciclo percepción–acción**, donde las decisiones están predeterminadas por reglas o condiciones específicas, lo que limita su capacidad de adaptación. Este tipo de comportamiento fue predominante en los primeros sistemas basados en reglas y agentes clásicos, los cuales carecían de aprendizaje y capacidad de generalización (Sapkota et al., 2026).

Un segundo nivel corresponde a la **autonomía adaptativa**, en la cual los agentes incorporan mecanismos de aprendizaje que les permiten modificar su comportamiento en función de la experiencia. En este contexto, enfoques como el aprendizaje por refuerzo multiagente permiten a los sistemas **optimizar decisiones a partir de la interacción continua con el entorno**, mejorando su desempeño en escenarios dinámicos. Esta forma de autonomía introduce la capacidad de ajuste progresivo, aunque sigue limitada por el diseño de los modelos y el entorno de entrenamiento (Yadav et al., 2023).

En un nivel más avanzado se encuentra la **autonomía deliberativa**, donde el agente es capaz de **planificar, evaluar alternativas y tomar decisiones orientadas a objetivos**. A diferencia de los sistemas reactivos, estos agentes no responden únicamente a estímulos inmediatos, sino que construyen secuencias de acción basadas en metas específicas. Este tipo de autonomía implica la integración de módulos de razonamiento, planificación y evaluación, lo que permite al agente operar en entornos más complejos (Piccialli et al., 2025).

Asimismo, la evolución hacia sistemas más complejos introduce la **autonomía colaborativa**, característica de los sistemas multiagente, donde múltiples entidades interactúan para alcanzar objetivos comunes. En este nivel, la autonomía no es individual, sino distribuida, ya que los agentes deben **coordinar, comunicarse y negociar decisiones** dentro de un sistema más amplio. Este tipo de autonomía resulta clave en contextos industriales y sistemas ciberfísicos, donde la interacción entre múltiples componentes es esencial (Karnouskos et al., 2020).

El nivel más avanzado corresponde a la **autonomía agéntica o estratégica**, en la cual los sistemas integran múltiples capacidades dentro de una arquitectura coherente que les permite operar de manera prolongada sin supervisión constante. Estos sistemas son capaces de **descomponer objetivos complejos en sub-tareas, seleccionar herramientas, coordinar procesos y ajustar su comportamiento en función de resultados**, lo que representa una transición hacia sistemas verdaderamente autónomos. Este nivel se observa en la evolución reciente de la IA agéntica, donde los sistemas pasan de ser asistentes reactivos a **colaboradores proactivos** (Bandi et al., 2025).

La clasificación por niveles de autonomía permite entender que la diferencia entre sistemas no radica únicamente en su capacidad tecnológica, sino en la forma en que organizan su comportamiento. Mientras que los niveles inferiores dependen de reglas o respuestas predefinidas, los niveles superiores integran **memoria, planificación, aprendizaje y coordinación**, lo que les permite sostener acción en el tiempo y adaptarse a condiciones cambiantes (Sapkota et al., 2026).

Otro aspecto relevante es que la autonomía no es estática, sino evolutiva, ya que los sistemas pueden transitar entre niveles mediante la incorporación de nuevas capacidades. Por ejemplo, un sistema reactivo puede evolucionar hacia uno adaptativo al integrar aprendizaje, o hacia uno deliberativo al incorporar planificación. Esta progresión refleja la transformación de la inteligencia artificial hacia modelos más complejos y autónomos (Yadav et al., 2023).

Asimismo, la clasificación según nivel de autonomía tiene implicaciones directas en el diseño y evaluación de sistemas agénticos, ya que permite establecer criterios claros para determinar el grado de independencia del agente. En contextos industriales, por ejemplo, **la autonomía puede ser limitada deliberadamente para garantizar control y seguridad, mientras que en otros contextos se busca maximizar la capacidad de operación independiente** (Karnouskos et al., 2020).

Finalmente, la clasificación según nivel de autonomía establece que la IA agéntica representa una evolución progresiva desde sistemas reactivos hacia estructuras capaces de **planificar, aprender, coordinar y actuar de manera autónoma en entornos complejos**, consolidando un nuevo paradigma en el que la inteligencia se define por la capacidad de sostener acción organizada en el tiempo (Bandi et al., 2025)

Clasificación según complejidad estructural

La **clasificación según complejidad estructural** permite diferenciar los sistemas agénticos en función del grado de **organización interna, integración de componentes y nivel de interdependencia entre sus módulos**, lo que constituye un criterio fundamental para comprender la evolución de la inteligencia artificial hacia arquitecturas más sofisticadas. A diferencia de la clasificación por autonomía, este enfoque no se centra en la independencia operativa, sino en la forma en que los sistemas están diseñados y estructurados internamente (Sapkota et al., 2026).

En un nivel más básico se encuentran los sistemas de **estructura simple o monolítica**, caracterizados por integrar sus funciones dentro de un único bloque operativo. En este tipo de arquitectura, los componentes de percepción, decisión y acción se encuentran poco diferenciados, lo que limita la flexibilidad del sistema. Estos modelos suelen ser eficientes en tareas específicas, pero presentan dificultades para escalar o adaptarse a entornos complejos, debido a la falta de modularidad (Bandi et al., 2025).

Un segundo nivel corresponde a los sistemas de **estructura modular**, en los cuales el agente se organiza en componentes diferenciados que cumplen funciones específicas, como percepción, razonamiento, memoria y acción. Esta separación permite una mayor claridad en el diseño y facilita la incorporación de nuevas capacidades, ya que cada módulo puede ser modificado o mejorado sin afectar la totalidad del sistema. La modularidad representa un avance significativo en la complejidad estructural, al permitir una mayor escalabilidad (Piccialli et al., 2025).

En un nivel más avanzado se encuentran los sistemas de **estructura jerárquica**, donde los componentes se organizan en distintos niveles de control. En este tipo de arquitectura, los niveles superiores definen objetivos y estrategias, mientras que los niveles inferiores ejecutan acciones específicas. Esta organización permite gestionar sistemas complejos mediante la **descomposición de tareas en subprocesos**, lo que mejora la eficiencia y el control del comportamiento del agente (Yadav et al., 2023).

Asimismo, la evolución hacia estructuras más complejas da lugar a los sistemas de **estructura distribuida o multiagente**, en los cuales múltiples agentes interactúan dentro de un mismo sistema. En este nivel, la complejidad no solo reside en la estructura interna de cada agente, sino en la **interacción entre múltiples entidades autónomas**, lo que introduce dinámicas de coordinación, comunicación y negociación. Este tipo de arquitectura es fundamental en entornos industriales y sistemas

ciberfísicos, donde diferentes componentes deben operar de manera conjunta (Karnouskos et al., 2020).

En el nivel más avanzado se encuentran los sistemas de **estructura agéntica integrada**, caracterizados por la combinación de múltiples agentes especializados dentro de una arquitectura coordinada que permite gestionar tareas complejas de manera autónoma. En estos sistemas, la complejidad estructural se incrementa al integrar **planificación, memoria persistente, razonamiento y mecanismos de coordinación**, lo que permite al sistema operar como un ecosistema inteligente (Bandi et al., 2025).

En este marco, la complejidad estructural no se limita al número de componentes, sino a la forma en que estos se relacionan entre sí. A medida que los sistemas evolucionan, se incrementa la **interdependencia entre módulos**, lo que permite generar comportamientos más sofisticados, pero también introduce desafíos en términos de diseño, control y estabilidad del sistema (Sapkota et al., 2026).

Otro aspecto relevante es que el aumento en la complejidad estructural permite mejorar la capacidad del sistema para operar en entornos dinámicos, ya que facilita la integración de múltiples fuentes de información y la coordinación de diversas acciones. Sin embargo, esta complejidad también implica mayores requerimientos en términos de recursos computacionales y mecanismos de control, lo que debe ser considerado en el diseño de sistemas agénticos (Piccialli et al., 2025). La clasificación según complejidad estructural permite identificar una evolución clara en el desarrollo de la inteligencia artificial, desde sistemas simples y centralizados hacia arquitecturas distribuidas y altamente integradas. Esta evolución refleja la necesidad de diseñar sistemas capaces de enfrentar problemas cada vez más complejos, donde la integración de múltiples componentes resulta indispensable (Yadav et al., 2023).

Desde una perspectiva funcional, esta clasificación permite diferenciar entre sistemas que operan de manera aislada y aquellos que **organizan su comportamiento mediante estructuras complejas y coordinadas**, lo que constituye un rasgo distintivo de la IA agéntica. Esta diferenciación resulta clave para comprender cómo los sistemas evolucionan en términos de capacidad operativa y diseño estructural (Karnouskos et al., 2020). La clasificación según complejidad estructural establece que la inteligencia agéntica se desarrolla a través de un proceso progresivo en el que los sistemas incrementan su nivel de organización interna, pasando de estructuras simples a arquitecturas altamente integradas que permiten **sostener comportamiento complejo, coordinado y adaptativo en entornos dinámicos**, consolidando así la base estructural de la IA agéntica (Bandi et al., 2025).

Clasificación según organización del comportamiento

La **clasificación según organización del comportamiento** permite distinguir los sistemas agénticos en función de la manera en que estructuran sus acciones a lo largo del tiempo, lo que implica analizar no solo qué hacen, sino **cómo organizan sus procesos de acción, decisión y adaptación**. Este enfoque reconoce que el comportamiento no es homogéneo entre agentes, sino que puede variar en términos de complejidad, continuidad e integración operativa, constituyendo un criterio clave para comprender la evolución de la IA agéntica (Sapkota et al., 2026).

En un nivel más básico se encuentran los sistemas con **comportamiento reactivo**, caracterizados por una organización basada en respuestas inmediatas ante estímulos del entorno. En estos sistemas, la acción se activa directamente a partir de la percepción, sin mediación de procesos complejos de evaluación o planificación. Esta forma de organización conductual se distingue por su simplicidad y eficiencia en entornos controlados, pero presenta limitaciones en escenarios dinámicos donde se requiere adaptación (Bandi et al., 2025).

Un segundo nivel corresponde a los sistemas con **comportamiento adaptativo**, en los cuales el agente es capaz de modificar su acción en función de la experiencia. En este caso, la organización del comportamiento incorpora mecanismos de aprendizaje que permiten ajustar respuestas a lo largo del tiempo, generando una estructura más flexible. Este tipo de comportamiento introduce la capacidad de mejora continua, aunque todavía se encuentra condicionado por la estructura del modelo de aprendizaje (Yadav et al., 2023).

En un nivel más avanzado se encuentran los sistemas con **comportamiento deliberativo**, donde la acción no se limita a la reacción o adaptación, sino que se organiza a partir de procesos de evaluación, planificación y selección de alternativas. En este tipo de agentes, el comportamiento se construye mediante la **anticipación de escenarios y la definición de estrategias**, lo que permite operar en entornos complejos con mayor grado de control y coherencia (Piccialli et al., 2025).

Asimismo, la evolución del comportamiento da lugar a los sistemas con **comportamiento coordinado**, característicos de entornos multiagente, donde múltiples entidades organizan sus acciones de manera conjunta. En este nivel, la organización del comportamiento no es individual, sino colectiva, lo que implica la existencia de mecanismos de comunicación, negociación y sincronización entre agentes. Esta forma de organización resulta fundamental en sistemas industriales y distribuidos (Karnouskos et al., 2020).

De manera complementaria, los sistemas más avanzados presentan un **comportamiento agéntico integrado**, en el cual el agente es capaz de articular percepción, memoria, decisión y acción dentro de una estructura coherente que se mantiene a lo largo del tiempo. Este tipo de organización conductual se caracteriza por

la capacidad de sostener procesos complejos, gestionar múltiples tareas y adaptarse a condiciones cambiantes sin perder coherencia operativa (Bandi et al., 2025).

La organización del comportamiento puede entenderse como un continuo que va desde estructuras simples y fragmentadas hasta sistemas altamente integrados. A medida que se incrementa la complejidad, el comportamiento deja de ser una secuencia de respuestas aisladas y se convierte en un **proceso estructurado y continuo**, lo que permite al agente sostener acción en entornos dinámicos (Sapkota et al., 2026). Otro aspecto clave, es que la organización del comportamiento no depende exclusivamente de la capacidad tecnológica del sistema, sino de la forma en que se integran sus componentes. Esto implica que sistemas con capacidades similares pueden presentar comportamientos distintos en función de su organización interna, lo que resalta la importancia del diseño en la inteligencia agéntica (Piccialli et al., 2025).

Asimismo, la clasificación según organización del comportamiento permite identificar una transición clara en la evolución de la IA, desde modelos basados en respuestas simples hacia sistemas capaces de **gestionar procesos complejos, coordinados y orientados a objetivos**. Esta transición refleja el cambio hacia arquitecturas más sofisticadas, donde el comportamiento constituye el eje central del funcionamiento del sistema (Yadav et al., 2023). A nivel funcional, esta clasificación permite diferenciar entre sistemas que ejecutan acciones de manera aislada y aquellos que **organizan su comportamiento como un proceso integrado**, lo que constituye un rasgo distintivo de la IA agéntica. Esta diferenciación resulta clave para comprender cómo los sistemas evolucionan en términos de capacidad operativa y adaptabilidad (Karnouskos et al., 2020).

La clasificación según organización del comportamiento establece que la inteligencia agéntica se define por la capacidad de estructurar la acción dentro de un proceso continuo, coherente y adaptativo, lo que permite al sistema operar de manera efectiva en entornos complejos. Esta perspectiva consolida la idea de que el comportamiento no es un resultado aislado, sino una **estructura dinámica que integra percepción, decisión y acción dentro de una lógica operativa unificada** (Bandi et al., 2025).

Agentes deliberativos

El subtema de **agentes deliberativos** examina aquellas configuraciones agénticas que estructuran su comportamiento mediante procesos de **evaluación, planificación y toma de decisiones orientadas a objetivos**. A diferencia de agentes reactivos, estos sistemas organizan la acción dentro de una lógica anticipatoria que permite sostener coherencia en el tiempo. Este enfoque introduce una dimensión cognitiva funcional en la arquitectura agéntica, en la cual la acción no se limita a responder a estímulos, sino que se construye a partir de la articulación de criterios internos. En este

sentido, los agentes deliberativos representan una forma avanzada de organización del comportamiento inteligente.

Naturaleza de la deliberación en sistemas agénticos

La **naturaleza de la deliberación en sistemas agénticos** se define por la capacidad de los agentes para **estructurar procesos de decisión que implican evaluación, planificación y selección de acciones en función de objetivos**, superando esquemas de respuesta inmediata característicos de sistemas reactivos. En este sentido, la deliberación constituye un proceso interno mediante el cual el agente organiza su comportamiento a partir de la construcción de representaciones del entorno y la anticipación de posibles consecuencias de sus acciones (Sapkota et al., 2026). La deliberación introduce una diferencia fundamental respecto a modelos tradicionales, al incorporar **procesos de razonamiento orientados a objetivos**, en los cuales las decisiones no se derivan directamente de estímulos, sino de la evaluación de alternativas. Esto implica que el agente no actúa de manera automática, sino que organiza su comportamiento mediante la comparación de posibles cursos de acción, seleccionando aquellos que resultan más adecuados para alcanzar metas específicas (Bandi et al., 2025). La deliberación se sustenta en la integración de componentes como **memoria, planificación y mecanismos de evaluación**, los cuales permiten al agente construir una estructura interna que soporta la toma de decisiones. La memoria proporciona información relevante sobre experiencias previas, mientras que la planificación permite organizar acciones futuras, generando una continuidad en el comportamiento del sistema (Piccialli et al., 2025).

Por otro lado, la deliberación implica la capacidad de **evaluar múltiples escenarios antes de ejecutar una acción**, lo que introduce un proceso de decisión basado en la comparación de resultados potenciales. Este proceso permite reducir la incertidumbre y mejorar la efectividad del comportamiento del agente, al considerar las posibles consecuencias de cada alternativa antes de su implementación (Yadav et al., 2023). Otro factor clave es la **dimensión temporal de la deliberación**, ya que las decisiones no se limitan al presente inmediato, sino que consideran implicaciones futuras. Esto permite al agente estructurar su comportamiento en términos de secuencias de acción, organizando tareas en función de objetivos que pueden requerir múltiples etapas para su cumplimiento (Bandi et al., 2025). La deliberación adquiere una mayor complejidad en sistemas multiagente, donde los procesos de decisión incluyen mecanismos de **coordinación, negociación y comunicación entre múltiples entidades**. En este contexto, la deliberación no es exclusivamente individual, sino que se construye a partir de la interacción entre agentes, lo que permite generar soluciones colectivas en entornos distribuidos (Karnouskos et al., 2020). La deliberación también implica la capacidad de gestionar **contextos dinámicos y condiciones de incertidumbre**, en los cuales la información disponible puede ser incompleta o cambiante. Frente a estas condiciones, el agente debe ajustar sus decisiones en función de nueva información, manteniendo coherencia en su comportamiento a pesar de la variabilidad del entorno (Yadav et al., 2023).

Otro aspecto relevante es que la deliberación permite al agente mantener **alineación entre objetivos y acciones**, asegurando que el comportamiento se encuentre orientado hacia metas definidas. Esta alineación resulta fundamental para evitar decisiones inconsistentes o desarticuladas, garantizando que el sistema opere dentro de una lógica estructurada (Piccialli et al., 2025). La deliberación transforma la naturaleza del comportamiento del agente, pasando de esquemas basados en respuestas inmediatas a procesos en los cuales la acción se organiza como resultado de **evaluación, planificación y selección estratégica**. Esta transformación constituye un elemento central en la evolución de la inteligencia artificial hacia modelos agénticos más sofisticados (Sapkota et al., 2026).

La deliberación introduce la posibilidad de **gestionar múltiples objetivos simultáneamente**, lo que implica que el agente puede priorizar tareas y asignar recursos en función de diferentes criterios. Esta capacidad resulta clave en entornos complejos donde las decisiones deben equilibrar diversas variables y restricciones (Bandi et al., 2025). La naturaleza de la deliberación en sistemas agénticos establece que la inteligencia artificial no se define únicamente por la capacidad de procesar información, sino por la posibilidad de **organizar el comportamiento mediante procesos internos de evaluación y planificación**, lo que permite sostener acción coherente en entornos dinámicos. En este sentido, la deliberación constituye el núcleo que diferencia a los sistemas agénticos de otros modelos de inteligencia artificial, al introducir una estructura interna que permite transformar información en acción estratégica (Piccialli et al., 2025).

Planificación y evaluación de alternativas

La **planificación y evaluación de alternativas** constituye uno de los procesos centrales en los sistemas agénticos deliberativos, en tanto permite estructurar la acción mediante la **anticipación de escenarios, la generación de opciones y la selección de cursos de acción orientados a objetivos**. A diferencia de los sistemas reactivos, donde la acción se deriva directamente de estímulos, los agentes deliberativos organizan su comportamiento a partir de procesos internos que implican análisis, comparación y decisión (Sapkota et al., 2026). La planificación implica la capacidad del agente para **descomponer objetivos complejos en secuencias de acciones**, lo que permite abordar problemas de manera estructurada. Este proceso no se limita a la ejecución inmediata, sino que requiere la construcción de trayectorias que conectan el estado actual con un estado deseado, considerando restricciones, recursos disponibles y posibles contingencias (Bandi et al., 2025).

Así, la planificación se encuentra estrechamente vinculada con la **representación interna del entorno**, ya que el agente debe construir modelos que le permitan simular posibles resultados antes de actuar. Esta capacidad de simulación constituye un elemento clave para la toma de decisiones, ya que permite anticipar consecuencias y reducir la incertidumbre asociada a la acción (Piccialli et al., 2025). La evaluación de alternativas introduce un proceso mediante el cual el agente compara diferentes cursos de acción en función de criterios definidos, como eficiencia, costo, riesgo o

cumplimiento de objetivos. Este proceso implica la existencia de mecanismos de **valoración y priorización**, que permiten seleccionar la opción más adecuada entre múltiples posibilidades (Yadav et al., 2023).

La evaluación no es un proceso aislado, sino parte integral de la planificación, ya que cada alternativa generada debe ser analizada en función de sus posibles resultados. Esta integración permite al agente construir decisiones informadas, evitando acciones impulsivas o desarticuladas frente a la complejidad del entorno (Sapkota et al., 2026). La **naturaleza iterativa de la planificación**, ya que los planes no son estáticos, sino que pueden ser ajustados en función de nueva información o cambios en el entorno. Esta característica permite al agente mantener flexibilidad en su comportamiento, adaptando sus decisiones a condiciones dinámicas sin perder coherencia en su acción (Bandi et al., 2025).

La planificación y evaluación de alternativas se sustentan en la integración de múltiples componentes del sistema, como **memoria, razonamiento y percepción**, lo que permite construir decisiones más robustas. La memoria aporta experiencias previas, la percepción actualiza la información del entorno y el razonamiento permite estructurar las alternativas, generando un proceso de decisión completo (Piccialli et al., 2025). En sistemas multiagente, este proceso adquiere una mayor complejidad al incorporar mecanismos de **coordinación y negociación**, donde las alternativas no se generan de manera individual, sino como resultado de la interacción entre múltiples agentes. En este contexto, la planificación implica considerar no solo las acciones propias, sino también las de otros agentes, lo que introduce una dimensión colectiva en la toma de decisiones (Karnouskos et al., 2020).

La evaluación de alternativas debe considerar **condiciones de incertidumbre**, donde la información disponible puede ser incompleta o ambigua. En estos escenarios, los agentes utilizan mecanismos probabilísticos o heurísticos para estimar resultados y seleccionar la opción más viable, lo que permite mantener la operatividad del sistema en entornos complejos (Yadav et al., 2023). La planificación y evaluación de alternativas permiten transformar la acción en un proceso organizado, donde las decisiones se construyen a partir de **criterios explícitos y comparaciones sistemáticas**. Esta organización constituye un rasgo distintivo de la inteligencia agéntica, diferenciándola de modelos basados en respuestas directas (Sapkota et al., 2026).

Otro elemento importante, es la capacidad del agente para **priorizar objetivos y gestionar recursos**, lo que implica que la planificación no solo considera qué hacer, sino también cómo hacerlo de manera eficiente. Esta capacidad resulta fundamental en entornos donde los recursos son limitados y las decisiones deben optimizar resultados (Bandi et al., 2025). La planificación y evaluación de alternativas establecen que la inteligencia agéntica se caracteriza por la capacidad de **organizar la acción mediante procesos internos de anticipación, comparación y selección**, lo que permite sostener comportamiento coherente en entornos dinámicos. Este proceso

consolida la deliberación como un elemento central en la arquitectura del agente, permitiendo transformar información en acción estratégica (Piccialli et al., 2025).

Deliberación y coherencia del comportamiento

La **deliberación y coherencia del comportamiento** constituyen un eje fundamental en la inteligencia agéntica, en tanto permiten comprender cómo los sistemas organizan sus acciones de manera consistente a lo largo del tiempo. En este sentido, la deliberación no solo implica la capacidad de tomar decisiones complejas, sino también la de **mantener una estructura interna que garantice continuidad y consistencia en la acción**, evitando respuestas fragmentadas o contradictorias (Sapkota et al., 2026). La coherencia del comportamiento se deriva de la capacidad del agente para **integrar decisiones dentro de una lógica operativa unificada**, donde cada acción se encuentra alineada con objetivos previamente definidos. Esta integración permite que el comportamiento no sea una secuencia de eventos aislados, sino un proceso estructurado en el que las decisiones mantienen una dirección clara (Bandi et al., 2025). La deliberación actúa como el mecanismo que permite construir dicha coherencia, al introducir procesos de **evaluación, planificación y selección de alternativas** que organizan la acción del agente. A través de estos procesos, el sistema puede asegurar que sus decisiones no sean arbitrarias, sino que respondan a criterios definidos, lo que contribuye a la estabilidad del comportamiento (Piccialli et al., 2025). La coherencia del comportamiento se encuentra estrechamente vinculada con la **capacidad de anticipación**, ya que el agente debe considerar las implicaciones futuras de sus acciones para evitar inconsistencias. Esta anticipación permite al sistema mantener continuidad en su comportamiento, alineando decisiones presentes con objetivos futuros (Yadav et al., 2023).

Otro aspecto a considerar y clave, es la relación entre deliberación y **temporalidad del comportamiento**, en la medida en que la coherencia se construye a lo largo del tiempo. Esto implica que las decisiones no pueden evaluarse de manera aislada, sino como parte de una secuencia en la que cada acción influye en las siguientes. En este sentido, la deliberación permite estructurar el comportamiento como un proceso continuo (Bandi et al., 2025). La coherencia del comportamiento depende de la capacidad del agente para mantener **consistencia interna en sus procesos de decisión**, lo que implica evitar contradicciones entre acciones o cambios abruptos en la estrategia. Esta consistencia resulta fundamental para garantizar que el sistema opere de manera confiable en entornos dinámicos (Sapkota et al., 2026).

En sistemas multiagente, la coherencia adquiere una dimensión adicional al requerir **alineación entre múltiples agentes**, lo que implica que las decisiones individuales deben integrarse dentro de un comportamiento colectivo. Este proceso requiere mecanismos de coordinación y comunicación que permitan mantener coherencia a nivel del sistema completo (Karnouskos et al., 2020). La deliberación permite gestionar **conflictos entre objetivos o alternativas**, lo que resulta clave para mantener coherencia en contextos complejos. Cuando el agente enfrenta múltiples opciones, la evaluación estructurada de alternativas permite seleccionar aquellas que mejor se

alinean con los objetivos del sistema, evitando decisiones contradictorias (Piccialli et al., 2025).

La capacidad del agente para **adaptar su comportamiento sin perder coherencia**, lo que implica que los cambios en la estrategia deben integrarse dentro de la lógica operativa del sistema. Esta capacidad permite responder a nuevas condiciones sin generar rupturas en el comportamiento, manteniendo estabilidad en la acción (Yadav et al., 2023). La coherencia del comportamiento representa un indicador clave de la inteligencia agéntica, ya que permite diferenciar entre sistemas que responden de manera desorganizada y aquellos que **estructuran su acción mediante procesos deliberativos consistentes**. Esta diferenciación resulta fundamental para comprender el valor de la deliberación en la organización del comportamiento (Sapkota et al., 2026). La coherencia no se limita a la alineación con objetivos, sino que también implica la capacidad de **gestionar múltiples procesos simultáneamente**, manteniendo consistencia entre ellos. Esto resulta particularmente relevante en sistemas complejos, donde el agente debe equilibrar diferentes tareas sin comprometer la estabilidad del comportamiento (Bandi et al., 2025). La deliberación y coherencia del comportamiento establecen que la inteligencia agéntica se define por la capacidad de **organizar la acción de manera continua, consistente y orientada a objetivos**, lo que permite al agente operar de manera efectiva en entornos dinámicos. En este sentido, la coherencia constituye el resultado de procesos deliberativos bien estructurados, consolidando la base del comportamiento agéntico (Piccialli et al., 2025).

Configuraciones arquitectónicas de la IA agéntica

El subtema de **configuraciones arquitectónicas de la IA agéntica** analiza las distintas formas en que los componentes del agente se estructuran para sostener comportamiento coherente. Este enfoque permite comprender cómo la organización interna del sistema determina su capacidad de integración, adaptación y continuidad operativa. A diferencia de enfoques centrados en funciones aisladas, las configuraciones arquitectónicas enfatizan la articulación estructural entre percepción, decisión, acción y memoria. En este sentido, el análisis arquitectónico no solo describe la composición del sistema, sino que explica cómo su diseño influye directamente en la forma en que se organiza el comportamiento dentro de la inteligencia artificial agéntica.

Arquitecturas modulares y su integración funcional

Las **arquitecturas modulares en sistemas agénticos** representan un principio fundamental de diseño orientado a organizar la inteligencia artificial mediante la **separación funcional de sus componentes**, permitiendo que cada módulo cumpla una función específica dentro del sistema. Esta estructura posibilita distinguir claramente entre procesos como la percepción, el razonamiento, la memoria y la acción, facilitando la comprensión y el control del comportamiento del agente. En este

sentido, la modularidad no es únicamente una estrategia técnica, sino una forma de estructurar la complejidad del sistema en unidades manejables y especializadas (Durga, 2025). La modularidad implica que los distintos componentes del agente operan de manera diferenciada, pero dentro de una arquitectura coordinada que permite su integración. Esta organización favorece la **escalabilidad del sistema**, ya que los módulos pueden ser modificados, reemplazados o ampliados sin afectar el funcionamiento global. Así, los sistemas agénticos pueden evolucionar progresivamente, incorporando nuevas capacidades sin comprometer su estabilidad operativa (Li, 2026).

Asimismo, la modularidad permite la **especialización funcional**, donde cada módulo se optimiza para realizar tareas específicas. Por ejemplo, el módulo de percepción puede enfocarse en la captura y procesamiento inicial de datos, mientras que el módulo de decisión se encarga de evaluar alternativas y seleccionar acciones. Esta especialización incrementa la eficiencia del sistema, ya que cada componente opera bajo condiciones diseñadas para su función particular (Durga, 2025). La **integración funcional** se convierte en el elemento que articula los distintos módulos dentro de una estructura coherente. A diferencia de una simple agregación de componentes, la integración implica la existencia de **flujos de información organizados**, donde los datos circulan entre módulos de manera estructurada, permitiendo que el sistema opere como una unidad funcional. Esta interconexión es esencial para garantizar que las decisiones del agente se basen en información completa y actualizada (Xie et al., 2025).

De manera complementaria, la integración funcional establece relaciones de dependencia entre los módulos, donde el funcionamiento de cada componente se encuentra vinculado al de los demás. Por ejemplo, la información generada por la percepción alimenta el proceso de decisión, mientras que los resultados de la acción retroalimentan la memoria del sistema. Este proceso configura un **ciclo operativo continuo**, en el cual cada módulo contribuye al comportamiento global del agente (Li, 2026).

En sistemas más avanzados, la integración funcional permite la coordinación simultánea de múltiples módulos, lo que posibilita la gestión de tareas complejas que requieren la interacción de diferentes capacidades. Esta coordinación resulta clave en entornos dinámicos, donde el agente debe procesar información, tomar decisiones y ejecutar acciones de manera concurrente, manteniendo coherencia en su comportamiento (Xie et al., 2025).

Por otro lado, en el contexto de sistemas multiagente, la modularidad se extiende hacia una **distribución funcional entre múltiples agentes**, donde cada entidad actúa como un módulo dentro de un sistema más amplio. En este caso, la integración funcional no solo ocurre dentro de un agente, sino también entre distintos agentes, lo que permite construir sistemas distribuidos capaces de abordar problemas complejos mediante la colaboración (Maldonado et al., 2024).

Esta distribución introduce una mayor complejidad en la arquitectura del sistema, ya que la integración debe considerar tanto la comunicación interna como la **coordinación externa entre agentes**. En este sentido, la arquitectura modular se convierte en la base para el desarrollo de sistemas escalables y adaptativos, capaces de operar en entornos altamente dinámicos (Maldonado et al., 2024).

Un elemento central en la integración funcional es el papel de la **memoria como componente articulador**, ya que permite conectar las distintas fases del proceso agéntico. La memoria no solo almacena información, sino que facilita la continuidad del comportamiento, permitiendo que los módulos operen sobre una base de conocimiento común. Esto resulta fundamental para garantizar coherencia en la acción del sistema y para sostener procesos de aprendizaje (Zhang et al., 2025).

Desde una perspectiva de diseño, las arquitecturas modulares permiten la **reconfiguración dinámica del sistema**, lo que implica que los módulos pueden reorganizarse en función de nuevas condiciones o requerimientos. Esta capacidad resulta esencial en entornos cambiantes, donde el agente debe adaptarse sin perder estabilidad operativa. La modularidad, en este sentido, no solo facilita la construcción del sistema, sino también su evolución (Durga, 2025).

Asimismo, la modularidad favorece la **interoperabilidad entre sistemas**, ya que los módulos pueden diseñarse para interactuar con otros sistemas o plataformas. Esto amplía el alcance del agente, permitiéndole integrarse en ecosistemas tecnológicos más amplios, lo que resulta clave en contextos industriales y organizacionales donde múltiples sistemas deben operar de manera conjunta (Xie et al., 2025).

Las arquitecturas modulares y su integración funcional permiten comprender que la inteligencia agéntica no depende únicamente de la capacidad de procesamiento, sino de la forma en que los componentes del sistema se organizan e interactúan. En este sentido, la modularidad y la integración constituyen los pilares que permiten construir sistemas capaces de **operar de manera coherente, adaptativa y escalable en entornos complejos**, consolidando así la base arquitectónica de la IA agéntica (Li, 2026).

Arquitecturas integradas y coherencia sistémica

Las **arquitecturas integradas en sistemas agénticos** representan una evolución respecto a los enfoques modulares, al centrarse en la **articulación total de los componentes dentro de una unidad funcional coherente**. Mientras que la modularidad permite organizar la complejidad mediante la separación de funciones, la integración busca **garantizar que dichas funciones operen de manera interdependiente**, configurando un sistema en el que el comportamiento emerge de la coordinación entre sus partes (Li, 2026). La integración no implica la eliminación de los módulos, sino su incorporación dentro de una estructura donde las fronteras funcionales se vuelven dinámicas. Esto significa que los procesos de percepción, decisión, acción y memoria no operan de manera aislada, sino que se encuentran

sincronizados dentro de un flujo continuo de información, lo que permite al sistema sostener coherencia en su comportamiento (Durga, 2025).

Así, la **coherencia sistémica** se configura como una propiedad emergente de la arquitectura integrada, en tanto el comportamiento del agente no depende de un componente específico, sino de la **interacción coordinada entre todos los elementos del sistema**. Esta coherencia permite que las acciones del agente mantengan consistencia a lo largo del tiempo, evitando fragmentación o contradicción en su operación (Li, 2026).

Asimismo, la integración funcional implica la existencia de **interdependencia estructural entre componentes**, donde cada módulo depende de los demás para operar correctamente. Por ejemplo, la percepción no solo alimenta la decisión, sino que se ajusta en función de la memoria y de los resultados de la acción, generando un sistema en el que los procesos se retroalimentan continuamente (Xie et al., 2025). En este sentido, la arquitectura integrada transforma la lógica del sistema, pasando de una organización basada en componentes independientes a una estructura en la que el comportamiento se define por la **relación entre los componentes**. Esta transformación resulta clave para comprender la evolución de la inteligencia artificial hacia modelos agénticos más complejos (Durga, 2025).

Otro aspecto fundamental es la **continuidad operativa del sistema**, ya que la integración permite que el agente mantenga una línea de acción coherente a lo largo del tiempo. Esta continuidad se logra mediante la sincronización de los procesos internos, lo que permite al sistema responder a cambios en el entorno sin perder estabilidad en su comportamiento (Li, 2026). En sistemas multiagente, la coherencia sistémica adquiere una dimensión ampliada, al requerir la integración de múltiples agentes dentro de una estructura coordinada. En este contexto, la arquitectura integrada no solo implica la coherencia interna de cada agente, sino también la **coherencia colectiva del sistema**, donde las acciones individuales contribuyen a un comportamiento global consistente (Maldonado et al., 2024).

Esta dimensión colectiva introduce la necesidad de mecanismos de **coordinación, comunicación y sincronización**, que permitan mantener la coherencia del sistema a pesar de la diversidad de agentes. La integración, en este sentido, se convierte en un proceso distribuido que articula múltiples entidades dentro de una misma lógica operativa (Maldonado et al., 2024).

Un elemento central en la coherencia sistémica es el papel de la **memoria como eje integrador**, ya que permite conectar los distintos procesos del sistema a lo largo del tiempo. La memoria no solo almacena información, sino que actúa como un mecanismo que asegura la continuidad del comportamiento, permitiendo que las decisiones actuales se fundamenten en experiencias previas (Zhang et al., 2025). Esta función integradora de la memoria resulta esencial para evitar inconsistencias en el comportamiento del agente, ya que permite mantener una referencia común para todos los módulos. De este modo, la memoria contribuye a la construcción de una

coherencia temporal, donde las acciones del sistema se articulan dentro de una secuencia lógica (Zhang et al., 2025).

Desde una perspectiva de diseño, las arquitecturas integradas permiten construir sistemas capaces de operar en entornos complejos mediante la **sincronización de múltiples procesos simultáneos**. Esta capacidad resulta fundamental en contextos donde el agente debe gestionar información diversa, tomar decisiones rápidas y ejecutar acciones coordinadas, todo ello sin perder coherencia en su comportamiento (Xie et al., 2025).Se destaca que, laintegración favorece la **adaptabilidad del sistema**, ya que permite ajustar el comportamiento en función de nuevas condiciones sin generar rupturas en la operación. Esta adaptabilidad no se basa en la modificación de un módulo específico, sino en la reorganización del sistema como un todo, lo que refleja un nivel más avanzado de inteligencia agéntica (Durga, 2025).

Las arquitecturas integradas y la coherencia sistémica permiten comprender que la inteligencia agéntica no se define por la suma de capacidades individuales, sino por la forma en que estas se articulan dentro de una estructura coherente. En este sentido, la integración constituye el principio que transforma la arquitectura del agente en un sistema capaz de **sostener comportamiento continuo, consistente y adaptativo en entornos dinámicos**, consolidando así el núcleo operativo de la IA agéntica (Li, 2026).

Arquitecturas híbridas como transición estructural

Las **arquitecturas híbridas** en sistemas agénticos representan una evolución estructural que surge de la necesidad de integrar múltiples capacidades dentro de un mismo sistema operativo, particularmente en contextos donde la autonomía y la adaptabilidad son esenciales. A diferencia de modelos tradicionales, estos sistemas no se limitan a estructuras rígidas, sino que combinan componentes como planificación, memoria, razonamiento y ejecución dentro de un marco dinámico que permite la toma de decisiones autónomas. Esta integración responde a la necesidad de operar en entornos cambiantes, donde los agentes deben aprender, actuar y ajustarse continuamente (**Durga, 2025**).

Las arquitecturas híbridas no solo combinan elementos estructurales, sino que establecen un **modelo de procesamiento orientado a tareas**, donde el agente recibe información multimodal, la procesa mediante modelos de lenguaje y ejecuta acciones apoyadas en herramientas externas. Este enfoque implica que la arquitectura no es estática, sino que se configura en función del flujo de información, integrando memoria, razonamiento y acción como procesos interdependientes. Así, la hibridez se manifiesta como una **estructura orientada a tareas y contexto**, lo que permite una mayor flexibilidad operativa (**Li, 2026**). En términos estructurales, la transición hacia arquitecturas híbridas implica superar la fragmentación funcional de los modelos modulares, integrando procesos que antes operaban de manera separada. La evidencia empírica muestra que los sistemas basados en inteligencia artificial requieren la integración de generación de soluciones, evaluación y verificación dentro de un mismo flujo operativo, especialmente en contextos complejos como el diseño

estructural. Esta integración permite que los sistemas no solo generen resultados, sino que los validen y ajusten en tiempo real, lo que constituye un avance hacia una **arquitectura operativa continua y adaptativa** (**Xie et al., 2025**).

Desde el punto de vista funcional, uno de los elementos clave de las arquitecturas híbridas es la incorporación de la **memoria como mecanismo estructural central**, ya que permite almacenar, recuperar y reutilizar información en procesos de decisión. La memoria no solo cumple una función de almacenamiento, sino que actúa como un sistema que conecta experiencias pasadas con decisiones futuras, permitiendo que el agente evolucione en su interacción con el entorno. En este sentido, la arquitectura híbrida no solo integra componentes, sino que articula temporalmente el comportamiento del agente (**Zhang et al., 2025**).

En consecuencia, la arquitectura híbrida introduce una lógica de funcionamiento basada en ciclos de interacción entre percepción, decisión y acción, donde cada proceso se retroalimenta mediante mecanismos de aprendizaje y evaluación. Esta dinámica permite que el sistema no solo ejecute tareas, sino que reflexione sobre su desempeño y ajuste sus estrategias en consecuencia. Así, la hibridez no implica únicamente diversidad estructural, sino la capacidad de operar bajo un esquema de **retroalimentación continua y autorregulación** (**Durga, 2025**).

Sistémicamente ajblando, la flexibilidad estructural que caracteriza a estas arquitecturas se manifiesta en la capacidad de integrar múltiples fuentes de información y adaptarse a distintos contextos operativos. Esto implica que el sistema puede reorganizar sus procesos internos en función de las condiciones del entorno, lo que le permite mantener coherencia en su comportamiento incluso en situaciones de incertidumbre. En este sentido, la arquitectura híbrida constituye un modelo que combina estabilidad estructural con capacidad de cambio, consolidando una lógica de **adaptabilidad contextual del sistema** (**Li, 2026**).

Asimismo, la integración de procesos en arquitecturas híbridas permite abordar problemas complejos mediante la coordinación de múltiples funciones dentro de un mismo sistema. En contextos como el diseño estructural automatizado, la combinación de generación de soluciones y verificación normativa demuestra que la eficiencia del sistema depende de la interacción entre sus componentes, más que de su funcionamiento aislado. Esto refuerza la idea de que la inteligencia agéntica no reside en módulos individuales, sino en la **integración funcional de procesos heterogéneos** (**Xie et al., 2025**). Las arquitecturas híbridas representan un cambio conceptual en el diseño de sistemas de inteligencia artificial, al introducir la capacidad de evolución estructural como una propiedad fundamental. A través de la memoria, la interacción con el entorno y los mecanismos de retroalimentación, estos sistemas pueden modificar su comportamiento y estructura en función de la experiencia acumulada. En consecuencia, la hibridez no es únicamente una característica técnica, sino el fundamento de una **inteligencia artificial capaz de adaptarse, aprender y evolucionar de manera autónoma** (**Zhang et al., 2025**).

Arquitecturas emergentes

El subtema de **arquitecturas emergentes** examina las configuraciones avanzadas de la IA agéntica caracterizadas por la **distribución, interacción y coevolución de múltiples agentes**. Estas arquitecturas superan los modelos centrados en agentes individuales, introduciendo dinámicas colectivas que permiten sostener comportamiento organizado a gran escala. En este sentido, los sistemas multiagente, distribuidos y ecosistémicos representan una expansión del concepto de agencia hacia formas relacionales y dinámicas. Este enfoque permite comprender cómo la inteligencia artificial evoluciona hacia estructuras en las que la coordinación, la interacción y la adaptación conjunta se convierten en elementos centrales para sostener coherencia en entornos complejos.

Sistemas multiagente y organización colectiva del comportamiento

Las **arquitecturas multiagente** constituyen una base fundamental para comprender la **organización colectiva del comportamiento** en sistemas agénticos, al permitir la interacción coordinada de múltiples entidades autónomas orientadas a un objetivo común. En estos sistemas, cada agente posee capacidades de percepción, decisión y acción, pero su funcionamiento adquiere sentido pleno únicamente cuando se integra dentro de una dinámica colectiva. Así, la inteligencia no se reduce al desempeño individual, sino que emerge de la interacción entre agentes, configurando una **inteligencia colectiva basada en la cooperación estructurada** (Maldonado et al., 2024). Desde esta perspectiva, los sistemas multiagente operan bajo un modelo de **organización distribuida**, en el cual no existe un control central que determine todas las decisiones. En su lugar, cada agente actúa con cierto grado de autonomía, pero ajusta su comportamiento en función de la información que recibe del entorno y de otros agentes. Este principio permite que el sistema sea más flexible y escalable, ya que puede adaptarse a cambios sin depender de una estructura rígida. En consecuencia, la coordinación se convierte en un proceso dinámico que articula múltiples decisiones individuales hacia un resultado global coherente (Durga, 2025).

Bajo este enfoque, la **organización colectiva del comportamiento** depende de los mecanismos de interacción que regulan la relación entre los agentes. Estos mecanismos incluyen procesos de comunicación, intercambio de información y coordinación de acciones, los cuales permiten que los agentes alineen sus objetivos individuales con el objetivo global del sistema. En este sentido, la coordinación no es un elemento accesorio, sino el núcleo que permite transformar comportamientos aislados en una acción colectiva coherente, consolidando una lógica de **interdependencia funcional entre agentes** (Maldonado et al., 2024).

Desde el punto de vista operativo, los sistemas multiagente incorporan mecanismos como algoritmos de consenso, asignación de tareas y formación de coaliciones, los cuales permiten gestionar la interacción entre agentes en entornos complejos. Estos

mecanismos facilitan la resolución de conflictos y la sincronización de acciones, lo que resulta esencial para mantener la coherencia del sistema. Así, la toma de decisiones deja de ser un proceso individual y se convierte en una **negociación distribuida**, donde cada agente contribuye al resultado final del sistema (Maldonado et al., 2024).

Los sistemas multiagente contemporáneos integran procesos de percepción, razonamiento y acción dentro de un marco colectivo que permite la construcción conjunta de soluciones. Esta integración implica que los agentes no solo reaccionan al entorno, sino que participan activamente en la generación de estrategias mediante procesos colaborativos. En consecuencia, la organización colectiva se configura como una propiedad emergente que surge de la interacción entre múltiples componentes autónomos, consolidando una **arquitectura distribuida orientada a la cooperación** (Li, 2026).

Un componente clave en esta dinámica es la **memoria en sistemas multiagente**, ya que permite a los agentes almacenar información sobre interacciones previas y utilizarla para mejorar su desempeño futuro. La memoria facilita la coordinación al proporcionar un contexto compartido que guía la toma de decisiones, lo que resulta especialmente relevante en entornos dinámicos. En este sentido, la memoria no solo cumple una función de almacenamiento, sino que actúa como un mecanismo que articula la continuidad del comportamiento colectivo, fortaleciendo la coherencia del sistema (Zhang et al., 2025).

Adaptativamente hablando, la organización colectiva del comportamiento implica la capacidad del sistema para ajustarse a condiciones cambiantes mediante procesos de aprendizaje y retroalimentación. Los sistemas multiagente modernos integran mecanismos que permiten evaluar el desempeño colectivo y modificar las estrategias de los agentes en función de los resultados obtenidos. Este proceso introduce una dimensión evolutiva en la organización del sistema, donde la inteligencia colectiva se construye a través de la **autoorganización adaptativa** y la mejora continua del comportamiento (Durga, 2025). La capacidad de los agentes para interactuar de manera eficiente permite abordar problemas complejos en diversos dominios, donde la coordinación es esencial para alcanzar objetivos globales. En contextos como redes inteligentes o sistemas distribuidos, la interacción entre múltiples agentes permite optimizar recursos y mejorar la eficiencia del sistema. Esto refuerza la idea de que la inteligencia agéntica no reside en componentes aislados, sino en la **integración funcional de múltiples agentes en un sistema coherente** (Maldonado et al., 2024).

Se debe considerar que, la organización colectiva del comportamiento en sistemas multiagente representa un cambio conceptual en el diseño de la inteligencia artificial, al desplazar el foco desde la capacidad individual hacia la interacción sistémica. Este enfoque permite comprender que la inteligencia emerge de la coordinación entre agentes, lo que redefine la manera en que se diseñan y analizan los sistemas agénticos. En consecuencia, los sistemas multiagente constituyen el fundamento de una **inteligencia distribuida capaz de adaptarse, coordinarse y evolucionar en entornos complejos** (Li, 2026).

Arquitecturas distribuidas y descentralización de la agencia

Las **arquitecturas distribuidas** constituyen el fundamento estructural de la **descentralización de la agencia** en sistemas agénticos, al permitir que múltiples entidades autónomas operen de manera coordinada sin depender de un núcleo central de control. A diferencia de los modelos centralizados, donde la inteligencia se concentra en una unidad dominante, las arquitecturas distribuidas redistribuyen las capacidades de percepción, decisión y acción entre múltiples agentes, configurando un sistema donde la agencia se fragmenta y se reconstruye a nivel colectivo. En este sentido, la agencia deja de ser una propiedad individual para convertirse en una **propiedad emergente del sistema distribuido** (Maldonado et al., 2024).

Desde esta perspectiva, la descentralización no implica ausencia de control, sino la transformación del control en un proceso distribuido que se ejerce a través de la interacción entre agentes. Mientras que los sistemas centralizados operan mediante jerarquías rígidas, los sistemas distribuidos funcionan mediante mecanismos de coordinación horizontal, donde cada agente contribuye a la toma de decisiones a partir de información local. Este cambio introduce una lógica en la cual la coherencia del sistema depende de la **capacidad de coordinación distribuida**, más que de la imposición de reglas desde un nivel superior (Durga, 2025).

A diferencia de esta concepción dinámica de la coordinación, los enfoques basados en frameworks de agentes subrayan la importancia de diseñar estructuras que permitan sostener dicha descentralización sin perder coherencia operativa. En este contexto, los sistemas distribuidos se configuran como redes funcionales donde cada agente ejecuta tareas específicas, pero mantiene la capacidad de interactuar con otros agentes para construir soluciones globales. Así, la descentralización no es simplemente una característica estructural, sino una condición que requiere un **framework organizativo capaz de articular múltiples niveles de interacción** (Li, 2026).

En términos operativos, la descentralización de la agencia exige la implementación de mecanismos que permitan alinear decisiones individuales sin recurrir a un controlador central. Entre estos mecanismos destacan los protocolos de comunicación, los algoritmos de consenso y la coordinación basada en objetivos compartidos, los cuales permiten que los agentes ajusten su comportamiento en función de la información disponible. En contraste con los modelos jerárquicos, donde las decisiones se imponen de manera vertical, los sistemas distribuidos operan mediante una **negociación distribuida del comportamiento**, en la que el resultado emerge de múltiples interacciones locales (Maldonado et al., 2024).

Sin embargo, la descentralización introduce desafíos relacionados con la coherencia del sistema, ya que la ausencia de un control central puede generar inconsistencias en el comportamiento colectivo. En este punto, la incorporación de mecanismos de memoria resulta fundamental, ya que permite a los agentes compartir información sobre interacciones previas y construir una base común de conocimiento.

A diferencia de los sistemas sin memoria, donde las decisiones son independientes, las arquitecturas que integran memoria permiten sostener una **coherencia distribuida basada en experiencia acumulada** (Zhang et al., 2025). Comparativamente hablando, mientras que la coordinación distribuida explica cómo los agentes interactúan en tiempo real, la memoria permite comprender cómo el sistema mantiene continuidad a lo largo del tiempo. Esta distinción introduce una dimensión temporal en la descentralización de la agencia, donde el comportamiento no depende únicamente del estado actual del sistema, sino de su historia de interacciones. En consecuencia, la agencia distribuida se configura como un proceso dinámico que combina interacción presente y aprendizaje acumulado, consolidando una lógica de **continuidad operativa en sistemas descentralizados** (Durga, 2025).

La descentralización de la agencia implica una capacidad adaptativa que permite al sistema reorganizarse en función de cambios en el entorno. Esta adaptabilidad no se limita a la modificación de parámetros individuales, sino que involucra la reconfiguración de las relaciones entre agentes, lo que permite mantener la coherencia del sistema incluso en condiciones de incertidumbre. En este sentido, las arquitecturas distribuidas constituyen una base para la **autoorganización adaptativa**, donde el sistema ajusta su estructura sin intervención externa (Li, 2026).

la descentralización de la agencia redefine el concepto mismo de inteligencia en sistemas artificiales, al desplazar el foco desde la capacidad individual hacia la interacción sistémica. En lugar de entender la inteligencia como una propiedad localizada, las arquitecturas distribuidas permiten concebirla como una propiedad que emerge de la coordinación entre múltiples agentes. Así, la inteligencia agéntica se configura como una **inteligencia distribuida capaz de coordinarse, aprender y evolucionar sin control centralizado**, consolidando un paradigma que trasciende los modelos tradicionales de inteligencia artificial (Zhang et al., 2025).

Ecosistemas agénticos y coevolución del comportamiento

Los **ecosistemas agénticos** representan una expansión conceptual de los sistemas multiagente al incorporar no solo la interacción entre agentes, sino también la evolución conjunta de sus comportamientos en función del entorno y de las dinámicas internas del sistema. A diferencia de configuraciones estáticas, donde los agentes operan bajo reglas predefinidas, los ecosistemas agénticos introducen una lógica en la que los agentes modifican sus estrategias a partir de la interacción continua con otros agentes y con el contexto. En este sentido, la inteligencia no se limita a la coordinación, sino que se transforma en un proceso de **coevolución del comportamiento dentro de sistemas interdependientes** (Maldonado et al., 2024).

Desde esta perspectiva, la coevolución implica que los cambios en el comportamiento de un agente afectan directamente a los demás, generando una dinámica de ajuste mutuo que redefine constantemente la estructura del sistema. Mientras que en los sistemas tradicionales la adaptación ocurre de manera individual, en los ecosistemas agénticos la adaptación es relacional, lo que significa que los

agentes evolucionan en función de sus interacciones. Esta condición introduce una lógica donde el comportamiento colectivo no es resultado de decisiones aisladas, sino de un proceso continuo de **adaptación interdependiente entre agentes** (Durga, 2025). A diferencia de esta dinámica relacional, los frameworks de agentes permiten comprender cómo estas interacciones se estructuran dentro de un sistema, proporcionando una base organizativa que sostiene la coevolución. En este contexto, los ecosistemas agénticos se configuran como redes donde los agentes no solo ejecutan tareas, sino que participan en la construcción de estrategias colectivas mediante procesos de intercambio de información. Así, la coevolución no ocurre de manera espontánea, sino que se encuentra mediada por un **framework que articula la interacción y el aprendizaje colectivo** (Li, 2026).

Bajo el esquema operativos, la coevolución del comportamiento se manifiesta a través de mecanismos de retroalimentación continua, donde los agentes ajustan sus acciones en función de los resultados obtenidos y de las respuestas de otros agentes. Estos mecanismos permiten que el sistema evolucione de manera progresiva, generando patrones de comportamiento cada vez más complejos. En contraste con sistemas estáticos, donde las reglas permanecen constantes, los ecosistemas agénticos operan bajo una lógica de **retroalimentación dinámica que impulsa la evolución del sistema** (Maldonado et al., 2024).

La coevolución no puede sostenerse sin la existencia de mecanismos que permitan conservar información sobre interacciones pasadas, lo que introduce el papel de la memoria en estos sistemas. La memoria permite que los agentes no solo respondan a estímulos presentes, sino que incorporen experiencias previas en su proceso de toma de decisiones, lo que favorece la continuidad del comportamiento colectivo. En este sentido, la coevolución se apoya en una **memoria distribuida que articula el aprendizaje y la adaptación en el tiempo** (Zhang et al., 2025).

Comparativamente, mientras que la adaptación individual permite a los agentes responder a cambios inmediatos, la coevolución introduce una dimensión más compleja en la que los agentes transforman sus estrategias en función de la evolución del sistema en su conjunto. Esta diferencia implica que los ecosistemas agénticos no solo reaccionan al entorno, sino que lo co-construyen a través de sus interacciones. En consecuencia, la inteligencia del sistema se configura como un proceso dinámico de **evolución conjunta entre agentes y entorno** (Durga, 2025). La coevolución del comportamiento implica una capacidad de autoorganización que permite al sistema generar estructuras emergentes sin intervención externa. Esta autoorganización se manifiesta en la formación de patrones colectivos, como la especialización de agentes o la distribución de tareas, los cuales surgen a partir de la interacción continua. En este sentido, los ecosistemas agénticos representan una forma avanzada de organización donde la estructura del sistema no está predefinida, sino que emerge de la dinámica de interacción, consolidando una lógica de **autoorganización evolutiva** (Li, 2026).

Los ecosistemas agénticos redefinen el concepto de inteligencia artificial al introducir la coevolución como un elemento central del comportamiento. En lugar de

concebir la inteligencia como una capacidad estática, estos sistemas permiten entenderla como un proceso en constante transformación, donde los agentes aprenden, se adaptan y evolucionan conjuntamente. Así, la inteligencia agéntica se configura como una **propiedad emergente de sistemas que coevolucionan a través de la interacción, la memoria y la adaptación continua**, estableciendo un paradigma que trasciende los modelos tradicionales de inteligencia artificial (Zhang et al., 2025).

Conclusiones

El análisis desarrollado en este capítulo permite afirmar que la inteligencia artificial agéntica no puede comprenderse únicamente desde una perspectiva conceptual, sino que exige una aproximación estructural que explique cómo se organiza internamente el comportamiento inteligente. En este sentido, la arquitectura se consolida como el eje central que articula la capacidad del sistema para integrar percepción, decisión, acción y memoria dentro de una lógica operativa coherente. **La inteligencia agéntica emerge, por tanto, no como resultado de componentes aislados, sino de su integración estructural y funcional dentro de un sistema organizado** .

Uno de los principales aportes del capítulo radica en demostrar que los componentes del agente no operan de manera independiente, sino como una **estructura interdependiente que sostiene la continuidad, adaptabilidad y orientación del comportamiento**. La percepción establece los límites de lo que el sistema puede conocer, la decisión transforma la información en acción potencial, y la acción, integrada con la memoria, permite cerrar el ciclo operativo mediante aprendizaje y retroalimentación. **Esta integración configura un sistema dinámico en el que cada componente depende de los demás para mantener coherencia operativa**.

Asimismo, el capítulo evidencia que la toma de decisión estructurada constituye el núcleo operativo del agente, ya que permite transformar información en comportamiento orientado a objetivos. **A diferencia de los sistemas tradicionales basados en reglas, los sistemas agénticos incorporan procesos de evaluación, anticipación y selección de alternativas**, lo que les permite operar en entornos complejos e inciertos. Esta capacidad introduce una lógica de comportamiento no lineal, donde las decisiones se ajustan continuamente en función del contexto.

En relación con la acción y la memoria, se establece que la inteligencia agéntica se sostiene en un proceso de retroalimentación continua. **La memoria no solo almacena información, sino que estructura la experiencia del sistema, permitiendo aprendizaje y adaptación progresiva**, mientras que la acción valida las decisiones mediante su impacto en el entorno. Esta relación convierte al agente en un sistema evolutivo capaz de reorganizar su comportamiento a partir de la experiencia acumulada.

Por otro lado, el análisis de las tipologías de agentes permite identificar que la inteligencia agéntica se desarrolla como un continuo que va desde estructuras simples hasta arquitecturas altamente integradas. **La evolución desde sistemas reactivos hacia sistemas deliberativos, colaborativos y estratégicos refleja una transformación en la forma en que se organiza el comportamiento**, pasando de respuestas inmediatas a procesos complejos de planificación, coordinación y adaptación.

En términos arquitectónicos, el capítulo demuestra que la modularidad, la integración y la hibridez constituyen principios fundamentales en el diseño de sistemas agénticos. **Las arquitecturas modulares permiten gestionar la complejidad mediante la especialización funcional, mientras que las arquitecturas integradas garantizan la coherencia sistémica del comportamiento**, y las arquitecturas híbridas introducen una capacidad de adaptación estructural que permite al sistema evolucionar en función del entorno. Este tránsito refleja una tendencia hacia sistemas cada vez más flexibles, escalables y autónomos.

Finalmente, las arquitecturas emergentes, particularmente los sistemas multiagente, distribuidos y ecosistémicos, redefinen el concepto de inteligencia al desplazarlo hacia una dimensión colectiva. **La inteligencia deja de ser una propiedad individual para convertirse en una propiedad emergente de la interacción entre múltiples agentes**, lo que introduce dinámicas de coordinación, descentralización y coevolución del comportamiento. En este contexto, la agencia se configura como un fenómeno relacional, donde la coherencia del sistema depende de la capacidad de los agentes para interactuar, adaptarse y evolucionar conjuntamente.

En síntesis, el capítulo establece que la IA agéntica debe entenderse como una arquitectura dinámica en la que la organización interna del sistema determina su capacidad para sostener comportamiento coherente, adaptativo y continuo. **La inteligencia no reside en los componentes, sino en la forma en que estos se articulan dentro de una estructura funcional integrada**, consolidando así una nueva comprensión de la inteligencia artificial como organización del comportamiento en entornos complejos. Ver **Tabla 2**

Tabla 2. Arquitectura y estructuración de la IA agéntica

Concepto	Definición	Diferenciación	Ventajas	Limitaciones	Referencias
Arquitectura agéntica	Estructura organizativa que integra componentes funcionales del agente (percepción, decisión,	Se diferencia de arquitecturas tradicionales por su enfoque en comportamiento integrado y no modular aislado	Permite coherencia sistémica y funcionamiento continuo	Alta complejidad en diseño e integración	Russell & Norvig (2022); Poole & Mackworth (2017)

	acción y memoria)				
Percepción en agentes	Proceso mediante el cual el sistema interpreta información del entorno para generar representación operativa	A diferencia de sensores pasivos, implica procesamiento contextual de información	Permite interacción contextualizada con el entorno	Dependencia de calidad y ruido de datos	Dorri et al. (2018); Wang (2025)
Toma de decisiones	Mecanismo que transforma información en acciones mediante evaluación de alternativas	Se diferencia de reglas fijas por su capacidad de adaptación y evaluación dinámica	Permite comportamiento orientado a objetivos	Riesgo de incertidumbre y errores de evaluación	Russell (2019); Acharya et al. (2025)
Acción y ejecución	Implementación de decisiones en el entorno mediante respuestas operativas	A diferencia de ejecución automática, implica coherencia con objetivos del sistema	Permite validar decisiones en contextos reales	Dependencia del entorno y sus restricciones	Abou Ali et al. (2026); Sapkota et al. (202

Fuente: Recopilación y elaboración. propia

CAPÍTULO 3. Diseño de la IA agéntica

. El diseño de la **inteligencia artificial agéntica** implica un cambio profundo en la forma en que se conciben los sistemas inteligentes, desplazando el enfoque desde la optimización de funciones aisladas hacia la estructuración del comportamiento como unidad analítica central. En este marco, diseñar no significa únicamente configurar componentes técnicos, sino establecer las condiciones bajo las cuales un sistema puede sostener coherencia en la acción dentro de entornos dinámicos. Este enfoque introduce una dimensión estructural en la cual la inteligencia se manifiesta en la capacidad de organizar procesos de manera integrada, lo que redefine los criterios tradicionales de diseño.

El presente capítulo tiene como propósito **analizar los principios, modelos y fundamentos que permiten construir sistemas agénticos desde una perspectiva rigurosa, diferenciándolos de aproximaciones convencionales centradas en el procesamiento de información**. En este sentido, el diseño se entiende como un proceso conceptual que articula percepción, decisión y acción dentro de una lógica operativa coherente, lo que implica considerar no solo la funcionalidad del sistema, sino también su capacidad para sostener comportamiento organizado en contextos variables.

Asimismo, el capítulo delimita su **alcance epistemológico** al situar el diseño como una práctica que integra teoría y aplicación, evitando reduccionismos técnicos y promoviendo una comprensión más amplia de la inteligencia artificial. Esta perspectiva permite establecer un continuum conceptual que conecta los fundamentos teóricos con las configuraciones operativas, sentando las bases para el análisis de los subtemas posteriores, en los cuales se abordarán los distintos niveles de diseño que estructuran la agencia artificial.

Principios estructurales del diseño agéntico

El subtema aborda los principios fundamentales del diseño en la **IA agéntica**, enfocándose en los **criterios estructurales** que permiten organizar el comportamiento de los sistemas inteligentes. Se analizan las bases conceptuales que sustentan la construcción de agentes capaces de sostener coherencia en la acción, integrando múltiples procesos dentro de ua misma lógica operativa. Este enfoque permite diferenciar el diseño agéntico de las aproximaciones tradicionales, destacando la importancia de la i**ntegración, la continuidad y la adaptabilidad como elementos centrales** en la estructuración del comportamiento inteligente en contextos dinámicos.

Diseño basado en la organización del comportamiento

El **diseño basado en la organización del comportamiento** constituye un cambio paradigmático en la construcción de sistemas agénticos, al desplazar el foco desde la arquitectura estática hacia la estructuración dinámica de las acciones del agente. En este enfoque, el comportamiento no es un resultado derivado de la estructura, sino el elemento primario que define cómo se organiza el sistema. Así, los agentes son concebidos como entidades que **planifican, ejecutan, verifican y reajustan acciones en ciclos iterativos**, integrando memoria, razonamiento y objetivos dentro de un mismo flujo operativo orientado a metas (Bandi et al., 2025). Desde esta perspectiva, el diseño de sistemas agénticos implica la articulación de componentes que permitan sostener procesos conductuales complejos, como la descomposición de tareas, la planificación jerárquica y la ejecución iterativa en múltiples etapas. A diferencia de los modelos tradicionales, donde la lógica de operación es rígida y predeterminada, los sistemas actuales organizan su comportamiento en función de metas dinámicas y contextuales. En este sentido, el comportamiento deja de ser una consecuencia y se convierte en una estructura en sí misma, configurando una lógica donde la **organización emerge del flujo continuo de acciones orientadas a objetivos** (Sapkota et al., 2026).

A diferencia de esta concepción centrada en la dinámica individual, los enfoques de inteligencia distribuida subrayan que el comportamiento debe entenderse como resultado de la interacción entre múltiples agentes dentro de un entorno compartido. En este contexto, el diseño basado en la organización del comportamiento se expande hacia sistemas donde múltiples entidades coordinan sus acciones para resolver problemas complejos. Esto implica que el comportamiento colectivo no es simplemente

agregado, sino construido a partir de interacciones estructuradas, consolidando una lógica de **organización conductual distribuida y colaborativa** (Piccialli et al., 2025). Operativamente hablando, la organización del comportamiento requiere mecanismos que permitan la toma de decisiones en entornos dinámicos e inciertos, lo que introduce el papel del aprendizaje en la estructuración del agente. En particular, los enfoques de aprendizaje por refuerzo multiagente muestran que los agentes pueden desarrollar estrategias conductuales mediante la interacción continua con el entorno y con otros agentes. Este proceso permite que el comportamiento no sea predefinido, sino emergente, evolucionando a partir de la experiencia y la retroalimentación, consolidando una lógica de **aprendizaje conductual adaptativo y emergente** (Yadav et al., 2023).

No obstante, la organización del comportamiento no puede sostenerse únicamente en procesos de aprendizaje, ya que requiere una estructura que permita integrar capacidades heterogéneas dentro del agente. En este punto, los frameworks de agentes destacan la importancia de diseñar sistemas que articulen percepción, razonamiento, memoria y acción dentro de un mismo esquema operativo. Esta integración permite que el agente procese información multimodal, tome decisiones informadas y ejecute acciones coherentes en contextos complejos, consolidando una lógica de **integración funcional orientada al comportamiento inteligente** (Li, 2026).

Comparativamente, mientras que los modelos tradicionales organizan el sistema a partir de su arquitectura interna, el diseño basado en el comportamiento invierte esta lógica, estructurando la arquitectura en función de las dinámicas conductuales del agente. Esta diferencia implica que el sistema no se define por sus componentes, sino por la forma en que estos interactúan para generar comportamiento significativo. En consecuencia, el diseño agéntico se configura como un proceso donde la estructura emerge de la interacción entre funciones, consolidando una lógica de **arquitectura emergente basada en comportamiento dinámico** (Sapkota et al., 2026).

Asimismo, la organización del comportamiento implica la capacidad del agente para operar en ciclos de retroalimentación continua, donde cada acción es evaluada en función de sus resultados y ajustada en consecuencia. Este proceso permite que el sistema mejore progresivamente su desempeño, adaptándose a condiciones cambiantes sin intervención externa directa. En este sentido, el comportamiento no es lineal, sino iterativo y reflexivo, integrando planificación, ejecución y evaluación dentro de un mismo ciclo, lo que configura una lógica de **comportamiento autoajustable y reflexivo** (Bandi et al., 2025). El diseño basado en la organización del comportamiento redefine el concepto de inteligencia artificial al introducir la adaptabilidad como principio estructural del sistema. En lugar de concebir la inteligencia como una propiedad estática, este enfoque permite entenderla como un proceso dinámico que emerge de la interacción entre el agente, otros agentes y el entorno. Así, la inteligencia agéntica se configura como una **propiedad emergente de sistemas capaces de organizar, adaptar y evolucionar su comportamiento en función de objetivos, contexto y experiencia acumulada**, estableciendo un nuevo paradigma en el diseño de sistemas inteligentes (Piccialli et al., 2025).

Integración estructural como criterio de diseño

La **integración estructural** se configura como un criterio fundamental en el diseño de sistemas agénticos, al establecer la forma en que los distintos componentes del agente —percepción, razonamiento, memoria y acción— se articulan dentro de una arquitectura coherente. A diferencia de los enfoques modulares clásicos, en los que cada componente opera de manera relativamente aislada, los sistemas agénticos contemporáneos requieren una integración que permita la coordinación dinámica de procesos en función de objetivos complejos. En este sentido, el diseño no se limita a ensamblar módulos, sino que implica la construcción de un sistema donde la **coherencia emerge de la interacción funcional entre sus partes** (Sapkota et al., 2026). La integración estructural responde a la necesidad de superar las limitaciones de los sistemas tradicionales, caracterizados por su rigidez y baja adaptabilidad. Mientras que los agentes clásicos operaban bajo reglas predefinidas y estructuras estáticas, los sistemas actuales incorporan capacidades de aprendizaje, memoria persistente y razonamiento contextual, lo que exige una integración más profunda entre sus componentes. Este cambio implica que el agente no solo ejecuta tareas, sino que **coordina múltiples procesos cognitivos dentro de un mismo flujo operativo**, permitiendo una mayor flexibilidad en entornos dinámicos (Bandi et al., 2025).

A diferencia de esta visión centrada en la coordinación interna, los enfoques de inteligencia distribuida destacan que la integración estructural también debe extenderse al entorno en el que opera el agente. En este contexto, los sistemas agénticos se conciben como entidades que interactúan con otros agentes y con el entorno mediante procesos de comunicación, aprendizaje y toma de decisiones autónoma. Esto implica que la integración no solo ocurre dentro del agente, sino también entre agentes, configurando una lógica de **integración sistémica que articula múltiples niveles de interacción** (Piccialli et al., 2025). la integración estructural se manifiesta en la capacidad del sistema para coordinar procesos como la planificación, la ejecución y la evaluación dentro de ciclos continuos de retroalimentación. Este enfoque permite que el agente no solo actúe, sino que también evalúe sus acciones y ajuste su comportamiento en función de los resultados obtenidos. En contraste con los sistemas lineales, donde las decisiones se ejecutan sin revisión, los sistemas integrados operan bajo una lógica de **retroalimentación continua que garantiza la coherencia del comportamiento** (Bandi et al., 2025).

Sin embargo, la integración estructural no puede sostenerse únicamente mediante la coordinación de procesos en tiempo real, lo que introduce la necesidad de mecanismos que permitan conservar y utilizar información a lo largo del tiempo. En este sentido, la memoria se convierte en un componente clave de la integración, ya que permite conectar experiencias pasadas con decisiones futuras. Esta capacidad es particularmente relevante en entornos dinámicos, donde la adaptación depende de la acumulación de conocimiento. Así, la integración estructural se fortalece mediante una **memoria funcional que articula la continuidad del comportamiento** (Piccialli et al., 2025). Mientras que los modelos tradicionales organizan el sistema a partir de su arquitectura interna, la integración estructural en sistemas agénticos invierte esta

lógica, priorizando la interacción entre componentes como base del diseño. Esta diferencia implica que la arquitectura no es un fin en sí misma, sino un medio para sostener procesos dinámicos de comportamiento. En consecuencia, el diseño agéntico se configura como un proceso en el que la estructura emerge de la interacción entre funciones, consolidando una lógica de **arquitectura emergente basada en integración funcional** (Sapkota et al., 2026).

Asimismo, la integración estructural permite abordar problemas complejos mediante la coordinación de múltiples agentes que interactúan dentro de un entorno compartido. En este contexto, los sistemas multiagente utilizan mecanismos de aprendizaje y comunicación para alinear sus acciones y optimizar resultados colectivos. Esto implica que la integración no solo mejora el desempeño individual del agente, sino que también permite la construcción de comportamientos colectivos más eficientes, consolidando una lógica de **integración distribuida orientada a la cooperación** (Yadav et al., 2023).

La integración estructural redefine el concepto de diseño en inteligencia artificial al introducir la coherencia como criterio central del sistema. En lugar de concebir el diseño como la simple combinación de componentes, este enfoque permite entenderlo como un proceso de articulación funcional que garantiza la interacción efectiva entre las distintas capacidades del agente. Así, la inteligencia agéntica se configura como una **propiedad emergente de sistemas que integran de manera coherente percepción, decisión, memoria y acción**, estableciendo un paradigma que trasciende los modelos tradicionales de diseño en inteligencia artificial (Li, 2026).

Adaptabilidad y coherencia como principios de diseño

La **adaptabilidad y la coherencia** constituyen principios fundamentales en el diseño de sistemas agénticos, en la medida en que permiten equilibrar la capacidad del sistema para responder a entornos dinámicos sin comprometer la consistencia de su comportamiento. A diferencia de los sistemas tradicionales, caracterizados por estructuras rígidas y reglas predefinidas, los sistemas agénticos contemporáneos integran mecanismos que les permiten modificar sus estrategias en función del contexto. Este cambio implica que la inteligencia artificial evoluciona hacia modelos capaces de **ajustarse dinámicamente a condiciones cambiantes manteniendo continuidad operativa**, lo que redefine los criterios de diseño en sistemas inteligentes (Sapkota et al., 2026). La adaptabilidad no debe entenderse como una simple capacidad de ajuste, sino como un proceso estructurado mediante el cual el agente reorganiza su comportamiento en función de objetivos, información disponible y retroalimentación del entorno. Los sistemas agénticos integran capacidades como planificación, memoria persistente y razonamiento contextual, lo que permite que las decisiones se ajusten de manera informada. En este sentido, la adaptabilidad se configura como una propiedad dinámica que permite la evolución del sistema, consolidando una lógica de **comportamiento flexible orientado a metas complejas** (Bandi et al., 2025).

Sin embargo, la adaptabilidad introduce un riesgo inherente de desorganización si no se encuentra regulada por un principio que garantice la estabilidad del sistema. En este contexto, la coherencia emerge como el mecanismo que permite mantener la continuidad del comportamiento, asegurando que las decisiones del agente sean consistentes con sus objetivos y con su trayectoria operativa. Desde una perspectiva sistémica, la coherencia implica la integración funcional de los distintos componentes del agente, evitando contradicciones internas y consolidando una lógica de **consistencia estructural en la toma de decisiones** (Piccialli et al., 2025). A diferencia de esta concepción centrada en la estabilidad interna, los enfoques de aprendizaje multiagente muestran que la adaptabilidad también depende de la interacción entre múltiples agentes en entornos compartidos. En estos sistemas, cada agente ajusta su comportamiento en función de las acciones de otros agentes, lo que genera dinámicas complejas de coordinación y aprendizaje. Este proceso implica que la adaptabilidad no es únicamente individual, sino relacional, configurando una lógica de **adaptación distribuida basada en interacción y optimización colectiva** (Yadav et al., 2023).

La coherencia se sostiene mediante la integración de los distintos componentes del agente dentro de un marco organizativo que articula percepción, razonamiento, memoria y acción. Esta integración permite que el agente procese información de manera consistente, tome decisiones fundamentadas y ejecute acciones coherentes en diferentes contextos. En este sentido, la coherencia no es una propiedad estática, sino el resultado de una arquitectura que permite mantener la estabilidad funcional del sistema frente a la variabilidad del entorno, consolidando una lógica de **coherencia estructural orientada a la continuidad operativa** (Li, 2026). La adaptabilidad introduce variabilidad en el comportamiento del sistema, mientras que la coherencia actúa como un mecanismo regulador que limita dicha variabilidad para evitar la pérdida de control. Esta relación no debe entenderse como una oposición, sino como una complementariedad necesaria para el funcionamiento del sistema. La adaptabilidad permite responder a la incertidumbre, mientras que la coherencia asegura que dichas respuestas mantengan una dirección consistente. En consecuencia, el diseño agéntico se configura como un equilibrio dinámico entre **cambio (adaptabilidad) y estabilidad (coherencia)** (Sapkota et al., 2026).

La interacción entre adaptabilidad y coherencia se manifiesta en los ciclos de retroalimentación que caracterizan a los sistemas agénticos. Estos ciclos permiten que el agente evalúe sus acciones, incorpore nueva información y ajuste su comportamiento sin perder consistencia. En este proceso, la memoria desempeña un papel central al permitir que el sistema conserve información sobre experiencias previas y las utilice en decisiones futuras. De esta manera, la adaptabilidad se articula con la coherencia a través de una lógica de **retroalimentación continua que sostiene la estabilidad dinámica del sistema** (Bandi et al., 2025). La integración de adaptabilidad y coherencia redefine el diseño de sistemas inteligentes al introducir la necesidad de concebir la inteligencia como un proceso simultáneamente dinámico y estructurado. En lugar de entender la inteligencia como una propiedad fija, este enfoque permite concebirla como una capacidad emergente que resulta de la

interacción entre procesos de ajuste y mecanismos de estabilidad. Así, la inteligencia agéntica se configura como una **propiedad emergente de sistemas capaces de adaptarse sin perder coherencia, integrando aprendizaje, memoria y acción en un marco estructural consistente**, estableciendo un nuevo paradigma en el diseño de sistemas inteligentes (Piccialli et al., 2025).

Modelado del comportamiento en sistemas agénticos

El **modelado del comportamiento en sistemas agénticos** se refiere al proceso de **diseñar y formalizar cómo un agente toma decisiones y actúa en función de su estado interno y del entorno**. A diferencia de enfoques tradicionales, donde el comportamiento se define mediante reglas fijas, en los sistemas agénticos este se construye como una **dinámica estructurada de transición entre estados**, permitiendo mayor adaptabilidad y coherencia operativa. Este modelado integra elementos como **memoria, contexto, objetivos y mecanismos de decisión**, los cuales interactúan para generar respuestas consistentes ante situaciones cambiantes. Asimismo, incorpora representaciones tanto discretas como continuas, especialmente en arquitecturas basadas en modelos de lenguaje, donde el comportamiento depende de inferencias contextuales. En este sentido, el modelado no solo describe acciones, sino que **establece la lógica interna que guía la evolución del agente**, garantizando su funcionamiento autónomo, flexible y orientado a objetivos dentro de entornos complejos.

Formalización del comportamiento como sistema de estados operativos

Desde una perspectiva estrictamente constructiva, el modelado del comportamiento en sistemas agénticos debe entenderse como la **formalización explícita de las transiciones entre estados operativos del agente**, donde cada estado representa una configuración interna relevante del sistema y cada transición define su evolución bajo condiciones específicas. En este enfoque, el comportamiento deja de ser interpretado como una simple salida observable para convertirse en una **estructura interna diseñada que organiza la dinámica del agente**, lo que permite su análisis, control y optimización. Así, el modelado conductual se posiciona como el núcleo del diseño agéntico, al definir la forma en que el sistema responde de manera coherente ante variaciones del entorno (Amancharla et al., 2025).

En términos analíticos, esta formalización implica concebir el comportamiento como una **función de transición condicional**, en la cual el agente evalúa su estado actual y determina la acción más adecuada dentro de un conjunto de alternativas posibles. Este proceso requiere integrar variables como objetivos, restricciones del entorno y capacidades del sistema, lo que transforma el modelado en un problema de **representación estructurada y evaluación multi-criterio**. De esta manera, el

comportamiento se construye como un sistema dinámico susceptible de simulación y validación, lo que permite anticipar su desempeño antes de su implementación y ajustar su diseño para garantizar coherencia operativa (Li, 2026).

Desde un enfoque contemporáneo, particularmente en sistemas basados en modelos de lenguaje, la formalización del comportamiento incorpora **representaciones contextuales dinámicas**, lo que permite al agente reinterpretar su estado en función de información semántica. En estos sistemas, las decisiones no dependen exclusivamente de reglas discretas, sino de la interacción entre contexto, memoria y generación de inferencias. Esto implica que el modelado debe considerar **espacios de representación continuos**, donde el comportamiento emerge de gradientes contextuales y no de estructuras rígidas, ampliando la capacidad adaptativa del agente en entornos complejos (Wang et al., 2024).

En términos estructurales, el modelado del comportamiento exige la definición de una arquitectura interna compuesta por **múltiples capas funcionales interdependientes**, donde cada capa cumple un rol específico en la transformación del estado del agente. Estas capas pueden incluir procesos de interpretación, evaluación, decisión y ejecución, los cuales deben operar de manera coordinada para evitar inconsistencias. En este sentido, el comportamiento se convierte en un problema de **coherencia estructural interna**, donde la calidad del sistema depende de la correcta articulación entre sus componentes más que de la complejidad individual de cada uno (Sapkota et al., 2026).

Desde una perspectiva de control, la formalización del comportamiento también implica la incorporación de **mecanismos de restricción que delimiten el espacio de estados posibles del agente**, evitando comportamientos no deseados o incoherentes. Estas restricciones pueden materializarse en reglas normativas, umbrales de decisión o funciones de penalización que orientan la evolución del sistema hacia resultados consistentes con sus objetivos. En este sentido, el modelado no solo busca generar comportamiento, sino también **regularlo y estabilizarlo**, garantizando que el agente opere dentro de parámetros definidos incluso en entornos dinámicos (Biswas & Talukdar, 2025). Lla formalización del comportamiento también requiere considerar la dimensión temporal del sistema, ya que las decisiones del agente no solo dependen de su estado actual, sino también de su trayectoria previa. Esto implica incorporar mecanismos de memoria que permitan almacenar y recuperar información relevante, lo que transforma el comportamiento en un proceso acumulativo y evolutivo. De esta manera, el modelado debe integrar estructuras que permitan representar la historia del agente, facilitando la construcción de patrones de comportamiento más eficientes a lo largo del tiempo (Zhang et al., 2025).

Desde un punto de vista operativo, la formalización del comportamiento permite trasladar los modelos teóricos a sistemas implementables capaces de operar en tiempo real. Esto implica considerar aspectos como eficiencia computacional, robustez y escalabilidad, ya que el comportamiento del agente depende de su capacidad para procesar información y ejecutar decisiones bajo condiciones variables. En este

contexto, el comportamiento se consolida como una **estructura diseñada que articula teoría, modelado y ejecución**, definiendo la forma en que el agente interactúa con su entorno y cumple sus objetivos de manera autónoma y coherente (Amancharla et al., 2025).

Representación funcional del comportamiento y procesos de decisión

El **modelado del comportamiento en sistemas agénticos** requiere trascender la formalización de estados para centrarse en la definición de los **procesos de decisión** que permiten al agente seleccionar acciones en contextos dinámicos. En este nivel, el comportamiento se entiende como una función operativa que transforma información en acción, mediante mecanismos internos de evaluación estructurada. Así, el comportamiento no es únicamente una transición entre estados, sino una **configuración funcional de decisión**, donde el agente integra múltiples variables para actuar de manera coherente (Li, 2026). El comportamiento del agente se articula mediante la interacción entre **planificación, memoria operativa y ejecución**, lo que permite estructurar decisiones en función del contexto. La planificación genera alternativas, la memoria proporciona información relevante y la ejecución concreta la acción. Esta articulación configura una **arquitectura decisional integrada**, en la cual el comportamiento emerge como resultado de la coordinación entre procesos internos especializados (Wang et al., 2024).

En el ámbito de la optimización, la toma de decisiones puede modelarse como un proceso de **evaluación multi-criterio**, donde el agente selecciona la acción que mejor satisface sus objetivos bajo restricciones específicas. Este enfoque implica comparar alternativas considerando variables como eficiencia, riesgo y contexto, lo que convierte el comportamiento en un problema de optimización estructurada. De esta manera, el agente no actúa de forma reactiva, sino que realiza una selección fundamentada entre opciones posibles (Biswas & Talukdar, 2025). La decisión requiere la articulación de múltiples capas funcionales que operan de manera coordinada. Estas capas incluyen la interpretación del entorno, la evaluación de opciones y la selección de acciones, lo que permite estructurar el comportamiento como una secuencia organizada de procesos internos. En este sentido, el comportamiento se configura como una **estructura jerárquica funcional**, donde cada nivel contribuye a la coherencia global del sistema (Sapkota et al., 2026).

En el plano adaptativo, el comportamiento del agente incorpora la capacidad de ajustar sus decisiones en función de cambios en el entorno. Este proceso implica la reconfiguración de estrategias en tiempo real, lo que introduce una dimensión dinámica en la toma de decisiones. La adaptación no se limita a responder a estímulos, sino que implica la capacidad de modificar los criterios de decisión en función de nuevas condiciones, consolidando una **adaptación decisional dinámica** (Chen, 2025).

Desde una perspectiva de aprendizaje, el comportamiento se fortalece mediante procesos de **aprendizaje continuo**, en los cuales el agente mejora sus decisiones a partir de la experiencia acumulada. Este enfoque implica que las decisiones futuras se ven influenciadas por resultados pasados, lo que permite construir estrategias más eficientes. El aprendizaje introduce una dimensión evolutiva en la toma de decisiones, donde el comportamiento se perfecciona progresivamente (Yadav et al., 2023).

En el plano relacional, la toma de decisiones se ve influida por la interacción con otros agentes, lo que implica que las decisiones individuales no son completamente independientes. Este fenómeno introduce una lógica en la que el comportamiento se configura como un proceso de **ajuste relacional**, donde las acciones se alinean con dinámicas colectivas. En este contexto, el agente debe considerar no solo sus objetivos, sino también el comportamiento de otros actores (Yuan & Xie, 2026).

La toma de decisiones requiere integrar mecanismos de memoria que permitan recuperar información relevante para la selección de acciones. Esta memoria no solo almacena datos, sino que contribuye a la construcción de contexto, lo que permite decisiones más informadas y coherentes. La incorporación de memoria transforma el comportamiento en un proceso acumulativo, donde la experiencia influye en la acción presente (Zhang et al., 2025).

Por tanto, desde una perspectiva operativa, la representación funcional del comportamiento permite implementar sistemas capaces de tomar decisiones en tiempo real. Esto implica considerar eficiencia, robustez y escalabilidad, ya que el comportamiento depende de la capacidad del sistema para procesar información de manera efectiva. En este sentido, el modelado conductual se consolida como una integración de **representación, decisión y ejecución**, que define la capacidad del agente para operar de manera autónoma en entornos complejos (Li, 2026).

Evaluación, coherencia y validación del comportamiento agéntico

El **modelado del comportamiento en sistemas agénticos** requiere incorporar mecanismos de evaluación que permitan verificar la consistencia entre las acciones del agente y sus objetivos operativos. En este nivel, el comportamiento deja de entenderse únicamente como ejecución para convertirse en un objeto de análisis sujeto a criterios de calidad, donde la coherencia, la estabilidad y la adaptabilidad se posicionan como dimensiones clave. La evaluación se configura así como un proceso sistemático orientado a garantizar la **validez operativa del comportamiento** en entornos dinámicos (Amancharla et al., 2025). La evaluación implica contrastar las decisiones del agente con las condiciones del entorno y con su propio estado interno. Este proceso permite identificar desviaciones entre lo esperado y lo ejecutado, lo que resulta fundamental para ajustar el modelo conductual. La coherencia se configura como un criterio central, ya que asegura que las decisiones del agente mantengan una

correspondencia lógica con sus objetivos y restricciones, evitando comportamientos inconsistentes o erráticos (Li, 2026).

En el plano cognitivo, la validación del comportamiento depende de la capacidad del sistema para mantener alineación entre sus procesos internos de planificación, memoria y ejecución. La desarticulación entre estos componentes puede generar respuestas incoherentes, por lo que la evaluación debe considerar la **integridad funcional del sistema**. Este enfoque implica analizar el comportamiento como resultado de la coordinación entre múltiples procesos internos que deben operar de manera consistente (Wang et al., 2024).

Desde la perspectiva de la optimización, la evaluación del comportamiento implica medir el desempeño del agente en términos de eficiencia, cumplimiento de objetivos y uso adecuado de recursos. Este proceso permite identificar áreas de mejora y ajustar las estrategias del agente para optimizar su funcionamiento. La evaluación se convierte así en un mecanismo de **retroalimentación estructurada**, que orienta la mejora continua del comportamiento a través de ciclos iterativos de ajuste (Biswas & Talukdar, 2025). la coherencia del comportamiento depende de la correcta articulación entre los distintos componentes del sistema. La evaluación debe considerar cómo interactúan estos componentes, identificando posibles fallas en su coordinación. En este sentido, el comportamiento se analiza como una estructura integrada cuya calidad depende de la **coherencia interna del sistema**, más que del desempeño aislado de cada elemento (Sapkota et al., 2026).

Desde una perspectiva adaptativa, la validación del comportamiento debe considerar la capacidad del agente para operar en entornos caracterizados por la incertidumbre. Esto implica evaluar no solo la consistencia de las decisiones, sino también su flexibilidad frente a cambios en el entorno. La evaluación se configura como un proceso dinámico que permite ajustar el comportamiento en función de nuevas condiciones, consolidando la **adaptabilidad como criterio de calidad conductual** (Chen, 2025).

En el plano relacional, la coherencia del comportamiento también se evalúa en **función de su interacción** con otros agentes dentro del sistema. En este contexto, es necesario verificar que las acciones individuales contribuyan a la estabilidad del sistema y no generen conflictos que afecten su funcionamiento. Esto introduce una dimensión colectiva en la evaluación, donde el comportamiento del agente se analiza en términos de su impacto en el sistema global (Yuan & Xie, 2026). La validación del comportamiento requiere incorporar la dimensión temporal, lo que implica analizar la evolución del agente a lo largo del tiempo. La memoria permite identificar patrones de comportamiento y **evaluar su consistencia en diferentes momentos**, lo que resulta fundamental para detectar tendencias y ajustar estrategias. Este enfoque permite comprender el comportamiento como un proceso acumulativo, donde la experiencia influye en la calidad de las decisiones futuras (Zhang et al., 2025).

Finalmente, desde una perspectiva operativa, la evaluación del comportamiento permite **garantizar que los sistemas agénticos funcionen de manera eficiente** en entornos reales. Esto implica considerar aspectos como robustez, escalabilidad y capacidad de respuesta, asegurando que el agente pueda operar de manera autónoma sin comprometer su coherencia. En este sentido, la validación se consolida como un proceso esencial para el diseño de sistemas inteligentes confiables, donde el comportamiento no solo se ejecuta, sino que se analiza, ajusta y perfecciona de manera continua (Amancharla et al., 2025).

Diseño funcional y organización de componentes del agente

El **diseño funcional y la organización de componentes del agente** constituyen la base estructural que permite a los sistemas agénticos operar de manera autónoma y coherente en entornos dinámicos. Este diseño implica la integración de elementos como **modelo central, memoria y herramientas externas**, los cuales definen la arquitectura del agente **(Li, 2026).** A nivel operativo, el sistema se articula mediante un **ciclo continuo de percepción, decisión y acción**, garantizando la continuidad del comportamiento (Amancharla et al., 2025). Desde la perspectiva cognitiva, funciones como la **planificación, memoria operativa y ejecución** organizan los procesos internos del agente (Wang et al., 2024). Asimismo, en el marco de la **IA agéntica**, la incorporación de **colaboración multiagente y descomposición dinámica de tareas** amplía su capacidad adaptativa (Sapkota et al., 2026). Finalmente, la integración de **razonamiento, planificación y acción autónoma** permite su aplicación efectiva en contextos reales (Biswas & Talukdar, 2025).

Estructura funcional del agente

El **diseño funcional de los agentes inteligentes** se configura como un proceso de estructuración orientado a garantizar la operación continua del sistema mediante la articulación coherente de sus componentes. A diferencia de los modelos tradicionales, donde las funciones se ejecutan de manera aislada, los sistemas agénticos requieren una organización basada en la **continuidad operativa**, en la cual el comportamiento emerge de ciclos recurrentes de interacción con el entorno. Esta perspectiva implica que el agente no es una entidad estática, sino un sistema dinámico cuya funcionalidad depende de la persistencia de su operación a lo largo del tiempo (**Amancharla et al., 2025**). El agente puede definirse como una configuración compuesta por elementos diferenciados que cumplen funciones específicas dentro del sistema global. En este sentido, la arquitectura del agente se organiza en torno a tres componentes fundamentales: el **modelo central**, encargado del procesamiento; la **memoria**, responsable de almacenar y recuperar información; y las **herramientas externas**, que permiten extender las capacidades del sistema más allá de su estructura interna. Esta definición delimita con precisión la composición del agente, estableciendo una base ontológica clara para su diseño (Li, 2026).

En un nivel distinto, correspondiente a la organización interna de los procesos, el diseño funcional debe considerar cómo se articulan las funciones cognitivas del agente. En este marco, la estructura interna se organiza en torno a procesos como la **planificación**, la **memoria operativa** y la **ejecución**, los cuales permiten transformar información en acción. La planificación anticipa posibles cursos de acción, la memoria operativa sostiene el contexto inmediato y la ejecución materializa las decisiones. Este nivel define la **lógica cognitiva del agente**, es decir, la forma en que sus procesos internos se coordinan para producir comportamiento coherente (Wang et al., 2024).

A nivel paradigmático, el diseño funcional se ve influido por la transición hacia sistemas de **IA agéntica**, los cuales introducen una nueva forma de concebir la organización de los componentes. En este contexto, el agente deja de ser una entidad aislada para convertirse en parte de sistemas más amplios caracterizados por la **colaboración multiagente** y la **descomposición dinámica de tareas**. Esta transformación implica que la estructura funcional ya no se limita a la eficiencia individual, sino que debe considerar la coordinación entre múltiples agentes, lo que redefine el alcance del diseño (Sapkota et al., 2026**).** En el nivel aplicado, el diseño funcional adquiere relevancia en la medida en que permite la ejecución efectiva de tareas en entornos reales. En este plano, la organización de componentes debe facilitar la integración de **razonamiento, planificación y acción autónoma**, asegurando que el agente pueda operar de manera eficiente sin intervención externa. La funcionalidad del sistema se evalúa entonces en términos de su capacidad para transformar decisiones en resultados concretos, lo que introduce una dimensión pragmática en el diseño (Biswas & Talukdar, 2025).

La articulación de estos niveles permite comprender que el diseño funcional no es un proceso unidimensional, sino una configuración compleja en la que cada dimensión cumple un rol específico. La separación entre niveles —operativo, estructural, cognitivo, paradigmático y aplicado— garantiza que cada componente conceptual aporte una función diferenciada, evitando redundancias y fortaleciendo la claridad del modelo. En este sentido, la **ortogonalidad del diseño** no solo es un criterio metodológico, sino una condición necesaria para la construcción de sistemas coherentes (Amancharla et al., 2025).

Asimismo, la estructura funcional del agente debe contemplar la capacidad de operar en entornos caracterizados por incertidumbre y variabilidad constante. Esto implica que los componentes no solo deben estar correctamente definidos, sino también organizados de tal manera que permitan la **adaptación continua** del sistema. La memoria, en este contexto, no solo almacena información, sino que contribuye a la construcción de contexto; la planificación no solo anticipa acciones, sino que permite reconfigurar estrategias; y la ejecución no solo actúa, sino que transforma el entorno, generando nuevas condiciones de operación (Wang et al., 2024).

En síntesis, el **diseño funcional de los agentes inteligentes** se configura como un proceso de integración estructural en el que la inteligencia emerge de la interacción entre sus componentes organizados en distintos niveles. La **coherencia operativa del**

sistema depende de la correcta articulación entre operación, estructura, cognición, paradigma y aplicación, lo que permite sostener comportamiento autónomo en entornos complejos. De este modo, la ortogonalidad máxima no solo garantiza claridad conceptual, sino que constituye la base para el desarrollo de sistemas agénticos robustos y funcionalmente consistentes (Li, 2026).

Organización de componentes y lógica de integración

La **organización de componentes en sistemas agénticos** implica establecer una lógica de integración que permita coordinar los distintos elementos del sistema dentro de un flujo operativo coherente. A diferencia de los enfoques tradicionales basados en modularidad independiente, los agentes inteligentes requieren una estructura en la que los componentes interactúan de manera continua, generando un comportamiento unificado. En este sentido, la organización no solo define la disposición de los elementos, sino también las reglas de interacción que garantizan la **coherencia funcional del sistema** (Amancharla et al., 2025). La integración de componentes se fundamenta en la articulación entre el **modelo central, la memoria y las herramientas externas**, los cuales deben operar de manera coordinada para ampliar las capacidades del agente. Este esquema permite que el sistema no se limite a procesar información internamente, sino que pueda acceder a recursos externos y construir conocimiento acumulativo. La lógica de integración en este nivel radica en establecer relaciones funcionales claras entre estos elementos, evitando fragmentaciones en el comportamiento (Li, 2026).

En el plano cognitivo, la organización de componentes se orienta a la coordinación de procesos internos que permiten transformar información en acción. En este contexto, funciones como la **planificación, la memoria operativa y la ejecución** deben integrarse mediante mecanismos que aseguren su sincronización. La planificación define cursos de acción, la memoria sostiene el contexto y la ejecución materializa las decisiones. La correcta articulación de estos procesos constituye la base de la **integración cognitiva del agente**, permitiendo respuestas consistentes y adaptativas (Wang et al., 2024). La lógica de integración se amplía hacia configuraciones distribuidas en las que múltiples agentes interactúan entre sí. En este contexto, la **colaboración multiagente** y la **descomposición dinámica de tareas** introducen una nueva forma de organización en la que los componentes ya no se limitan a un solo agente, sino que se distribuyen entre varias entidades. Esta integración requiere mecanismos de coordinación que permitan gestionar la interdependencia entre agentes, garantizando la coherencia del sistema en su conjunto (Sapkota et al., 2026).

En el nivel aplicado, la organización de componentes debe facilitar la ejecución efectiva de tareas mediante la integración de **razonamiento, planificación y acción autónoma**. Esta integración no solo permite la operación del agente, sino que también garantiza que las decisiones puedan traducirse en acciones concretas dentro del entorno. La lógica de integración en este nivel se orienta a la **eficiencia operativa**,

asegurando que los procesos internos del agente se alineen con los objetivos del sistema (Biswas & Talukdar, 2025).

La articulación de estos niveles permite comprender que la organización de componentes no es un proceso lineal, sino una configuración compleja en la que cada dimensión cumple un rol específico. La separación entre niveles —operativo, estructural, cognitivo, paradigmático y aplicado— asegura que cada elemento aporte una función diferenciada, evitando redundancias y fortaleciendo la claridad del modelo. En este sentido, la **ortogonalidad en la organización** se convierte en un principio fundamental para el diseño de sistemas agénticos coherentes (Amancharla et al., 2025). Asimismo, la lógica de integración debe considerar la capacidad del sistema para operar en entornos caracterizados por la incertidumbre. Esto implica que los componentes deben organizarse de manera que permitan la **adaptación continua**, mediante la actualización de información, la reconfiguración de planes y la ejecución de nuevas acciones. La integración no solo conecta componentes, sino que también habilita la capacidad del agente para responder a condiciones cambiantes sin perder coherencia operativa (Wang et al., 2024).

Por lo tantop, la **organización de componentes en sistemas agénticos** se define como un proceso de integración multidimensional que articula elementos estructurales, cognitivos y operativos dentro de una lógica coherente. La eficiencia del sistema depende de la capacidad de estos componentes para interactuar de manera coordinada, permitiendo la generación de comportamiento autónomo y adaptativo. La **coherencia funcional emergente** es, en última instancia, el resultado de una integración correctamente diseñada, donde cada componente contribuye de forma específica al funcionamiento del agente (Li, 2026).

Emergencia de comportamiento y coherencia operativa

La **emergencia del comportamiento en sistemas agénticos** constituye una propiedad sistémica derivada de la organización funcional de sus componentes, más que de la acción aislada de alguno de ellos. En este sentido, el comportamiento del agente no se programa directamente, sino que surge como resultado de la interacción continua entre percepción, procesamiento y acción dentro de un ciclo operativo persistente. Esta característica introduce una concepción en la que la inteligencia no es un atributo estático, sino una **dinámica emergente**, sostenida por la continuidad del sistema en el tiempo (Amancharla et al., 2025). La coherencia operativa depende de la correcta articulación entre los elementos que componen al agente, particularmente el **modelo central, la memoria y las herramientas externas**. La integración de estos componentes permite que el sistema mantenga consistencia en su comportamiento, al combinar procesamiento interno con acceso a información acumulada y capacidades externas. En este nivel, la coherencia no se define por la estabilidad, sino por la capacidad del sistema para mantener una **estructura funcional consistente** a pesar de la variabilidad del entorno (Li, 2026).

En el plano cognitivo, la emergencia del comportamiento se encuentra directamente relacionada con la coordinación entre procesos como la **planificación, la memoria operativa y la ejecución**. Estos procesos permiten al agente transformar información en decisiones y acciones de manera organizada. La planificación establece objetivos y estrategias, la memoria proporciona el contexto necesario y la ejecución materializa las decisiones. La interacción entre estos elementos da lugar a una **coherencia cognitiva**, en la cual las acciones del agente se mantienen alineadas con su estado interno (Wang et al., 2024). La emergencia del comportamiento adquiere una dimensión adicional en los sistemas de **IA agéntica**, donde múltiples agentes interactúan dentro de un entorno compartido. En este contexto, la coherencia operativa no depende únicamente de un agente, sino de la coordinación entre varios sistemas autónomos. La **colaboración multiagente** y la **descomposición dinámica de tareas** permiten que el comportamiento global emerja de la interacción entre agentes individuales, lo que introduce una lógica de inteligencia distribuida (Sapkota et al., 2026).

En un nivel de aplicaciómn, la coherencia operativa se manifiesta en la capacidad del agente para ejecutar tareas de manera efectiva en contextos reales. La integración de **razonamiento, planificación y acción autónoma** permite que el sistema no solo responda al entorno, sino que actúe de forma estratégica. En este sentido, el comportamiento del agente se evalúa en función de su capacidad para generar resultados consistentes, lo que refleja la adecuada articulación entre sus procesos internos (Biswas & Talukdar, 2025).

La comprensión de la emergencia del comportamiento requiere reconocer que la coherencia operativa no es una propiedad dada, sino el resultado de una organización funcional adecuada. En este sentido, la interacción entre los distintos niveles —operativo, estructural, cognitivo, paradigmático y aplicado— permite que el sistema mantenga consistencia en su comportamiento, incluso en condiciones de incertidumbre. La **coherencia sistémica** emerge cuando estos niveles se encuentran correctamente alineados (Amancharla et al., 2025). La coherencia operativa implica la capacidad del sistema para adaptarse sin perder continuidad en su comportamiento. Esto requiere que los componentes del agente estén organizados de manera que permitan la actualización constante de información, la reconfiguración de estrategias y la ejecución de nuevas acciones. En este sentido, la coherencia no se opone al cambio, sino que se define como la capacidad de mantener consistencia a través de la transformación (Wang et al., 2024).

En resumen, la **emergencia del comportamiento inteligente** en sistemas agénticos es el resultado de la interacción coordinada entre sus componentes dentro de una estructura funcional integrada. La coherencia operativa depende de la capacidad del sistema para articular sus procesos internos y externos de manera consistente, permitiendo la generación de comportamiento autónomo en entornos complejos. De este modo, la inteligencia del agente no reside en un componente específico, sino en la **integración funcional** que permite la emergencia de comportamiento coherente (Li, 2026).

Arquitecturas distribuidas y descentralización de la agencia

Las **arquitecturas distribuidas y la descentralización de la agencia** representan un cambio fundamental en el diseño de sistemas inteligentes, al sustituir estructuras centralizadas por redes de **agentes autónomos interconectados**. En este enfoque, la inteligencia emerge de la **interacción coordinada entre múltiples agentes**, lo que permite mayor **escalabilidad, flexibilidad y resiliencia**. En entornos industriales, esta lógica favorece la **modularidad y reconfiguración dinámica** de sistemas ciberfísicos, donde cada agente controla recursos específicos. Desde la perspectiva del aprendizaje, el paradigma de **multi-agent reinforcement learning** posibilita decisiones simultáneas en entornos compartidos, promoviendo la adaptación colectiva Asimismo, la transición hacia **IA agéntica** incorpora **colaboración multiagente y descomposición dinámica de tareas**, ampliando la capacidad de resolución de problemas complejos. En conjunto, la descentralización redefine la agencia como una **propiedad sistémica emergente**.

Fundamentos de la arquitectura distribuida

Las **arquitecturas distribuidas en sistemas agénticos** representan un cambio paradigmático en la forma en que se conciben los sistemas inteligentes, al sustituir estructuras centralizadas por configuraciones donde múltiples agentes autónomos interactúan en un entorno compartido. Este enfoque permite que la inteligencia se distribuya entre diversas entidades, eliminando la dependencia de un único punto de control. En consecuencia, la arquitectura distribuida se configura como un modelo que favorece la **flexibilidad, adaptabilidad y robustez sistémica**, elementos fundamentales para operar en entornos complejos (Piccialli et al., 2025).

Bajo un enfoque industrial, Karnouskos et al. (2020) establecen que los sistemas multiagente constituyen un habilitador clave para los **sistemas ciberfísicos industriales**, donde cada agente representa y controla un recurso específico dentro del sistema. Esta correspondencia entre agentes y recursos físicos permite construir estructuras altamente modulares, en las que cada componente puede operar de manera autónoma sin perder coherencia con el sistema global. La modularidad, en este contexto, se convierte en un principio estructural que facilita la **reconfiguración dinámica** de los sistemas productivos (Karnouskos et al., 2020). La descentralización implica que cada agente posee capacidades de percepción, decisión y acción, lo que le permite operar de manera independiente dentro del sistema. Esta autonomía distribuida no elimina la necesidad de coordinación, pero sí redefine la forma en que se organiza el control. En lugar de una jerarquía centralizada, el sistema se estructura como una red de agentes interconectados que interactúan mediante mecanismos de comunicación y cooperación.

Por su parte, Yadav et al. (2023) destacan que los sistemas de **multi-agent reinforcement learning (MARL)** constituyen un enfoque fundamental para comprender el funcionamiento de arquitecturas distribuidas, ya que permiten modelar escenarios en los que múltiples agentes toman decisiones de manera simultánea en un entorno compartido. Este enfoque introduce la noción de que el comportamiento del sistema no puede entenderse a partir de un solo agente, sino como el resultado de la interacción entre múltiples entidades que aprenden de manera conjunta (Yadav et al., 2023). Sapkota et al. (2026) señalan que la evolución hacia sistemas de **IA agéntica** implica una transición desde agentes individuales hacia configuraciones caracterizadas por la **colaboración multiagente, la autonomía coordinada y la descomposición dinámica de tareas**. Esta transformación redefine la arquitectura del sistema, ya que la inteligencia deja de ser una propiedad individual para convertirse en una propiedad distribuida que emerge de la interacción entre agentes (Sapkota et al., 2026). Li (2026) propone un marco general en el que los agentes se configuran como entidades que integran **procesamiento, memoria y herramientas externas**, lo que les permite operar dentro de entornos distribuidos. Esta integración es fundamental para la arquitectura distribuida, ya que permite que los agentes no solo procesen información local, sino que también accedan a recursos externos y compartan información con otros agentes, ampliando así sus capacidades operativas (Li, 2026).

La combinación de estos elementos permite comprender que la arquitectura distribuida no es simplemente una cuestión de diseño técnico, sino una reconfiguración profunda de la lógica de los sistemas inteligentes. La descentralización introduce una estructura en la que los agentes pueden actuar de manera autónoma, pero también cooperar para alcanzar objetivos comunes. Esta dualidad entre autonomía y cooperación constituye uno de los principios fundamentales de los sistemas agénticos. (Karnouskos et al., 2020). La arquitectura distribuida permite abordar problemas complejos que no pueden resolverse mediante enfoques centralizados, ya que facilita la descomposición de tareas en unidades más pequeñas que pueden ser gestionadas por distintos agentes. Este enfoque no solo mejora la eficiencia del sistema, sino que también permite una mayor adaptabilidad frente a cambios en el entorno (Yadav et al., 2023).

En resumen, las **arquitecturas distribuidas y la descentralización de la agencia** configuran un modelo en el que la inteligencia emerge de la interacción entre múltiples agentes autónomos. La modularidad, la autonomía, la cooperación y la capacidad de adaptación continua constituyen los pilares de este enfoque, permitiendo la construcción de sistemas capaces de operar de manera eficiente y resiliente en entornos dinámicos. La descentralización, por tanto, no solo redefine la estructura del sistema, sino que también establece las condiciones para la emergencia de nuevas formas de inteligencia colectiva (Piccialli et al., 2025).

Coordinación, comunicación y lógica descentralizada

La **coordinación en sistemas distribuidos** constituye el mecanismo mediante el cual múltiples agentes autónomos logran alinear sus acciones sin depender de un control centralizado. A diferencia de los enfoques estructurales, donde el énfasis recae en la arquitectura del sistema, la coordinación se centra en los **procesos de interacción**, es decir, en cómo los agentes intercambian información y ajustan su comportamiento en función de otros. En este sentido, la coherencia del sistema emerge de la **sincronización de acciones distribuidas**, lo que convierte a la coordinación en un fenómeno dinámico y no estructural (Karnouskos et al., 2020).

Desde la perspectiva de la comunicación interagente, los sistemas distribuidos requieren el uso de **protocolos semánticos** que permitan a los agentes interpretar correctamente la información intercambiada. Estos protocolos no solo transmiten datos, sino que codifican significados, intenciones y estados del sistema. En este nivel, la comunicación se configura como un proceso interpretativo que posibilita la comprensión mutua entre agentes, lo que resulta esencial para la coordinación efectiva en entornos complejos (Piccialli et al., 2025).

En el ámbito del aprendizaje, la interacción entre agentes introduce una dimensión adaptativa en la que las decisiones individuales se ven influenciadas por el comportamiento colectivo. El paradigma de **multi-agent reinforcement learning** establece que cada agente ajusta su estrategia en función de las recompensas obtenidas y de las acciones de otros agentes, lo que genera un proceso de **aprendizaje interdependiente**. Este tipo de aprendizaje permite la emergencia de patrones cooperativos sin necesidad de una programación explícita (Yadav et al., 2023).

Desde una lógica funcional, la descentralización implica que los agentes deben ser capaces de **negociar objetivos y coordinar tareas** en ausencia de una autoridad central. Este proceso de negociación se basa en el intercambio de información sobre estados, capacidades y metas, lo que permite a los agentes tomar decisiones alineadas con el sistema global. La negociación, en este sentido, actúa como un mecanismo de regulación que sustituye al control jerárquico (Sapkota et al., 2026).

En el plano operativo, la coordinación se ve reforzada por la capacidad de los agentes para compartir y utilizar información contextual mediante sistemas de **memoria distribuida**. Esta memoria permite que los agentes mantengan consistencia en sus decisiones al acceder a información generada por otros agentes, facilitando la alineación de acciones en entornos dinámicos. La memoria, en este nivel, no es un componente estructural, sino un **recurso para la interacción** (Li, 2026). La lógica descentralizada introduce la necesidad de mecanismos de **resolución de conflictos**, ya que los agentes pueden tener objetivos divergentes o información incompleta. Estos mecanismos permiten identificar inconsistencias y ajustar las decisiones de manera que se preserve la coherencia del sistema. La resolución de conflictos se convierte así

en un proceso fundamental para mantener la estabilidad del comportamiento colectivo (Karnouskos et al., 2020).

En entornos caracterizados por alta incertidumbre, la coordinación debe ser capaz de adaptarse a cambios constantes en la información disponible. Esto implica que los agentes deben operar bajo una lógica de **adaptación continua**, en la que las decisiones se revisan y ajustan en función de nuevas condiciones. La coordinación, en este contexto, no es un estado fijo, sino un proceso dinámico que evoluciona en el tiempo (Yadav et al., 2023). La **coordinación, comunicación y lógica descentralizada** constituyen el núcleo de interacción en sistemas agénticos distribuidos. A diferencia de la estructura, que define la organización del sistema, estos procesos determinan cómo los agentes se relacionan entre sí para generar comportamiento colectivo. La coherencia del sistema emerge cuando los agentes logran comunicarse, negociar, aprender y resolver conflictos de manera efectiva, consolidando la descentralización como un modelo basado en la **interacción inteligente** distribuida (Sapkota et al., 2026)

Emergencia sistémica y comportamiento colectivo en sistemas descentralizados

La **emergencia sistémica en arquitecturas distribuidas** se refiere a la capacidad de los sistemas agénticos para generar comportamientos globales que no pueden explicarse a partir del análisis aislado de sus componentes. En este sentido, el comportamiento del sistema no es programado directamente, sino que surge de la interacción continua entre agentes autónomos. Esta propiedad introduce una nueva forma de comprender la inteligencia artificial, en la que la **inteligencia colectiva** emerge como resultado de dinámicas relacionales más que de estructuras individuales **(Sapkota et al., 2026).** La emergencia del comportamiento se encuentra asociada a la interacción reiterada entre agentes que operan bajo reglas locales, pero cuyos efectos se proyectan a nivel global. Este fenómeno implica que pequeñas variaciones en la **acción de un agente pueden amplificarse a través del sistema, generando patrones complejos de comportamiento**. En este contexto, la inteligencia del sistema se manifiesta como una propiedad distribuida que no puede localizarse en un punto específico (Yadav et al., 2023).

En el ámbito industrial, los sistemas ciberfísicos basados en agentes permiten que el comportamiento global emerja de la interacción entre unidades autónomas que controlan recursos específicos. Esta configuración facilita la adaptación del sistema a cambios en el entorno, ya que cada agente puede ajustar su comportamiento en función de condiciones locales, contribuyendo a la estabilidad del sistema en su conjunto. La emergencia, en este sentido, se vincula con la capacidad del sistema para **autoorganizarse (Karnouskos et al., 2020).** La emergencia del comportamiento también se relaciona con la capacidad de los sistemas para alcanzar **escalabilidad funcional**, es decir, la posibilidad de incrementar el número de agentes sin comprometer la coherencia del sistema. En arquitecturas distribuidas, la incorporación

de nuevos agentes no requiere rediseñar el sistema, ya que la lógica de funcionamiento se basa en reglas de interacción que se mantienen consistentes independientemente del tamaño del sistema (Piccialli et al., 2025).

En un enfoque operativo, la emergencia del comportamiento implica que los agentes pueden generar soluciones a problemas complejos mediante la coordinación indirecta, sin necesidad de planificación centralizada. Este tipo de comportamiento se observa en sistemas donde los agentes responden a señales del entorno y de otros agentes, generando patrones de acción que resultan coherentes a nivel global. La **autoorganización** se convierte así en un principio fundamental de la agencia descentralizada (Sapkota et al., 2026). La emergencia sistémica introduce una dimensión temporal en el comportamiento de los agentes, ya que los patrones emergentes se desarrollan a lo largo del tiempo mediante procesos de interacción continua. Esta característica implica que el comportamiento del sistema no puede evaluarse únicamente en un momento específico, sino que debe analizarse como un proceso dinámico en evolución. La temporalidad, en este sentido, se convierte en un factor clave para comprender la inteligencia distribuida **(Yadav et al., 2023).**

En este contexto, la descentralización permite que los sistemas sean más resilientes frente a fallos o perturbaciones, ya que la pérdida de uno o varios agentes no compromete el funcionamiento global. Esta capacidad de resiliencia se deriva de la distribución de funciones entre múltiples agentes, lo que evita la existencia de puntos únicos de falla. La emergencia del comportamiento, por tanto, está estrechamente vinculada con la **robustez sistémica (Karnouskos et al., 2020).** la emergencia del comportamiento en sistemas distribuidos redefine la noción de control en la inteligencia artificial. En lugar de imponer reglas globales, el diseño se orienta a establecer condiciones que permitan la generación de comportamientos deseables a partir de interacciones locales. Esta aproximación implica que el diseñador no controla directamente el sistema, sino que configura el entorno en el que los agentes interactúan, permitiendo que la inteligencia emerja de manera autónoma **(Piccialli et al., 2025**).

Dado lo anterior, la **emergencia sistémica y el comportamiento colectivo** constituyen el resultado más avanzado de la descentralización de la agencia, donde la inteligencia no se diseña de manera directa, sino que surge de la interacción entre agentes autónomos. La **autoorganización, la escalabilidad, la resiliencia y la temporalidad** configuran las bases de este fenómeno, consolidando a las arquitecturas distribuidas como el fundamento de una nueva forma de inteligencia artificial basada en la **interacción compleja y distribuida** (Sapkota et al., 2026).

Ecosistemas agénticos y coevolución del comportamiento

Los **ecosistemas agénticos y la coevolución del comportamiento** representan una fase avanzada en la inteligencia artificial, donde múltiples agentes autónomos interactúan dentro de entornos compartidos generando dinámicas adaptativas. En estos sistemas, la inteligencia no reside en un agente individual, sino que emerge de la **interacción continua, el aprendizaje colectivo y la retroalimentación distribuida (Sapkota et al., 2026).** A través de enfoques como el **multi-agent reinforcement learning**, los agentes ajustan sus estrategias en función del comportamiento de otros, configurando procesos de **adaptación interdependiente (Yadav et al., 2023).** Asimismo, la coevolución implica que las decisiones de cada agente transforman el entorno, afectando a su vez a los demás agentes **(Wang et al., 2024).** Este fenómeno se amplifica mediante mecanismos de **refinamiento iterativo y colaboración multiagente** (Yuan & Xie, 2026). En conjunto, estos ecosistemas configuran una inteligencia emergente, dinámica y distribuida

Fundamentos de los ecosistemas agénticos y coevolución del comportamiento

Los **ecosistemas agénticos** se configuran como entornos en los que múltiples agentes autónomos interactúan de manera continua, generando dinámicas colectivas que no pueden explicarse a partir de unidades individuales. A diferencia de los sistemas tradicionales, donde la inteligencia se localiza en componentes específicos, en estos ecosistemas la inteligencia emerge como una **propiedad sistémica distribuida**, resultado de la interacción entre agentes que perciben, deciden y actúan en un entorno compartido. Esta transformación conceptual implica que el comportamiento del sistema no es predefinido, sino que se desarrolla a partir de procesos relacionales en evolución (Sapkota et al., 2026). Los ecosistemas agénticos se sostienen en arquitecturas donde las capacidades de percepción, memoria y acción se integran dentro de marcos que permiten la interacción con otros agentes. En este contexto, los sistemas basados en modelos de lenguaje han demostrado que los agentes pueden operar en entornos abiertos mediante la combinación de **memoria, planificación y ejecución**, lo que les permite adaptarse a condiciones cambiantes y participar en dinámicas colectivas complejas **(Wang et al., 2024).**

En el ámbito del aprendizaje, la coevolución del comportamiento se fundamenta en la lógica del **multi-agent reinforcement learning**, donde múltiples agentes aprenden simultáneamente a partir de su interacción con el entorno y con otros agentes. Este proceso introduce una dinámica en la que el aprendizaje es inherentemente interdependiente, ya que las decisiones de cada agente afectan las condiciones de aprendizaje de los demás. La coevolución se manifiesta así como un proceso de **adaptación mutua**, en el cual las estrategias evolucionan en función de un entorno compartido (Yadav et al., 2023). Los ecosistemas agénticos pueden entenderse como

sistemas en los que múltiples agentes participan en procesos de mejora continua mediante la interacción. En este sentido, la coevolución implica la existencia de mecanismos de **optimización distribuida**, donde cada agente ajusta su comportamiento en función de los resultados obtenidos y de las acciones de otros agentes. Esta dinámica permite que el sistema evolucione hacia configuraciones más eficientes sin necesidad de un control centralizado **(Du et al., 2026).**

A nivel de interacción, la coevolución del comportamiento se ve reforzada por procesos de **refinamiento iterativo**, en los cuales los agentes ajustan sus decisiones a partir de retroalimentación continua. Este proceso no solo mejora el desempeño individual, sino que también contribuye a la evolución del sistema en su conjunto. La incorporación de múltiples agentes en estos procesos amplía la capacidad del sistema para explorar soluciones, generando comportamientos más complejos y adaptativos **(Yuan & Xie, 2026).**

Desde una perspectiva adaptativa, los ecosistemas agénticos operan en entornos caracterizados por la incertidumbre, donde la información es incompleta o cambiante. En este contexto, la coevolución del comportamiento implica que los agentes deben ser capaces de ajustar sus estrategias en función de nuevas condiciones, lo que requiere mecanismos de **adaptación continua**. Esta capacidad permite que el sistema mantenga coherencia en su comportamiento a pesar de la variabilidad del entorno **(Chen, 2025).**

En el ámbito organizacional, los ecosistemas agénticos introducen una nueva forma de interacción entre humanos y sistemas inteligentes, donde la coevolución del comportamiento se extiende más allá de los agentes artificiales. En este sentido, la inteligencia emerge de la interacción entre diferentes tipos de actores, configurando sistemas híbridos en los que la **colaboración humano-IA** redefine la organización del trabajo y la toma de decisiones **(Raisch & Krakowski, 2021).** La coevolución del comportamiento tiene implicaciones en el diseño del trabajo, ya que la integración de sistemas inteligentes en entornos organizacionales transforma la forma en que se estructuran las tareas y las relaciones laborales. Este proceso implica la necesidad de diseñar sistemas que permitan la interacción efectiva entre agentes humanos y artificiales, favoreciendo la **adaptación socio-técnica** y la optimización conjunta del sistema **(Parker & Grote, 2022).**

Desde una perspectiva sistémica, los ecosistemas agénticos pueden entenderse como redes dinámicas en las que la inteligencia emerge de la interacción entre múltiples actores. Este enfoque permite comprender que la coevolución del comportamiento no es un fenómeno aislado, sino un proceso que involucra múltiples niveles de análisis, desde la interacción entre agentes hasta la configuración del sistema en su conjunto. En este sentido, la inteligencia colectiva se configura como una propiedad emergente de la **interacción distribuida** (Miller & Davenport, 2021).

Así, los **ecosistemas agénticos y la coevolución del comportamiento** constituyen un paradigma en el que la inteligencia se entiende como un proceso

dinámico, emergente y relacional. La interacción, el aprendizaje colectivo, la optimización distribuida y la adaptación continua configuran las bases de este enfoque, consolidando una visión en la que el comportamiento del sistema evoluciona de manera constante a partir de sus propias dinámicas internas.

Mecanismos de coevolución y dinámica adaptativa en ecosistemas agénticos

La **coevolución del comportamiento en ecosistemas agénticos** se sostiene en un conjunto de mecanismos dinámicos que permiten la transformación continua de las estrategias de los agentes en función de la interacción con otros. A diferencia de los fundamentos estructurales, este nivel se centra en los **procesos de cambio adaptativo**, mediante los cuales el comportamiento evoluciona a través de ciclos reiterados de interacción. En este sentido, la coevolución se configura como un fenómeno dinámico en el que los agentes modifican sus decisiones en función de las condiciones emergentes del sistema (Sapkota et al., 2026).

Uno de los mecanismos centrales es el **aprendizaje interdependiente**, en el cual los agentes ajustan sus estrategias considerando las acciones de otros agentes dentro del mismo entorno. Este proceso implica que el aprendizaje no se desarrolla en un entorno estático, sino en un sistema dinámico donde cada decisión altera las condiciones futuras. En este contexto, la coevolución se manifiesta como una forma de **adaptación mutua**, en la que las estrategias se transforman de manera conjunta **(Yadav et al., 2023).** La coevolución se articula mediante procesos de **ajuste iterativo distribuido**, donde los agentes refinan sus decisiones a través de múltiples ciclos de interacción. Este mecanismo permite que el sistema evolucione hacia configuraciones más eficientes sin requerir supervisión centralizada. La optimización distribuida introduce una lógica en la que las mejoras individuales contribuyen al desempeño colectivo del sistema **(Du et al., 2026).**

La coevolución se apoya en procesos de **retroalimentación continua**, en los cuales los agentes reciben información sobre el impacto de sus acciones y ajustan su comportamiento en consecuencia. Esta retroalimentación permite la evaluación constante de estrategias, generando dinámicas evolutivas que incrementan la complejidad del sistema. La iteración de este proceso es fundamental para la generación de comportamientos adaptativos (Yuan & Xie, 2026).

Otro mecanismo clave es la **adaptación contextual**, mediante la cual los agentes modifican sus decisiones en función de las condiciones específicas del entorno. Este tipo de adaptación implica la capacidad de interpretar señales cambiantes y responder de manera flexible, lo que resulta esencial en sistemas caracterizados por la incertidumbre. La coevolución, en este sentido, se configura como un proceso de ajuste continuo a contextos dinámicos (Chen, 2025).

Cognitivamente hablando, los agentes incorporan mecanismos de **reflexión y autoevaluación**, lo que les permite analizar sus decisiones y mejorar su desempeño en iteraciones posteriores. Este proceso introduce una dimensión metacognitiva en la coevolución, donde los agentes no solo actúan, sino que también aprenden de sus propios errores. La capacidad de reflexión favorece la generación de estrategias más sofisticadas y adaptativas (Du et al., 2026). La coevolución se ve influida por la capacidad de los agentes para **alinear sus decisiones con las dinámicas del sistema**, lo que implica un proceso de ajuste continuo entre intereses individuales y colectivos. Este mecanismo permite evitar divergencias que podrían afectar la coherencia del sistema, favoreciendo la estabilidad del comportamiento colectivo (Yuan & Xie, 2026).

Asimismo, la coevolución implica procesos de **coadaptación**, en los cuales los agentes y el entorno evolucionan de manera simultánea. Este fenómeno refleja que el entorno no es un elemento pasivo, sino una dimensión activa que se transforma a partir de las acciones de los agentes. La coadaptación permite comprender la evolución del sistema como un proceso bidireccional en el que agentes y entorno se influyen mutuamente (Chen, 2025). Los **mecanismos de coevolución en ecosistemas agénticos** se fundamentan en procesos de aprendizaje interdependiente, optimización distribuida, retroalimentación continua, adaptación contextual, reflexión metacognitiva y coadaptación. Estos mecanismos permiten que el comportamiento del sistema evolucione de manera dinámica, consolidando una forma de inteligencia que emerge de la interacción constante entre sus componentes.

Validación y auditoría del desempeño en sistemas agénticos

La **validación del desempeño en sistemas agénticos** se configura como el proceso mediante el cual se comprueba empíricamente que el comportamiento del agente cumple con los objetivos definidos bajo condiciones específicas de operación. Este proceso implica someter al sistema a escenarios controlados donde se analizan sus resultados frente a criterios previamente establecidos. En este sentido, la validación se concibe como una **verificación empírica del comportamiento**, donde el desempeño del agente es contrastado con expectativas definidas (Chen et al., 2025). La validación requiere el diseño de protocolos de prueba que permitan evaluar el comportamiento del agente en diferentes condiciones. Estos protocolos incluyen escenarios de prueba que simulan situaciones reales para analizar la respuesta del sistema. La validación se configura así como un **proceso experimental estructurado**, donde el desempeño se evalúa mediante pruebas controladas (Du et al., 2026).

Por otro lado, a validación del desempeño implica analizar la capacidad del agente para mantener resultados consistentes frente a variaciones en las condiciones de entrada. Este proceso permite determinar si el sistema es capaz de operar de manera estable en diferentes contextos. La validación se configura así como una **evaluación de robustez del comportamiento**, donde se analiza la resistencia del sistema ante cambios (Yuan & Xie, 2026). Desde la perspectiva de la auditoría, la evaluación del desempeño incluye procesos que permiten revisar de manera sistemática el

comportamiento del agente para identificar posibles desviaciones respecto a los estándares definidos. Esta auditoría implica analizar registros de operación y resultados obtenidos, lo que permite verificar la coherencia del sistema. La auditoría se configura así como un **proceso de revisión estructurada del desempeño**, donde se examina la conformidad del comportamiento (Sawant, 2025).

A partir del ámbito de la certificación, la validación del desempeño permite establecer si el sistema cumple con criterios que lo hacen apto para su implementación en entornos reales. Este proceso implica comparar el desempeño del agente con estándares establecidos por organismos o marcos de referencia. La certificación se configura así como un **proceso de acreditación del desempeño**, donde se valida la calidad del sistema (World Economic Forum, 2025).

Desde la perspectiva de la reproducibilidad, la validación del desempeño implica garantizar que los resultados obtenidos por el agente puedan replicarse en condiciones similares. Este proceso permite asegurar que el comportamiento del sistema no es producto de condiciones específicas o aleatorias. La reproducibilidad se configura así como una **propiedad clave de la validación**, donde los resultados deben ser consistentes en distintos contextos (Chen et al., 2025).

A nivel de la trazabilidad, la auditoría del desempeño requiere la capacidad de rastrear las decisiones del agente para comprender cómo se generaron los resultados. Este proceso implica analizar los pasos que condujeron a una acción específica, lo que permite evaluar la coherencia del sistema. La trazabilidad se configura así como un **mecanismo de seguimiento del comportamiento**, donde se analiza la lógica detrás de las decisiones (Du et al., 2026). La validación del desempeño implica que los resultados y procesos del agente sean comprensibles y accesibles para su análisis. Este proceso permite garantizar que el comportamiento del sistema pueda ser evaluado de manera clara. La transparencia se configura así como una **condición de auditabilidad del sistema**, donde el desempeño puede ser interpretado de manera objetiva (Yuan & Xie, 2026).

A partir de la evaluación externa, la auditoría del desempeño implica la revisión del sistema por parte de entidades independientes que permiten validar su funcionamiento desde una perspectiva objetiva. Este proceso fortalece la credibilidad del sistema, ya que introduce un nivel adicional de verificación. La evaluación externa se configura así como un **proceso de validación independiente**, donde el desempeño se examina fuera del entorno de desarrollo (World Economic Forum, 2025). La validación y auditoría del desempeño consolidan un marco riguroso que permite garantizar la calidad del comportamiento de los sistemas agénticos. La integración de pruebas, revisión, certificación y evaluación externa permite construir sistemas confiables y verificables. En este sentido, la evaluación del desempeño se configura como el **proceso que asegura la validez y confiabilidad del agente**, consolidando su implementación en entornos reales (Sawant, 2025).

Emergencia, estabilidad y evolución sistémica en ecosistemas agénticos

La **emergencia sistémica en ecosistemas agénticos** constituye el fenómeno mediante el cual surgen patrones de comportamiento colectivo que no pueden explicarse a partir del análisis individual de los agentes, sino como resultado de su interacción continua. En este contexto, la inteligencia no se localiza en un componente específico, sino que se manifiesta como una **propiedad emergente distribuida**, generada por dinámicas relacionales que evolucionan en el tiempo. Esta perspectiva redefine la noción de agencia al situarla en el nivel del sistema, donde el comportamiento colectivo adquiere autonomía respecto a sus componentes individuales (Sapkota et al., 2026).

Desde el punto de vista del aprendizaje, la emergencia se encuentra vinculada con la capacidad de los sistemas multiagente para generar comportamientos adaptativos a partir de la interacción reiterada entre agentes. Este proceso implica que el comportamiento global no es diseñado directamente, sino que se configura progresivamente a través de dinámicas de interacción. La emergencia se entiende así como el resultado de una **adaptación colectiva sostenida**, en la cual los agentes contribuyen a la construcción de patrones coherentes en el sistema (Yadav et al., 2023). La estabilidad del ecosistema se construye mediante la interacción entre procesos que regulan la variabilidad del comportamiento. Los sistemas agénticos deben ser capaces de evolucionar sin perder coherencia operativa, lo que implica mantener un equilibrio entre cambio y consistencia. Este equilibrio se logra mediante mecanismos de **regulación distribuida**, en los cuales las decisiones individuales se ajustan en función de las dinámicas globales del sistema (Du et al., 2026).

La emergencia del comportamiento colectivo se ve influida por la repetición de ciclos de comunicación y ajuste entre agentes. Estos ciclos permiten la formación de patrones que, al consolidarse, generan estructuras dinámicas dentro del sistema. La interacción continua actúa como un mecanismo de **autoorganización**, mediante el cual el sistema desarrolla configuraciones coherentes sin necesidad de control centralizado (Yuan & Xie, 2026). Adaptativamente, la evolución del ecosistema se caracteriza por la capacidad de los agentes para responder a cambios en el entorno mediante la modificación de sus estrategias. Este proceso implica que el sistema no solo se adapta, sino que también se transforma a lo largo del tiempo, incorporando nuevas formas de comportamiento. La evolución se configura así como un proceso de **transformación adaptativa continua**, donde el sistema se redefine en función de sus propias dinámicas internas (Chen, 2025).

Desde una perspectiva organizacional, la emergencia y evolución de los ecosistemas agénticos se extiende a entornos híbridos en los que humanos y agentes artificiales interactúan. En este contexto, la estabilidad del sistema depende de la capacidad de integrar diferentes formas de inteligencia dentro de un marco común. La interacción entre actores humanos y sistemas inteligentes genera nuevas dinámicas

organizacionales que influyen directamente en la evolución del ecosistema (Raisch & Krakowski, 2021). La estabilidad del ecosistema requiere considerar el diseño del trabajo y las estructuras organizacionales que permiten la interacción entre agentes. En este sentido, la estabilidad no debe entenderse como rigidez, sino como la capacidad del sistema para mantener coherencia mientras evoluciona. La incorporación de principios de **adaptación socio-técnica** permite que la tecnología y la organización se transformen de manera conjunta, favoreciendo la sostenibilidad del sistema (Parker & Grote, 2022) .

En el plano sistémico, la evolución de los ecosistemas agénticos se caracteriza por trayectorias no lineales, donde pequeños cambios pueden generar efectos significativos en el comportamiento global. Esta característica implica que la evolución del sistema no sigue patrones predecibles, sino que se desarrolla a través de procesos complejos de interacción y ajuste. La inteligencia colectiva se configura así como una propiedad emergente de la **interacción distribuida y evolutiva**, donde el sistema se transforma continuamente (Miller & Davenport, 2021) .

Otro aspecto fundamental es la **resiliencia sistémica**, entendida como la capacidad del ecosistema para mantener su funcionamiento frente a perturbaciones. En sistemas distribuidos, la ausencia de un punto único de control permite que el sistema continúe operando incluso cuando algunos agentes fallan. Esta resiliencia se deriva de la distribución de funciones y de la capacidad de los agentes para reorganizarse en respuesta a cambios, lo que fortalece la estabilidad del sistema (Sapkota et al., 2026).

La emergencia, estabilidad y evolución sistémica se configuran como dimensiones interdependientes que permiten comprender el funcionamiento de los ecosistemas agénticos en su nivel más avanzado. La interacción entre autoorganización, regulación distribuida, transformación adaptativa y resiliencia configura un sistema capaz de evolucionar de manera coherente en entornos complejos. En este sentido, la inteligencia no se diseña de manera directa, sino que emerge como resultado de la **interacción dinámica, distribuida y evolutiva** entre los componentes del sistema (Sapkota et al., 2026).

Conclusiones

El Capítulo 3 desarrolla una comprensión profunda del diseño de la inteligencia artificial agéntica, estableciendo que este proceso representa un cambio paradigmático respecto a los enfoques tradicionales de la inteligencia artificial. En lugar de centrarse en la optimización de funciones aisladas o en la eficiencia de componentes individuales, el diseño agéntico se orienta hacia la **organización del comportamiento como eje central de la inteligencia**, lo que redefine los fundamentos conceptuales y operativos de los sistemas inteligentes .

En este contexto, diseñar sistemas agénticos implica ir más allá de la simple configuración técnica de módulos, para enfocarse en la creación de estructuras

capaces de sostener **coherencia operativa en entornos dinámicos y cambiantes**. Esto supone integrar de manera funcional los procesos de percepción, decisión, acción y memoria dentro de una lógica continua, donde el comportamiento no es un resultado secundario, sino la estructura primaria que define al sistema. **El diseño se convierte así en un proceso estructural que articula múltiples dimensiones en un flujo operativo orientado a objetivos**.

Uno de los aportes más relevantes del capítulo es el enfoque de diseño basado en la organización del comportamiento, el cual introduce una lógica en la que los agentes operan mediante ciclos iterativos de planificación, ejecución, evaluación y ajuste. Este enfoque permite que el comportamiento sea **dinámico, adaptativo y emergente**, en contraste con los sistemas tradicionales caracterizados por rigidez y determinismo. Además, se reconoce que este comportamiento no solo es individual, sino también colectivo, ya que en sistemas multiagente la inteligencia emerge de la interacción coordinada entre múltiples entidades.

El capítulo también enfatiza la importancia de la **integración estructural como criterio fundamental del diseño agéntico**. A diferencia de los enfoques modulares clásicos, donde los componentes operan de manera relativamente independiente, los sistemas agénticos requieren una articulación funcional que permita la coordinación dinámica de procesos. **La coherencia del sistema no depende de sus partes aisladas, sino de la forma en que estas interactúan dentro de una estructura integrada**. Esta integración se extiende tanto al interior del agente como a su interacción con otros agentes y con el entorno.

De manera complementaria, la **adaptabilidad y la coherencia** se presentan como principios esenciales y complementarios del diseño. La adaptabilidad permite al sistema ajustarse a condiciones cambiantes, mientras que la coherencia garantiza la continuidad y consistencia del comportamiento. **El diseño agéntico se configura como un equilibrio dinámico entre cambio y estabilidad**, lo que permite al sistema evolucionar sin perder dirección operativa.

En el ámbito del modelado, el capítulo establece que el comportamiento debe formalizarse como un sistema de estados operativos y procesos de decisión, integrando variables como contexto, memoria, objetivos y restricciones. **El comportamiento deja de ser una simple respuesta y se convierte en una estructura diseñada que organiza la dinámica del agente**, permitiendo su análisis, control y optimización. Asimismo, la evaluación y validación del comportamiento se consolidan como procesos fundamentales para garantizar la calidad, robustez y confiabilidad del sistema.

El diseño funcional del agente se organiza en múltiples niveles —operativo, estructural, cognitivo, paradigmático y aplicado—, lo que permite alcanzar **máxima ortogonalidad conceptual y funcional**. Esta estructura evita redundancias, fortalece la claridad del modelo y permite una comprensión integral del sistema.

Finalmente, el capítulo introduce las arquitecturas distribuidas y los ecosistemas agénticos como una evolución avanzada del diseño, donde la inteligencia emerge de la interacción entre múltiples agentes autónomos. **La descentralización, la coordinación y la coevolución del comportamiento permiten la construcción de inteligencia colectiva**, configurando sistemas más adaptativos, resilientes y escalables.

Así, el Capítulo 3 establece que el diseño de la IA agéntica es una **configuración estructural compleja orientada a la organización del comportamiento**, sentando las bases para el desarrollo de sistemas inteligentes capaces de operar de manera autónoma, coherente y adaptativa en entornos dinámicos y distribuidos. Ver **Tabla 3.**

Tabla 3. Diseño de la IA agéntica

Concepto	Definición	Diferenciación	Ventajas	Limitaciones	Referencias
Diseño agéntico	Proceso de configuración estructural orientado a organizar el comportamiento del sistema	Se diferencia de la arquitectura al centrarse en la lógica de construcción y no en la estructura	Permite diseñar sistemas coherentes y orientados a objetivos	Alta complejidad conceptual	Russell & Norvig (2022); Bandi et al. (2025)
Diseño basado en comportamiento	Enfoque que prioriza la organización del comportamiento sobre funciones aisladas	A diferencia del diseño tradicional, no parte de módulos sino de dinámicas conductuales	Permite adaptación y coherencia operativa	Difícil formalización inicial	Sapkota et al. (2026); Wang (2025)
Ciclo operativo del agente	Secuencia iterativa de percepción, decisión, acción y evaluación	Se diferencia de procesos lineales por su carácter continuo y dinámico	Permite aprendizaje y ajuste constante	Requiere monitoreo continuo	Russell (2019); Poole & Mackworth (2017)
Diseño orientado a objetivos	Configuración del sistema en función de metas que guían el comportamiento	A diferencia de sistemas reactivos, introduce dirección operativa	Permite acciones coherentes y alineadas	Dependencia de correcta definición de objetivos	Russell & Norvig (2022); Acharya et al. (2025)
Modelado del comportamiento	Representación formal del comportamiento mediante estados, decisiones y acciones	Se diferencia de modelado de datos al centrarse en dinámica operativa	Permite simulación y predicción	Complejidad en entornos variables	Sapkota et al. (2026); Zhang et al. (2025)
Integración funcional en diseño	Articulación de componentes para sostener	Se diferencia de integración arquitectónica	Garantiza consistencia del sistema	Alta dificultad de implementación	Bandi et al. (2025);

	coherencia del comportamiento	al centrarse en lógica operativa			Guidotti et al. (2018)
Adaptabilidad estructurada	Capacidad del sistema de ajustarse sin perder coherencia	Se diferencia de adaptación reactiva por su organización interna	Permite operar en entornos dinámicos	Riesgo de desviación del objetivo	Wang (2025); Vinuesa et al. (2020)
Evaluación del comportamiento	Proceso de validación de la coherencia y desempeño del sistema	Se diferencia de evaluación de resultados al centrarse en procesos	Permite mejorar diseño y confiabilidad	Requiere criterios complejos	Guidotti et al. (2018); Adadi & Berrada (2018)
Niveles de diseño agéntico	Estructuración en capas: operativo, estructural, cognitivo, paradigmático y aplicado	Se diferencia de modelos planos por su enfoque multinivel	Permite claridad y ortogonalidad	Incrementa complejidad conceptual	Bandi et al. (2025); Sapkota et al. (2026)
Diseño multiagente	Configuración de sistemas donde múltiples agentes interactúan	Se diferencia del diseño individual por su enfoque colectivo	Permite inteligencia distribuida	Problemas de coordinación	Rahwan et al. (2019); Vinuesa et al. (2020)
Ecosistemas agénticos	Entornos donde agentes interactúan, evolucionan y se adaptan	Se diferencia de sistemas cerrados por su carácter abierto	Permite aprendizaje colectivo	Difícil control del sistema	World Economic Forum (2025); OECD (2026)

CAPÍTULO 4. Implementación de sistemas agénticos

El tránsito desde la conceptualización hacia la **implementación de sistemas agénticos** representa un punto crítico en la consolidación de la inteligencia artificial contemporánea. Mientras que los capítulos anteriores han establecido los fundamentos teóricos, arquitectónicos y de diseño, este capítulo se enfoca en la materialización operativa de los agentes en entornos reales, donde la autonomía, la adaptabilidad y la interacción distribuida deben traducirse en sistemas funcionales, robustos y escalables. En este contexto, la implementación no constituye únicamente una fase técnica, sino una **integración sistémica de componentes, procesos y entornos de ejecución** que permiten la operación efectiva de agentes inteligentes .

La literatura reciente destaca que los sistemas agénticos se despliegan mediante arquitecturas complejas que integran **memoria, planificación, razonamiento y ejecución**, lo que permite abordar problemas dinámicos en entornos abiertos y altamente variables . Asimismo, los avances en optimización de agentes basados en modelos de lenguaje han demostrado que la implementación requiere mecanismos de ajuste continuo, incluyendo aprendizaje por refuerzo, refinamiento iterativo y estrategias híbridas que mejoran el desempeño en tareas complejas.

En el ámbito aplicado, los sistemas multiagente han sido identificados como una solución clave para la gestión de sistemas industriales y ciberfísicos, donde la distribución de funciones permite incrementar la eficiencia, la resiliencia y la capacidad de respuesta ante condiciones cambiantes . En este sentido, la implementación implica no solo el desarrollo de agentes individuales, sino la coordinación de múltiples entidades dentro de **ecosistemas distribuidos capaces de operar de manera autónoma y colaborativa** .

Sin embargo, este proceso también introduce desafíos relevantes, particularmente en términos de integración, escalabilidad, validación y gobernanza. La incorporación de marcos regulatorios y principios éticos se vuelve esencial para garantizar que los sistemas agénticos operen de manera segura, transparente y alineada con los valores sociales . De este modo, la implementación se configura como un proceso multidimensional que articula tecnología, organización y regulación, consolidando el paso definitivo hacia la operacionalización de la inteligencia artificial agéntica.

Ciclo de vida del agente

El **ciclo de vida del agente** en sistemas agénticos se concibe como una **estructura operativa iterativa** que articula la **percepción, el procesamiento, la toma de decisiones y la acción** en un flujo continuo. A diferencia de modelos lineales, este ciclo funciona mediante **iteraciones constantes**, donde cada interacción con el entorno actualiza el estado interno del agente y redefine su comportamiento. En este proceso, la **memoria y el razonamiento** permiten contextualizar la información, mientras que la **planificación y ejecución** transforman los datos en acciones coherentes. Asimismo, el ciclo incorpora mecanismos de **adaptación y aprendizaje**, lo que posibilita la mejora progresiva del desempeño en función de la experiencia acumulada. Esta dinámica convierte al agente en un sistema capaz de **operar de manera autónoma en entornos cambiantes**, manteniendo coherencia y eficiencia. En síntesis, el ciclo de vida constituye el **núcleo funcional de los sistemas agénticos**, ya que integra continuidad, adaptación y decisión dentro de una misma lógica operativa.

Estructuración del ciclo de vida del agente

El **ciclo de vida del agente en sistemas agénticos** se define como la estructura operativa que organiza la interacción continua entre el agente y su entorno, permitiendo transformar información en comportamiento mediante procesos iterativos. Este ciclo no responde a una lógica lineal, sino a una dinámica recurrente en la que cada estado operativo se actualiza constantemente a partir de nuevas entradas del entorno. En este sentido, el ciclo de vida constituye una **configuración dinámica de operación**, donde la **continuidad**, la **iteración** y la **actualización permanente del estado** son esenciales para la funcionalidad del sistema (Sawant, 2025). El ciclo de vida se sostiene en la integración de múltiples componentes funcionales que permiten la operación del agente. Estos componentes incluyen módulos de **percepción**,

memoria, **razonamiento** y **acción**, los cuales deben articularse de manera coherente para generar comportamiento inteligente. La integración de estos elementos configura una **estructura operativa interdependiente**, donde el funcionamiento del agente depende de la coordinación efectiva entre sus partes, evitando fragmentaciones funcionales que comprometan su desempeño (Piccialli et al., 2025).

En el plano del procesamiento, el ciclo de vida implica la actualización constante del estado interno del agente a partir de la información recibida del entorno. Este proceso permite que el agente mantenga una representación dinámica de su contexto, lo que resulta fundamental para la toma de decisiones en entornos cambiantes. La actualización continua introduce una lógica de **procesamiento adaptativo**, en la cual el comportamiento no es fijo, sino que se redefine en función de nuevas condiciones, consolidando una capacidad de respuesta flexible (Wang et al., 2024). El ciclo de vida incorpora mecanismos que permiten transformar información en decisiones estructuradas. Estos mecanismos incluyen procesos de **planificación**, **evaluación de alternativas** y **selección de acciones**, lo que permite al agente operar con autonomía en escenarios complejos. El comportamiento se configura así como el resultado de una **deliberación estructurada**, donde las decisiones emergen de la integración de múltiples variables y no de respuestas automáticas (Ozdemir, 2026).

El ciclo de vida a nivel de interacción, se amplía en entornos donde múltiples agentes operan simultáneamente, lo que introduce dinámicas de coordinación entre ciclos individuales. En estos contextos, la acción de un agente influye en el comportamiento de otros, generando una red de interdependencias que transforma el ciclo de vida en un proceso relacional. Esta dimensión colectiva permite comprender el sistema como una **configuración distribuida de comportamiento**, donde la inteligencia emerge de la interacción entre múltiples entidades (Karnouskos et al., 2020). El ciclo de vida incorpora mecanismos que permiten modificar el comportamiento del agente en función de la experiencia acumulada. Este proceso implica que cada iteración del ciclo contribuye al aprendizaje del agente, lo que mejora su desempeño en el tiempo. La adaptación se convierte así en un elemento estructural del ciclo de vida, consolidando una lógica de **evolución continua del comportamiento**, donde el agente ajusta sus respuestas con base en resultados previos (Yadav et al., 2023).

Además, el ciclo de vida incluye mecanismos de control que permiten verificar la coherencia entre las decisiones del agente y sus objetivos operativos. Estos mecanismos introducen una dimensión de evaluación dentro del ciclo, lo que permite ajustar el comportamiento en tiempo real. La integración de control y ejecución configura el ciclo de vida como un proceso de **autorregulación del sistema**, donde el agente mantiene consistencia, estabilidad y alineación con sus metas (Sayyad et al., 2024).

En el plano sistémico, el ciclo de vida se consolida como una estructura que permite la continuidad operativa del agente en entornos complejos. La combinación de percepción, procesamiento, decisión y acción dentro de un flujo iterativo permite

construir un sistema capaz de operar de manera autónoma. Esta estructura dinámica configura el ciclo de vida como el **núcleo funcional de los sistemas agénticos**, donde la inteligencia emerge de la interacción continua entre el agente y su entorno (Du et al., 2026). El ciclo de vida del agente no solo define la operación interna del sistema, sino que establece las condiciones para su escalabilidad y transferencia a distintos contextos de aplicación. La capacidad de mantener coherencia operativa en diferentes entornos permite que los agentes se adapten a múltiples dominios, lo que amplía su utilidad en aplicaciones industriales, científicas y sociales. En este sentido, el ciclo de vida se configura como una **estructura generalizable del comportamiento inteligente**, capaz de sostener la operación de sistemas complejos en escenarios diversos (Li, 2026).

Dinámica operativa y ejecución del ciclo de vida

La **dinámica operativa del ciclo de vida del agente** se define por la ejecución continua de procesos que permiten mantener una interacción constante con el entorno, donde cada acción genera nuevas condiciones que deben ser evaluadas en iteraciones posteriores. Esta dinámica implica que el agente no opera mediante eventos aislados, sino dentro de un flujo permanente de actualización, en el que la información se transforma en comportamiento de manera progresiva. En este sentido, el ciclo de vida se configura como una **estructura de ejecución iterativa**, donde la continuidad operativa es esencial para la funcionalidad del sistema (Du et al., 2026). La la dinámica del ciclo de vida se articula a través de una secuencia de fases interdependientes que incluyen **percepción, análisis, decisión y acción**, las cuales operan de manera coordinada. Estas fases no deben entenderse como etapas separadas, sino como componentes de un sistema integrado en el que la salida de una fase se convierte en la entrada de la siguiente. Esta articulación permite construir una **secuencia operativa coherente**, donde el comportamiento emerge de la interacción estructurada entre procesos internos (Maldonado et al., 2024).

En el plano del procesamiento en tiempo real, la dinámica operativa se caracteriza por la capacidad del agente para responder de manera inmediata a cambios en el entorno. Este procesamiento implica integrar información externa con datos internos, lo que permite generar respuestas adaptativas en contextos dinámicos. La ejecución en tiempo real introduce una lógica de **respuesta inmediata y contextualizada**, fundamental para la operación eficiente de sistemas agénticos en escenarios complejos (Wang et al., 2024). La ejecución del ciclo de vida se apoya en mecanismos que permiten evaluar alternativas y seleccionar acciones de manera estructurada. Estos mecanismos incluyen procesos de inferencia, análisis de contexto y priorización de decisiones, lo que permite al agente actuar con un alto grado de autonomía. El comportamiento se configura así como resultado de una **toma de decisiones dinámica**, donde las acciones se ajustan continuamente a las condiciones del entorno (Li, 2026).

En el ámbito de la coordinación, la dinámica operativa se complejiza en sistemas donde múltiples agentes interactúan simultáneamente, lo que introduce la necesidad

de sincronizar ciclos individuales. En estos contextos, la ejecución de un agente depende no solo de su estado interno, sino también del comportamiento de otros agentes, lo que genera dinámicas de interdependencia. Esta coordinación configura una **dinámica colectiva del comportamiento**, donde la estabilidad del sistema depende de la interacción entre sus componentes (Karnouskos et al., 2020). La dinámica del ciclo de vida incorpora mecanismos que permiten ajustar el comportamiento en función de los resultados obtenidos en iteraciones previas. Este proceso implica que la ejecución no es estática, sino que evoluciona en función de la experiencia acumulada. La adaptación introduce una dimensión de **aprendizaje continuo**, donde el agente mejora su desempeño mediante la incorporación de información histórica (Yadav et al., 2023).

En el plano de la optimización, la dinámica operativa del ciclo de vida se orienta a mejorar la eficiencia del agente en términos de tiempo de respuesta, uso de recursos y precisión en la toma de decisiones. Este proceso implica la incorporación de estrategias que permitan reducir la incertidumbre y maximizar el desempeño del sistema. La optimización se configura así como un elemento clave dentro de la ejecución, consolidando una lógica de **mejora progresiva del comportamiento** (Ozdemir, 2026). Se debe apreciar que la dinámica del ciclo de vida incluye mecanismos de supervisión que permiten evaluar el desempeño del agente durante la ejecución. Estos mecanismos introducen una dimensión de control en tiempo real, lo que permite detectar desviaciones y corregir el comportamiento de manera inmediata. La supervisión continua refuerza la estabilidad del sistema, garantizando la coherencia entre las decisiones y los objetivos del agente (Sayyad et al., 2024).

En conclusión, la dinámica operativa del ciclo de vida se configura como un proceso continuo de ejecución, evaluación y ajuste que permite al agente operar de manera autónoma en entornos complejos. Esta dinámica integra múltiples procesos dentro de una estructura coherente, consolidando una forma de inteligencia basada en la **iteración constante, la adaptación y la coordinación**, donde el comportamiento se construye a partir de la interacción continua con el entorno (Sawant, 2025).

Evolución, monitoreo y control del ciclo de vida

La **evolución del ciclo de vida del agente** constituye un componente esencial para garantizar la sostenibilidad operativa de los sistemas agénticos en entornos dinámicos. A diferencia de sistemas estáticos, los agentes requieren mecanismos que les permitan modificar su comportamiento a lo largo del tiempo, integrando procesos de aprendizaje, ajuste y mejora continua. En este sentido, el ciclo de vida se configura como una **estructura evolutiva del comportamiento**, donde cada iteración contribuye al perfeccionamiento de las decisiones del agente (Sawant, 2025).El ciclo de vida incorpora mecanismos que permiten observar el comportamiento del agente durante su ejecución. Este proceso implica la recopilación sistemática de datos sobre el desempeño, lo que permite identificar patrones, detectar anomalías y evaluar la consistencia del comportamiento. La capacidad de monitoreo introduce una dimensión

de **observabilidad del sistema**, fundamental para la gestión de entornos complejos donde el comportamiento debe ser constantemente evaluado (Zhang et al., 2025).

El ciclo de vida a nivel del control, incluye mecanismos que permiten intervenir en el comportamiento del agente cuando se detectan desviaciones respecto a los objetivos establecidos. Estas intervenciones pueden implicar ajustes en los parámetros de decisión, modificación de estrategias o reconfiguración de procesos internos. El control se convierte así en un elemento clave para garantizar la coherencia del sistema, configurando una lógica de **regulación del comportamiento inteligente** (Du et al., 2026).

La evolución del ciclo de vida se ve influida por la capacidad del agente para aprender de su experiencia. Este aprendizaje permite ajustar las decisiones futuras en función de resultados previos, lo que introduce una dimensión de mejora progresiva en el comportamiento. La evolución se configura así como un proceso de **aprendizaje acumulativo**, donde el agente refina sus estrategias a través de la iteración continua (Yadav et al., 2023). En el ámbito sistémico, la evolución del ciclo de vida se amplía en entornos multiagente, donde la interacción entre agentes influye en la dinámica global del sistema. En estos contextos, el monitoreo y control deben considerar no solo el comportamiento individual, sino también las dinámicas colectivas que emergen de la interacción. Esta situación introduce una lógica de **control distribuido**, donde la estabilidad del sistema depende de la coordinación entre múltiples agentes (Karnouskos et al., 2020).

Desde la perspectiva de la validación, el ciclo de vida incorpora mecanismos que permiten evaluar el desempeño del agente en términos de eficiencia, precisión y cumplimiento de objetivos. Este proceso de validación permite identificar áreas de mejora y ajustar el comportamiento en función de métricas definidas. La validación se configura así como un componente esencial del ciclo de vida, consolidando una lógica de **evaluación continua del desempeño** (Maldonado et al., 2024).

En el plano ético, el monitoreo y control del ciclo de vida adquieren relevancia en términos de responsabilidad, transparencia y gobernanza. La capacidad de supervisar el comportamiento del agente permite garantizar que sus acciones se alineen con principios normativos y valores sociales. En este sentido, el ciclo de vida no solo es un proceso técnico, sino también un mecanismo de **gobernanza del comportamiento inteligente**, que permite gestionar riesgos y asegurar la confianza en los sistemas agénticos (Hahn et al., 2026). La evolución del ciclo de vida implica la incorporación de mecanismos de resiliencia que permitan al agente operar de manera estable frente a perturbaciones. Estos mecanismos garantizan la continuidad del comportamiento incluso en condiciones adversas, fortaleciendo la robustez del sistema. La resiliencia se convierte así en un elemento clave dentro del ciclo de vida, configurando una lógica de **estabilidad adaptativa del sistema** (Piccialli et al., 2025).

Por último, la evolución, el monitoreo y el control del ciclo de vida consolidan una estructura operativa que permite a los sistemas agénticos adaptarse, mejorar y

sostener su funcionamiento en el tiempo. Este proceso integra observación, intervención y aprendizaje dentro de una dinámica continua, configurando una forma avanzada de inteligencia basada en la **capacidad de evolucionar de manera autónoma en entornos complejos**. De esta manera, el ciclo de vida se posiciona como el mecanismo central que garantiza la viabilidad, eficiencia y confiabilidad de los sistemas agénticos (Li, 2026).

Integración tecnológica

La **integración tecnológica en sistemas agénticos** se refiere al proceso mediante el cual los agentes inteligentes se **conectan, interactúan y operan dentro de infraestructuras digitales y entornos reales**, trascendiendo su lógica interna. Este proceso implica la vinculación con **sistemas externos, servicios digitales, bases de datos y plataformas tecnológicas**, lo que permite ampliar sus capacidades funcionales. Asimismo, la integración incluye el despliegue del agente en **infraestructuras como cloud, edge computing o entornos distribuidos**, garantizando escalabilidad y disponibilidad. Más allá del plano técnico, también supone su inserción en **entornos socio-técnicos**, donde interactúa con usuarios y procesos organizacionales. En este sentido, la integración tecnológica no es solo una cuestión de conexión, sino una **articulación entre inteligencia, infraestructura y contexto operativo**, que permite al agente actuar de manera efectiva en escenarios complejos. Así, se configura como el **puente entre la capacidad del agente y su aplicación en el mundo real**.

Integración del agente con sistemas y herramientas externas

La **integración del agente con sistemas y herramientas externas** constituye el proceso mediante el cual los sistemas agénticos trascienden su operación interna para interactuar con entornos digitales reales. Esta integración permite que el agente no solo procese información, sino que acceda a recursos externos, ejecute acciones sobre sistemas y genere efectos en contextos operativos concretos. En este sentido, la integración tecnológica se configura como una **extensión funcional del agente hacia el entorno**, donde la conexión con sistemas externos amplía sus capacidades más allá de su arquitectura interna (Maldonado et al., 2024). La integración implica la conexión del agente con **interfaces de programación de aplicaciones (APIs)** que permiten acceder a funcionalidades específicas, como bases de datos, motores de búsqueda o servicios especializados. Estas interfaces actúan como puntos de interacción que posibilitan la ejecución de tareas complejas mediante la combinación de recursos internos y externos. La utilización de APIs configura una **lógica de interoperabilidad funcional**, donde el agente opera como un coordinador de servicios distribuidos (Piccialli et al., 2025).

La integración tecnológica a nivel de la gestión de datos, permite al agente acceder, consultar y manipular información almacenada en sistemas externos. Esta capacidad implica la conexión con **bases de datos estructuradas y no estructuradas**, lo que

permite enriquecer el proceso de toma de decisiones con información contextual. El acceso a datos externos introduce una dimensión de **ampliación informacional del agente**, donde la calidad del comportamiento depende de la disponibilidad y pertinencia de los datos (Wang et al., 2024).

Operativamente, la integración con herramientas externas permite al agente ejecutar acciones en sistemas digitales, como la automatización de procesos, la gestión de tareas o la interacción con plataformas. Esta capacidad transforma al agente en un **ejecutor activo dentro de sistemas tecnológicos**, donde su comportamiento tiene efectos directos en el entorno. La ejecución externa se convierte así en un elemento clave para la implementación de sistemas agénticos en contextos reales (Li, 2026).

La integración tecnológica, en el ámbito de los sistemas distribuidos, permite que el agente interactúe con múltiples servicios de manera simultánea, lo que introduce una lógica de coordinación entre distintos recursos. Esta interacción requiere mecanismos que permitan gestionar la comunicación entre sistemas heterogéneos, asegurando la coherencia del flujo de información. La integración se configura así como una **red de conexiones funcionales**, donde el agente actúa como nodo articulador (Karnouskos et al., 2020). La integración con sistemas externos introduce desafíos relacionados con la protección de datos y la gestión de accesos. La necesidad de garantizar la integridad y confidencialidad de la información implica la incorporación de mecanismos de autenticación, autorización y control de accesos. La seguridad se convierte así en un componente esencial de la integración tecnológica, configurando una **estructura de control sobre la interacción externa del agente** (Ozdemir, 2026).

Es de considerar que, la integración tecnológica debe permitir la incorporación de nuevos servicios y herramientas sin afectar el funcionamiento del agente. Esta capacidad implica diseñar sistemas flexibles que puedan ajustarse a cambios en el entorno tecnológico, lo que resulta fundamental para su sostenibilidad. La adaptabilidad se configura así como un elemento clave de la integración, permitiendo la evolución del sistema en función de nuevas necesidades (Du et al., 2026). La integración con herramientas externas implica la necesidad de gestionar la consistencia de la información entre distintos sistemas. Esto requiere mecanismos que permitan sincronizar datos y garantizar la coherencia del comportamiento del agente. La consistencia se convierte así en un elemento fundamental para evitar errores y asegurar la confiabilidad del sistema, configurando una **lógica de coherencia intersistémica** (Sayyad et al., 2024).

Concluyendo, la integración del agente con sistemas y herramientas externas permite construir sistemas agénticos capaces de operar en entornos digitales complejos. La conexión con servicios, datos y plataformas transforma al agente en un actor que no solo procesa información, sino que interactúa activamente con el entorno. En este sentido, la integración tecnológica se consolida como el **puente entre la inteligencia del agente y su capacidad de acción en el mundo real**, permitiendo la implementación efectiva de sistemas agénticos en contextos diversos (Sawant, 2025).

Infraestructura tecnológica para el despliegue del agente

La **infraestructura tecnológica para el despliegue de sistemas agénticos** constituye el conjunto de recursos computacionales que permiten la operación efectiva del agente en entornos reales. A diferencia de la arquitectura interna del agente, la infraestructura define el entorno donde el sistema se ejecuta, incluyendo servidores, redes y plataformas que soportan su funcionamiento. En este sentido, la infraestructura se configura como la **base material de operación del agente**, donde se habilita su capacidad de interacción con el entorno digital (Du et al., 2026). la infraestructura permite desplegar agentes en entornos escalables que facilitan el acceso a recursos bajo demanda. Este modelo permite ajustar la capacidad de procesamiento en función de las necesidades del sistema, lo que resulta fundamental para aplicaciones que requieren alta disponibilidad. El uso de la nube configura una **infraestructura elástica**, donde el agente puede operar de manera continua sin restricciones físicas (Li, 2026).

En el plano del *edge computing*, la infraestructura se orienta a procesar información cerca del punto donde se generan los datos, lo que reduce la latencia y mejora la capacidad de respuesta del sistema. Este enfoque resulta especialmente relevante en entornos donde la inmediatez es crítica, como sistemas industriales o dispositivos conectados. El *edge computing* configura una **infraestructura distribuida de baja latencia**, que permite optimizar la ejecución del agente en tiempo real (Karnouskos et al., 2020). La infraestructura tecnológica depende de redes que permiten la comunicación entre los distintos componentes del sistema. Estas redes deben garantizar la transferencia eficiente de datos, lo que resulta esencial para la operación del agente. La conectividad se configura así como un elemento clave dentro de la infraestructura, permitiendo la **integración de sistemas distribuidos** (Wang et al., 2024).

La infraestructura permite coordinar múltiples nodos que operan de manera conjunta para soportar el funcionamiento del agente. Esta coordinación implica la gestión de recursos distribuidos, lo que permite construir sistemas escalables y resilientes. La infraestructura distribuida se configura así como una **red de soporte para sistemas agénticos**, donde la operación no depende de un único punto (Piccialli et al., 2025). La infraestructura tecnológica incluye sistemas que permiten gestionar grandes volúmenes de datos necesarios para la operación del agente. Estos sistemas deben garantizar la disponibilidad y consistencia de la información, lo que resulta fundamental para la toma de decisiones. El almacenamiento se configura así como un componente clave para la **gestión eficiente de información** dentro del sistema (Maldonado et al., 2024).

En el plano de la seguridad, la infraestructura debe incorporar mecanismos que protejan el sistema frente a amenazas externas. Esto incluye la implementación de protocolos de seguridad, sistemas de autenticación y mecanismos de control de accesos. La seguridad se convierte así en un elemento esencial de la infraestructura, configurando una **estructura de protección del sistema agéntico** (Ozdemir, 2026). De estra forma la infraestructura tecnológica debe permitir la expansión del sistema sin

comprometer su funcionamiento. Esto implica diseñar entornos que puedan adaptarse al crecimiento del número de agentes y de la carga de trabajo. La escalabilidad se configura así como una propiedad fundamental de la infraestructura, permitiendo la evolución del sistema en el tiempo (Sawant, 2025).

Asimismo, la infraestructura debe garantizar la disponibilidad del sistema, lo que implica la implementación de mecanismos que permitan mantener la operación incluso en caso de fallas. Esto incluye la redundancia de componentes y la capacidad de recuperación ante errores. La disponibilidad se configura así como un elemento clave para la **continuidad operativa del agente** (Du et al., 2026).

Finalmente, la infraestructura tecnológica para el despliegue del agente constituye el entorno que permite materializar la operación de los sistemas agénticos en contextos reales. La combinación de cómputo, conectividad, almacenamiento y seguridad permite construir sistemas robustos y eficientes. En este sentido, la infraestructura se consolida como el **soporte fundamental que habilita la implementación de agentes inteligentes en entornos complejos**, garantizando su funcionamiento, escalabilidad y sostenibilidad (Li, 2026).

Inserción del agente en entornos socio-técnicos reales

La **inserción de agentes en entornos socio-técnicos** representa la fase en la cual los sistemas agénticos dejan de operar únicamente en el plano tecnológico para integrarse en contextos donde interactúan con **usuarios, procesos organizacionales y estructuras sociales**. Este proceso implica que el agente no solo ejecute tareas, sino que participe activamente en sistemas donde la toma de decisiones está influida por factores humanos, institucionales y contextuales. En este sentido, la integración se configura como una **interacción entre tecnología y organización**, donde el agente se convierte en un actor dentro de sistemas complejos (Hahn et al., 2026). Desde la perspectiva de la interacción humano-agente, la integración tecnológica implica diseñar mecanismos que permitan la comunicación efectiva entre usuarios y sistemas agénticos. Esta interacción debe garantizar que el comportamiento del agente sea comprensible, predecible y alineado con las expectativas humanas. La relación humano-agente se configura así como una **interfaz socio-técnica**, donde la confianza y la interpretabilidad del sistema son elementos fundamentales para su aceptación (Piccialli et al., 2025).

La inserción del agente a nivel organizacional, implica su integración dentro de procesos de trabajo existentes, lo que requiere adaptar su funcionamiento a dinámicas institucionales. Este proceso implica que el agente no actúe de manera aislada, sino que se integre en flujos de trabajo donde interactúa con otros sistemas y actores humanos. La integración organizacional se configura así como una **articulación entre automatización y estructura institucional**, donde el agente contribuye a la eficiencia de los procesos (Maldonado et al., 2024).

Si se toma en cuenta la toma de decisiones, la inserción del agente en entornos socio-técnicos implica que sus acciones pueden influir directamente en resultados organizacionales. Esto introduce la necesidad de garantizar que las decisiones del agente sean coherentes con objetivos estratégicos y valores institucionales. En este contexto, el agente se configura como un **actor decisional dentro del sistema**, cuya operación debe ser supervisada y alineada con criterios definidos (Li, 2026). La integración socio-técnica permite la interacción entre múltiples agentes y humanos dentro de un mismo entorno. Esta interacción configura sistemas híbridos donde las capacidades humanas y tecnológicas se complementan. La colaboración se convierte así en un elemento central de la integración, configurando una **dinámica de cooperación humano-máquina**, donde el valor del sistema emerge de la combinación de capacidades (Karnouskos et al., 2020).

Éticamente, la inserción del agente en entornos socio-técnicos introduce desafíos relacionados con la responsabilidad, la transparencia y la equidad. La capacidad del agente para influir en decisiones humanas implica la necesidad de establecer mecanismos que garanticen el uso responsable de la tecnología. La ética se configura así como un componente fundamental de la integración, permitiendo gestionar los impactos sociales del sistema (Hahn et al., 2026). Por ejemplo, en el plano de la aceptación social, la integración de agentes en entornos reales depende de la percepción que los usuarios tienen del sistema. Factores como la confianza, la facilidad de uso y la percepción de utilidad influyen en la adopción de la tecnología. La aceptación se configura así como un elemento clave para la implementación, donde el éxito del sistema depende de su integración efectiva en el contexto social (Sawant, 2025).

Desde un enfoque adaptativo, la inserción del agente en entornos socio-técnicos requiere la capacidad de ajustarse a cambios en las condiciones organizacionales y sociales. Esto implica que el agente debe ser capaz de modificar su comportamiento en función de nuevas reglas, procesos o necesidades. La adaptabilidad se convierte así en un elemento esencial para la sostenibilidad del sistema en contextos dinámicos (Du et al., 2026). La integración socio-técnica implica la necesidad de establecer mecanismos de gobernanza que permitan supervisar el comportamiento del agente dentro del sistema. Estos mecanismos permiten garantizar que la operación del agente se mantenga alineada con normas y objetivos establecidos. La gobernanza se configura así como una **estructura de control del sistema**, donde la supervisión es clave para la estabilidad del entorno (Ozdemir, 2026).

Por lo tanto, la inserción del agente en entornos socio-técnicos consolida la transición de los sistemas agénticos desde el plano tecnológico hacia el plano operativo real. La interacción con usuarios, organizaciones y contextos sociales permite construir sistemas que no solo ejecutan tareas, sino que participan activamente en la configuración de procesos y decisiones. En este sentido, la integración socio-técnica se configura como el **factor que permite la materialización de la inteligencia agéntica en contextos reales**, consolidando su impacto en sistemas complejos (Wang et al., 2024).

Funcionamiento en entorno real

El **funcionamiento en entorno real de los sistemas agénticos** se caracteriza por la capacidad del agente para operar en contextos donde predominan la **incertidumbre, la variabilidad y la interacción con múltiples sistemas**. A diferencia de entornos controlados, el agente debe procesar información heterogénea, tomar decisiones bajo condiciones cambiantes y ejecutar acciones con impacto directo en sistemas operativos. Este funcionamiento implica una **adaptación continua al contexto**, donde el comportamiento se ajusta en función de restricciones externas como calidad de datos, latencia o dependencias tecnológicas. Asimismo, el agente interactúa con procesos organizacionales y plataformas digitales, integrándose como un actor operativo dentro de estructuras complejas. En este sentido, el funcionamiento en entorno real no solo depende de la capacidad interna del agente, sino de su habilidad para **responder, coordinarse y mantenerse estable dentro de condiciones dinámicas**, consolidando su papel como componente activo en sistemas socio-técnicos reales.

Operación del agente en contextos reales

El **funcionamiento de sistemas agénticos en entornos reales** implica la capacidad del agente para operar en contextos donde las condiciones no están completamente controladas, lo que introduce variables como incertidumbre, variabilidad y restricciones externas. En este sentido, el agente debe ser capaz de ejecutar acciones en escenarios donde la información es incompleta o dinámica, lo que transforma su operación en un proceso de **adaptación contextual continua**. Esta capacidad permite que los sistemas agénticos se desempeñen en aplicaciones reales, donde la interacción con el entorno define la efectividad del comportamiento (Collaco et al., 2026). El funcionamiento en condiciones reales requiere que el agente procese información proveniente de múltiples fuentes, incluyendo sensores, sistemas digitales y entradas humanas. Esta diversidad de información implica la necesidad de gestionar datos heterogéneos que pueden variar en calidad y disponibilidad. En este contexto, el agente desarrolla una **capacidad de interpretación contextual**, donde la percepción del entorno se convierte en un elemento clave para la toma de decisiones (Chen et al., 2025).

El funcionamiento en entorno real se manifiesta en sistemas donde los agentes operan en sectores como la industria, la salud o los servicios digitales, donde las decisiones tienen consecuencias directas. En estos escenarios, el agente no solo procesa información, sino que ejecuta acciones que afectan sistemas reales, lo que introduce una dimensión de responsabilidad operativa. Esta situación configura al agente como un **actor funcional dentro de sistemas complejos**, donde su

comportamiento tiene impacto directo en el entorno (Piccialli et al., 2025). El funcionamiento en entorno real implica que múltiples agentes operan simultáneamente, interactuando entre sí y con el entorno. Esta interacción genera dinámicas de coordinación que permiten resolver problemas complejos mediante la distribución de tareas. En este contexto, el sistema se configura como una **estructura colaborativa**, donde el comportamiento global emerge de la interacción entre agentes individuales (Karnouskos et al., 2020).

En el ámbito de la implementación práctica, el funcionamiento en entorno real requiere que los agentes sean capaces de operar en plataformas tecnológicas que soportan su ejecución. Estas plataformas permiten la integración de funciones que facilitan la operación del agente, como el acceso a datos o la ejecución de procesos automatizados. En este sentido, el agente se configura como una **entidad operativa integrada en plataformas digitales**, donde su desempeño depende de la infraestructura que lo soporta (Alqurni, 2026). El funcionamiento en entorno real implica que el agente debe ajustar su comportamiento en función de cambios en las condiciones del entorno. Este proceso implica la capacidad de modificar decisiones y estrategias en tiempo real, lo que permite mantener la coherencia operativa. La adaptación se configura así como una **respuesta dinámica a la variabilidad del entorno**, donde el agente ajusta su comportamiento continuamente (Collaco et al., 2026).

Al tomarse en cuenta al toma de decisiones, el funcionamiento en entorno real implica que las decisiones del agente deben considerar múltiples variables que pueden cambiar de manera inesperada. Este proceso requiere integrar información diversa para generar respuestas coherentes. En este sentido, la decisión se configura como un proceso de **evaluación contextual compleja**, donde el agente debe balancear diferentes factores para actuar de manera efectiva (Chen et al., 2025). El funcionamiento en entorno real implica que las acciones del agente tienen efectos directos sobre sistemas y procesos. Esta capacidad introduce la necesidad de garantizar la coherencia del comportamiento, ya que cualquier desviación puede generar consecuencias negativas. El agente se configura así como un **actor con impacto real**, donde la precisión y la consistencia son fundamentales (Piccialli et al., 2025).

En el ámbito de la coordinación sistémica, el funcionamiento en entorno real requiere que los agentes mantengan coherencia en sus interacciones, evitando conflictos que puedan afectar la estabilidad del sistema. Esta coordinación se convierte en un elemento clave para garantizar el funcionamiento adecuado del sistema, configurando una **dinámica de equilibrio entre agentes**, donde la estabilidad depende de la interacción coordinada (Karnouskos et al., 2020). el funcionamiento de los sistemas agénticos en entornos reales consolida la transición de la inteligencia artificial desde el plano conceptual hacia la aplicación práctica. La capacidad de operar en contextos dinámicos, interactuar con sistemas reales y generar impacto operativo permite construir soluciones efectivas en escenarios complejos. En este sentido, el agente se configura como una entidad capaz de **operar, adaptarse y generar valor**

en el mundo real, consolidando su relevancia en aplicaciones contemporáneas (Alqurni, 2026).

Interacción del agente con sistemas y procesos reales

La **interacción del agente con sistemas reales** constituye el mecanismo mediante el cual los sistemas agénticos se integran en entornos operativos donde existen procesos establecidos, flujos de trabajo y estructuras tecnológicas definidas. En este contexto, el agente no actúa de manera aislada, sino que se inserta dentro de sistemas que requieren coherencia funcional para operar correctamente. Esta interacción implica que el agente participe activamente en procesos que ya existen, configurando una **integración operativa dentro de sistemas reales** (Collaco et al., 2026). La interacción del agente implica la capacidad de conectarse con plataformas que gestionan información, ejecutan procesos y coordinan actividades. Esta conexión permite que el agente no solo acceda a datos, sino que también intervenga en su procesamiento y utilización. En este sentido, la interacción se configura como una **relación funcional entre el agente y los sistemas digitales**, donde el flujo de información determina el comportamiento operativo (Chen et al., 2025).

La interacción del agente con procesos reales implica su incorporación en dinámicas de trabajo donde existen reglas, procedimientos y objetivos definidos. Este proceso requiere que el agente ajuste su comportamiento a las condiciones del sistema, evitando interferencias que puedan afectar su funcionamiento. La interacción se configura así como una **adaptación del agente a estructuras organizacionales**, donde su operación debe alinearse con los procesos existentes (Piccialli et al., 2025). La interacción con sistemas reales implica que el agente debe operar en conjunto con otros agentes o sistemas tecnológicos, lo que introduce la necesidad de sincronizar acciones. Esta sincronización permite mantener la coherencia del sistema, evitando conflictos entre componentes. La coordinación se configura así como una **dinámica de interacción controlada**, donde la estabilidad del sistema depende de la alineación entre sus partes (Karnouskos et al., 2020).

La interacción del agente implica que sus acciones tienen efectos directos sobre los sistemas en los que opera. Esto significa que el agente no solo procesa información, sino que interviene en procesos que generan resultados concretos. Esta capacidad introduce una dimensión de **acción operativa sobre sistemas reales**, donde el comportamiento del agente influye en el desempeño del sistema (Alqurni, 2026). Desde una perspectiva de adaptación, la interacción del agente con sistemas reales requiere la capacidad de ajustar su comportamiento en función de cambios en los procesos o en el entorno. Este proceso implica que el agente debe ser capaz de modificar sus decisiones para mantener la coherencia con las condiciones del sistema. La adaptación se configura así como una **respuesta operativa a cambios en sistemas reales**, donde la flexibilidad es fundamental (Collaco et al., 2026).

La interacción con sistemas reales implica que las decisiones del agente deben considerar múltiples factores relacionados con el entorno en el que opera. Estos

factores incluyen restricciones operativas, objetivos del sistema y condiciones externas. En este sentido, la toma de decisiones se configura como un proceso de **evaluación contextual dentro de sistemas reales**, donde el agente debe actuar de manera coherente con el entorno (Chen et al., 2025). Así a nivel de la eficiencia, la interacción del agente con sistemas reales implica optimizar su desempeño dentro de los procesos en los que participa. Esta optimización permite mejorar la efectividad del sistema, reduciendo errores y aumentando la productividad. La eficiencia se configura así como una **optimización del comportamiento dentro de procesos reales**, donde el agente contribuye al rendimiento del sistema (Piccialli et al., 2025).

En el ámbito de la estabilidad, la interacción del agente con sistemas reales requiere mantener un comportamiento coherente a lo largo del tiempo, evitando desviaciones que puedan afectar el funcionamiento del sistema. Esta estabilidad depende de la capacidad del agente para coordinar sus acciones con otros componentes. La estabilidad se configura así como una **propiedad emergente de la interacción sistémica**, donde el comportamiento debe mantenerse consistente (Karnouskos et al., 2020). La interacción del agente con sistemas y procesos reales consolida su papel como un componente activo dentro de entornos operativos complejos. La capacidad de integrarse, adaptarse y ejecutar acciones permite construir sistemas donde los agentes contribuyen directamente al funcionamiento del entorno. En este sentido, el agente se configura como un **actor operativo dentro de sistemas reales**, donde su comportamiento tiene impacto directo en los resultados (Alqurni, 2026).

Restricciones y condiciones del entorno real

El **funcionamiento del agente en entornos reales** está condicionado por múltiples restricciones externas que limitan su desempeño operativo, obligándolo a actuar dentro de márgenes definidos por el entorno. Estas restricciones incluyen incertidumbre, variabilidad y dependencia de factores externos que no pueden ser controlados completamente por el sistema. En este sentido, el entorno real se configura como un espacio de **condicionamiento operativo**, donde el agente debe ajustar su comportamiento para mantener coherencia frente a condiciones cambiantes (Chen et al., 2025). Los entornos reales presentan estructuras altamente interconectadas donde múltiples elementos interactúan simultáneamente, lo que introduce restricciones relacionadas con la coordinación y la sincronización. Estas condiciones dificultan la operación del agente, ya que su comportamiento debe alinearse con dinámicas colectivas que no dependen exclusivamente de su lógica interna. La interacción sistémica se convierte así en una **limitación estructural del comportamiento**, donde el agente opera dentro de redes complejas (Karnouskos et al., 2020).

Los entornos reales, a nivel de la incertidumbre, se caracterizan por la presencia de información incompleta, ambigua o cambiante, lo que limita la capacidad del agente para construir representaciones precisas del contexto. Esta situación obliga al sistema a tomar decisiones bajo condiciones de conocimiento parcial, lo que introduce riesgos en su operación. La incertidumbre se configura así como una **restricción cognitiva del agente**, donde la acción se basa en estimaciones y no en certezas (Collaco et al.,

2026). La operación del agente depende de la infraestructura disponible, lo que introduce restricciones relacionadas con capacidad de procesamiento, conectividad y disponibilidad de recursos. Estas limitaciones condicionan el tipo de tareas que el agente puede ejecutar y afectan su desempeño en entornos complejos. La infraestructura se configura así como una **restricción material del sistema**, donde el agente debe operar dentro de capacidades tecnológicas definidas (Alqurni, 2026).

En el ámbito de la calidad de los datos, el funcionamiento en entorno real implica enfrentar problemas como ruido, inconsistencias y datos incompletos, lo que afecta la precisión de la toma de decisiones. Estas condiciones limitan la capacidad del agente para interpretar correctamente el entorno, generando posibles desviaciones en el comportamiento. La calidad de los datos se configura así como una **limitación informacional crítica**, donde la confiabilidad del sistema depende de la integridad de la información (Piccialli et al., 2025). Los entornos reales presentan cambios constantes que alteran las condiciones de operación del agente. Estos cambios pueden derivar de modificaciones en los sistemas, en los datos o en el contexto operativo, lo que obliga al agente a ajustar su comportamiento de manera continua. La variabilidad se configura así como una **restricción dinámica**, donde la estabilidad depende de la capacidad de adaptación del sistema (Collaco et al., 2026).

La latencia introduce limitaciones relacionadas con el tiempo de respuesta del agente, lo que puede generar desajustes entre la percepción del entorno y la acción ejecutada. Estos retrasos afectan la sincronización del sistema, lo que puede comprometer su efectividad en entornos donde la inmediatez es crítica. La latencia se configura así como una **restricción temporal del sistema**, donde la velocidad de procesamiento influye directamente en el desempeño (Chen et al., 2025). El funcionamiento en entorno real implica que el agente interactúa con sistemas que pueden fallar o comportarse de manera impredecible. Esta dependencia introduce riesgos que afectan la estabilidad del sistema, ya que su operación puede verse comprometida por factores fuera de su control. La dependencia se configura así como una **restricción sistémica**, donde el comportamiento del agente está condicionado por la confiabilidad del entorno (Karnouskos et al., 2020).

Los entornos reales operativamente hablando, involucran múltiples variables interdependientes que dificultan la predicción de resultados. Esta complejidad obliga al agente a gestionar simultáneamente diferentes factores, lo que incrementa la dificultad de la toma de decisiones. La complejidad se configura así como una **restricción estructural del entorno**, donde el agente opera en sistemas altamente interrelacionados (Piccialli et al., 2025).

Por último, las restricciones del entorno real consolidan un marco en el cual el agente debe operar bajo condiciones que limitan su autonomía y condicionan su desempeño. La combinación de incertidumbre, limitaciones tecnológicas, variabilidad y dependencia externa configura un escenario donde la efectividad del sistema depende de su capacidad para adaptarse a estos límites. En este sentido, el funcionamiento en entorno real se configura como un proceso de **operación bajo**

restricciones, donde el agente ajusta continuamente su comportamiento para mantener coherencia en contextos complejos (Collaco et al., 2026).

Evaluación del desempeño

La **evaluación del desempeño en sistemas agénticos** se define como el proceso sistemático mediante el cual se mide, valida y analiza el comportamiento del agente en relación con objetivos previamente establecidos. Este proceso implica el uso de **métricas e indicadores** que permiten cuantificar dimensiones como la precisión, la eficiencia y la consistencia del sistema. Asimismo, la evaluación incluye mecanismos de **benchmarking y comparación**, que permiten situar el rendimiento del agente frente a estándares o sistemas equivalentes. Más allá de la medición, incorpora procesos de **validación empírica y auditoría**, donde el comportamiento es sometido a pruebas y revisiones para garantizar su confiabilidad. En este sentido, la evaluación no solo mide resultados, sino que asegura la calidad del desempeño mediante verificación, reproducibilidad y transparencia. Así, se configura como el **mecanismo central para garantizar la efectividad, confiabilidad y mejora continua del comportamiento del agente** en contextos operativos reales.

Fundamentos de la evaluación del desempeño en sistemas agénticos

La **evaluación del desempeño en sistemas agénticos** se configura como un proceso estructurado orientado a medir la efectividad del comportamiento del agente en función de objetivos previamente establecidos. Este proceso implica definir criterios claros que permitan valorar aspectos como la precisión de las decisiones, la coherencia del comportamiento y la capacidad de cumplir metas específicas. En este sentido, la evaluación se concibe como una **estructura formal de medición del rendimiento**, donde el comportamiento del agente se traduce en indicadores observables y comparables (Sawant, 2025). La evaluación del desempeño requiere alinearse con estándares que permitan garantizar la comparabilidad y validez de los resultados en diferentes contextos. Estos marcos establecen parámetros que permiten medir la calidad del sistema en términos de confiabilidad, eficiencia y cumplimiento de objetivos. La evaluación se configura así como un **proceso de estandarización del desempeño**, donde los resultados se analizan en relación con criterios globales que permiten validar su consistencia (World Economic Forum, 2025).

En el plano metodológico, la evaluación implica transformar el comportamiento del agente en datos cuantificables que permitan su análisis objetivo. Este proceso requiere la definición de métricas que capturen dimensiones clave del desempeño, como la calidad de las decisiones, la velocidad de respuesta y la precisión de las acciones. La medición se configura así como una **traducción del comportamiento en indicadores cuantificables**, lo que permite evaluar el desempeño de manera sistemática (Chen et al., 2025). La evaluación permite determinar el grado en que el agente cumple con los objetivos definidos dentro del sistema. Este análisis implica comparar los resultados

obtenidos con metas previamente establecidas, lo que permite identificar desviaciones y áreas de mejora. La evaluación se configura así como un **proceso de verificación del cumplimiento**, donde el desempeño se mide en función de resultados concretos y verificables (Du et al., 2026).

En el ámbito del benchmarking, la evaluación del desempeño permite comparar el comportamiento del agente con otros sistemas o con versiones previas del mismo sistema. Este proceso facilita la identificación de mejoras en el rendimiento y la detección de posibles deficiencias. El benchmarking se configura así como una **herramienta de comparación estructurada**, donde el desempeño se analiza en relación con referentes que permiten contextualizar los resultados (Yuan & Xie, 2026). La evaluación implica verificar que el agente mantenga un comportamiento estable a lo largo del tiempo. Esta estabilidad es fundamental para garantizar la confiabilidad del sistema, ya que permite asegurar que el comportamiento no presenta variaciones significativas que afecten su desempeño. La evaluación se configura así como un **control de estabilidad del comportamiento**, donde se analiza la variabilidad del desempeño en diferentes condiciones (Sawant, 2025).

En el plano de la validez de las métricas, la evaluación del desempeño requiere garantizar que los indicadores utilizados reflejen de manera precisa el comportamiento del agente. Este proceso implica validar que las métricas sean pertinentes y representativas del desempeño real del sistema. La evaluación se configura así como un **proceso de validación de indicadores**, donde se asegura que las métricas utilizadas sean adecuadas para medir el comportamiento (Chen et al., 2025). La evaluación requiere que los resultados obtenidos sean reproducibles y consistentes en diferentes contextos. Este proceso permite garantizar que la medición del desempeño no esté sujeta a variaciones arbitrarias. La confiabilidad se configura así como una **propiedad esencial de la evaluación**, donde los resultados deben ser estables y verificables (World Economic Forum, 2025).

En el plano de la objetividad, la evaluación del desempeño requiere minimizar la influencia de factores subjetivos en la medición del comportamiento del agente. Esto implica utilizar métricas claras y definidas que permitan evaluar el desempeño de manera imparcial. La evaluación se configura así como un **proceso objetivo de medición**, donde los resultados se basan en datos y no en interpretaciones subjetivas (Du et al., 2026). La evaluación del desempeño permite comprender el comportamiento del agente dentro del sistema en el que opera, analizando cómo sus acciones contribuyen al logro de objetivos globales. Este enfoque permite integrar la evaluación del agente dentro de un contexto más amplio, donde el desempeño se analiza en relación con el sistema completo. La evaluación se configura así como un **proceso integral de análisis del comportamiento**, donde se considera la interacción entre el agente y su entorno evaluativo (Yuan & Xie, 2026).

Se concluye por tanto, que los fundamentos de la evaluación del desempeño consolidan un marco conceptual que permite medir de manera rigurosa el comportamiento de los sistemas agénticos. La integración de criterios, métricas,

validación y análisis permite construir un proceso de evaluación robusto y confiable. En este sentido, la evaluación se configura como el **mecanismo central para garantizar la calidad del desempeño del agente**, asegurando su efectividad en contextos complejos y dinámicos (Sawant, 2025).

Métricas e indicadores para la evaluación del desempeño

La **evaluación del desempeño en sistemas agénticos** se fundamenta en la definición de métricas que permitan medir de manera objetiva el comportamiento del agente en relación con los objetivos establecidos. Estas métricas constituyen instrumentos que traducen la operación del sistema en valores cuantificables, permitiendo analizar su rendimiento en distintas condiciones. En este sentido, las métricas se configuran como una **base cuantitativa de la evaluación**, donde el comportamiento del agente se representa mediante indicadores medibles (Du et al., 2026). La evaluación requiere identificar variables que reflejen el grado de cumplimiento de los objetivos del sistema. Estos indicadores permiten analizar dimensiones específicas del comportamiento, como precisión, eficiencia y consistencia. La evaluación se configura así como un **sistema estructurado de indicadores**, donde cada métrica aporta información relevante sobre el desempeño del agente (Sawant, 2025).

Una de las métricas fundamentales es la capacidad del agente para generar decisiones correctas en función de la información disponible. Esta métrica permite evaluar la calidad del comportamiento en términos de resultados obtenidos frente a resultados esperados. La precisión se configura así como un **indicador de exactitud del comportamiento**, donde se mide la efectividad de las decisiones (Chen et al., 2025). La evaluación del desempeño implica medir el uso de recursos necesarios para ejecutar las acciones del agente. Este análisis considera variables como tiempo de respuesta, consumo computacional y optimización de procesos. La eficiencia se configura así como un **indicador de rendimiento operativo**, donde se analiza la relación entre recursos utilizados y resultados obtenidos (Yuan & Xie, 2026).

Las métricas permiten evaluar la estabilidad del comportamiento del agente a lo largo del tiempo. Este análisis implica medir la variabilidad del desempeño en diferentes condiciones, lo que permite determinar la confiabilidad del sistema. La consistencia se configura así como un **indicador de estabilidad del desempeño**, donde se evalúa la uniformidad del comportamiento (Sawant, 2025). Las métricas permiten establecer comparaciones entre diferentes sistemas o configuraciones del agente. Este proceso facilita la identificación de mejoras o deficiencias en el desempeño. La evaluación se configura así como un **proceso de benchmarking estructurado**, donde los resultados se analizan en relación con referentes definidos (World Economic Forum, 2025).

En el plano del análisis de datos, las métricas permiten interpretar el comportamiento del agente a partir de los resultados obtenidos durante su operación. Este análisis permite identificar patrones, tendencias y posibles áreas de mejora. La

evaluación se configura así como un **proceso analítico del desempeño**, donde los datos se utilizan para comprender el comportamiento del sistema (Chen et al., 2025). Las métricas permiten supervisar el desempeño del agente durante su operación, facilitando la detección de desviaciones respecto a los objetivos establecidos. Este proceso permite mantener la coherencia del sistema mediante ajustes en tiempo real. La evaluación se configura así como un **mecanismo de monitoreo del desempeño**, donde las métricas permiten controlar el comportamiento del agente (Du et al., 2026).

En el ámbito de la validez, las métricas deben ser representativas del comportamiento real del agente, lo que implica asegurar que los indicadores utilizados reflejen adecuadamente el desempeño del sistema. Este proceso permite garantizar la pertinencia de la evaluación. La validez se configura así como una **propiedad fundamental de las métricas**, donde se asegura su adecuación para medir el comportamiento (Sawant, 2025). La estandarización en las las métricas, deben definirse de manera que permitan su aplicación en distintos contextos, lo que facilita la comparación de resultados. Este proceso implica establecer criterios uniformes para la medición del desempeño. La estandarización se configura así como un **elemento clave de la evaluación**, donde se garantiza la comparabilidad de los resultados (World Economic Forum, 2025). La evaluación del desempeño requiere combinar diferentes indicadores para obtener una visión completa del comportamiento del agente. Este enfoque permite analizar múltiples dimensiones del desempeño de manera simultánea. La evaluación se configura así como un **proceso multidimensional de medición**, donde se integran distintas métricas para evaluar el sistema (Yuan & Xie, 2026).

Las métricas e indicadores constituyen el núcleo de la evaluación del desempeño en sistemas agénticos, permitiendo medir, analizar y comparar el comportamiento del agente de manera objetiva. La correcta definición y aplicación de estas métricas permite construir sistemas más eficientes, confiables y optimizados. En este sentido, la evaluación se consolida como un **proceso riguroso de medición del desempeño**, donde los indicadores permiten comprender y mejorar el comportamiento del agente (Du et al., 2026).

Validación, control y mejora del desempeño

La **validación del desempeño en sistemas agénticos** se configura como el proceso mediante el cual se verifica que el comportamiento del agente cumple con los objetivos establecidos dentro del sistema. Este proceso implica comparar los resultados obtenidos con los criterios definidos previamente, lo que permite determinar si el agente actúa de manera adecuada. En este sentido, la validación se concibe como una **confirmación sistemática del rendimiento**, donde el desempeño se evalúa en función de su alineación con metas específicas (Chen et al., 2025). La evaluación del desempeño incorpora mecanismos que permiten supervisar el comportamiento del agente durante su operación. Este control implica monitorear continuamente las acciones del agente para detectar posibles desviaciones respecto a los objetivos establecidos. El control se configura así como un **proceso de regulación del**

comportamiento, donde se asegura la coherencia del sistema mediante la observación constante (Yuan & Xie, 2026).

La evaluación del desempeño, en la mejora continua, permite identificar áreas donde el comportamiento del agente puede optimizarse. Este proceso implica analizar los resultados obtenidos para ajustar estrategias futuras, lo que contribuye a mejorar el rendimiento del sistema. La mejora continua se configura así como un **proceso evolutivo del desempeño**, donde el sistema se adapta y optimiza de manera progresiva (Du et al., 2026). La evaluación, como factor de retroalimentación, genera información que se utiliza para ajustar el comportamiento del agente en función de los resultados obtenidos. Este proceso permite mejorar la calidad de las decisiones y reducir errores en la ejecución. La retroalimentación se configura así como un **mecanismo de ajuste del comportamiento**, donde los resultados guían la mejora del sistema (Sawant, 2025).

La validación del desempeño implica analizar el rendimiento del agente en términos de uso de recursos y cumplimiento de objetivos. Este análisis permite identificar oportunidades para optimizar el comportamiento del sistema. La eficiencia se configura así como un **criterio central de la evaluación**, donde se busca maximizar el rendimiento del agente (World Economic Forum, 2025). Así, la validación del desempeño requiere verificar que el agente mantenga un comportamiento estable a lo largo del tiempo. Esta estabilidad es fundamental para garantizar la confiabilidad del sistema, ya que permite asegurar que el comportamiento no presenta variaciones significativas. La consistencia se configura así como una **propiedad del desempeño validado**, donde se evalúa la uniformidad del comportamiento (Sawant, 2025).

La evaluación del desempeño, como monitoreo contínuo, implica la observación constante del comportamiento del agente para detectar posibles desviaciones. Este monitoreo permite realizar ajustes en tiempo real, lo que contribuye a mantener la coherencia del sistema. El monitoreo se configura así como un **proceso de supervisión continua**, donde el desempeño se evalúa de manera permanente (Du et al., 2026). Así, la validación del desempeño implica asegurar que los resultados obtenidos sean consistentes y reproducibles en diferentes condiciones. Este proceso permite garantizar que el sistema opera de manera estable y predecible. La confiabilidad se configura así como una **propiedad clave del desempeño**, donde el comportamiento del agente debe ser verificable (Chen et al., 2025).

La evaluación del desempeño, a nivel de gobernanza, permite establecer mecanismos que regulen el comportamiento del agente dentro de un marco definido. Estos mecanismos permiten asegurar que el agente actúe de acuerdo con criterios establecidos, evitando desviaciones que puedan afectar el sistema. La gobernanza se configura así como un **proceso de control estructurado del desempeño**, donde se establecen límites para la operación del agente (World Economic Forum, 2025). La validación del desempeño permite evaluar cómo el comportamiento del agente influye en el sistema en el que opera. Este análisis permite comprender la relación entre el desempeño del agente y los resultados del sistema. La evaluación se configura así

como un **proceso de análisis del impacto del desempeño**, donde se estudian los efectos del comportamiento (Yuan & Xie, 2026).

Finalomente, la validación, el control y la mejora del desempeño consolidan un marco integral que permite garantizar la calidad del comportamiento de los sistemas agénticos. La combinación de verificación, monitoreo, retroalimentación y optimización permite construir sistemas confiables y eficientes. En este sentido, la evaluación del desempeño se configura como el **mecanismo central para asegurar la calidad del comportamiento del agente**, consolidando su efectividad en entornos complejos (Sawant, 2025).

Gestión de riesgos

La gestión de riesgos en sistemas agénticos se define como el **proceso integral mediante el cual se identifican, evalúan, priorizan y mitigan las amenazas que pueden afectar el comportamiento y funcionamiento del agente en entornos reales**. Este proceso articula distintas fases analíticas y operativas que permiten comprender la exposición del sistema frente a eventos adversos, considerando tanto su probabilidad como su impacto. **La gestión del riesgo no se limita a la prevención, sino que implica una intervención estructurada sobre las condiciones que generan vulnerabilidad**, integrando dimensiones tecnológicas, éticas, operativas y de gobernanza. Asimismo, incorpora mecanismos de control y supervisión que permiten mantener el comportamiento del agente dentro de límites aceptables, garantizando su coherencia y estabilidad. En este sentido, la gestión del riesgo se configura como un **sistema dinámico de regulación del comportamiento**, orientado a anticipar, reducir y controlar amenazas en contextos complejos e interconectados

Identificación y tipología de riesgos en sistemas agénticos

La **identificación de riesgos en sistemas agénticos** constituye el proceso mediante el cual se reconocen las amenazas potenciales que pueden afectar el funcionamiento del agente en entornos reales. Este proceso implica analizar las condiciones operativas del sistema para detectar vulnerabilidades que puedan comprometer su comportamiento. En este sentido, la identificación se configura como un **mapeo estructurado de riesgos**, donde se delimitan las fuentes de amenaza que pueden incidir en el desempeño del agente (Amanchala, 2024). los sistemas agénticos operan en entornos altamente interconectados donde la dependencia tecnológica introduce múltiples fuentes de riesgo. Estas fuentes incluyen fallas en redes, plataformas y sistemas distribuidos, lo que incrementa la exposición a amenazas externas. En este contexto, la identificación del riesgo se configura como un **análisis de vulnerabilidades tecnológicas**, donde se examinan los puntos críticos del sistema (Kshetri, 2023).

La identificación de riesgos, éticamente implica reconocer las posibles consecuencias del comportamiento del agente sobre usuarios y sistemas. Estos

riesgos surgen de la capacidad del agente para tomar decisiones que pueden generar efectos no deseados. En este sentido, la identificación se configura como un **reconocimiento de riesgos éticos**, donde se consideran las implicaciones morales del comportamiento del sistema (Floridi & Sanders, 2004). La identificación de riesgos requiere alinearse con marcos regulatorios que permiten clasificar las amenazas en función de su naturaleza y relevancia. Estos marcos proporcionan una base para estructurar el análisis del riesgo, facilitando su comprensión. La identificación se configura así como un **proceso de clasificación normativa del riesgo**, donde se organizan las amenazas en categorías definidas (OECD, 2021).

Los sistemas agénticos, a nivel de gobernanza, operan en contextos donde la regulación aún está en desarrollo, lo que introduce riesgos asociados a la falta de control institucional. Esta situación implica que los agentes pueden enfrentarse a escenarios donde las amenazas no están completamente gestionadas. La identificación se configura así como un **reconocimiento de riesgos de gobernanza**, donde se analizan condiciones de incertidumbre regulatoria (World Economic Forum, 2025). Operativamente, los riesgos pueden manifestarse en forma de errores en la ejecución de tareas, lo que afecta directamente el desempeño del agente. Estos errores pueden derivarse de fallas en los datos, en los modelos o en los procesos internos del sistema. La identificación permite reconocer estos escenarios, configurando una **tipología de riesgos operativos**, donde se analizan posibles desviaciones del comportamiento esperado (Amanchala, 2024).

La interdependencia entre sistemas introduce riesgos que no pueden analizarse de manera aislada, ya que una falla en un componente puede afectar el funcionamiento global. Esta situación implica que el agente opera dentro de redes complejas donde los riesgos se propagan entre sistemas. La identificación se configura así como un **análisis de riesgos sistémicos**, donde se consideran las interacciones entre componentes (Kshetri, 2023). La identificación de riesgos implica organizar las amenazas en categorías que permitan su análisis posterior. Estas categorías pueden incluir riesgos técnicos, éticos, operativos y de gobernanza, lo que facilita su comprensión. La identificación se configura así como un **proceso de estructuración del riesgo**, donde se establecen tipologías claras para su gestión (OECD, 2021).

La identificación de riesgos permite anticipar posibles impactos negativos derivados del comportamiento del agente, lo que contribuye a prevenir daños en el entorno. Este proceso implica considerar no solo las consecuencias técnicas, sino también las implicaciones sociales. La identificación se configura así como un **proceso preventivo de análisis ético**, donde se evalúan posibles efectos del sistema (Floridi & Sanders, 2004). Así, la identificación y tipología de riesgos constituyen la base para la gestión efectiva del riesgo en sistemas agénticos, ya que permiten delimitar el conjunto de amenazas que deben ser analizadas y gestionadas. La capacidad de reconocer y clasificar riesgos facilita el desarrollo de estrategias posteriores para su evaluación y mitigación. En este sentido, la identificación se configura como el **fundamento de la seguridad del sistema**, permitiendo anticipar escenarios de riesgo en entornos complejos (World Economic Forum, 2025).

Evaluación y priorización del riesgo en sistemas agénticos

La **evaluación del riesgo en sistemas agénticos** se configura como el proceso mediante el cual se analiza la magnitud de las amenazas previamente identificadas, considerando tanto su probabilidad de ocurrencia como el impacto potencial sobre el sistema. Este proceso permite transformar el riesgo en un objeto analizable, facilitando su comprensión dentro de un marco estructurado. En este sentido, la evaluación se concibe como un **análisis sistemático del riesgo**, donde se dimensiona la exposición del sistema frente a eventos adversos (OECD, 2021). La evaluación del riesgo implica examinar las vulnerabilidades presentes en las infraestructuras digitales que soportan al agente, identificando los puntos donde el sistema es más susceptible a fallas o ataques. Este análisis permite comprender la relación entre vulnerabilidad y exposición al riesgo, configurando un **análisis de vulnerabilidad tecnológica**, donde se determina la fragilidad del sistema frente a amenazas externas (Kshetri, 2023).

La evaluación del riesgo implica analizar las posibles consecuencias del comportamiento del agente en términos de impacto sobre usuarios, organizaciones y sistemas. Este análisis permite determinar la gravedad del riesgo más allá de su dimensión técnica, considerando implicaciones sociales y morales. La evaluación se configura así como un **análisis del impacto ético del riesgo**, donde se valoran los efectos de las decisiones del agente (Floridi & Sanders, 2004). La evaluación permite analizar cómo los riesgos afectan el desempeño del agente en la ejecución de tareas, identificando escenarios donde el sistema puede fallar o desviarse de su comportamiento esperado. Este análisis permite comprender la relación entre riesgo y desempeño, configurando un **análisis del impacto operativo del riesgo**, donde se estudian las consecuencias sobre la ejecución (Amanchala, 2024).

La evaluación del riesgo implica considerar el contexto institucional en el que opera el sistema, incluyendo factores como regulación, supervisión y control. Este análisis permite comprender la exposición del sistema a riesgos derivados de la falta de gobernanza. La evaluación se configura así como un **análisis contextual del riesgo**, donde se consideran condiciones externas que afectan el funcionamiento del agente (World Economic Forum, 2025).

Desde la perspectiva probabilística, la evaluación implica estimar la frecuencia con la que una amenaza puede materializarse, lo que permite clasificar los riesgos en función de su recurrencia. Este proceso facilita la comprensión del riesgo en términos de su posibilidad de ocurrencia. La evaluación se configura así como un **análisis probabilístico del riesgo**, donde se estima la frecuencia de eventos adversos (OECD, 2021). La evaluación implica analizar la severidad de las consecuencias que un riesgo puede generar sobre el sistema, lo que permite diferenciar entre riesgos críticos y menores. Este análisis permite dimensionar la gravedad de las amenazas, configurando un **análisis de impacto del riesgo**, donde se establecen niveles de afectación (Kshetri, 2023).

Desde la perspectiva de la priorización, la evaluación permite ordenar los riesgos en función de su criticidad, combinando la probabilidad y el impacto para determinar cuáles requieren atención inmediata. Este proceso facilita la asignación de recursos para la gestión del riesgo. La evaluación se configura así como un **proceso de jerarquización del riesgo**, donde se establecen niveles de prioridad (World Economic Forum, 2025). La evaluación del riesgo implica analizar la interacción entre diferentes riesgos dentro del sistema, considerando cómo pueden amplificarse o mitigarse entre sí. Este análisis permite identificar riesgos complejos que no pueden analizarse de manera aislada. La evaluación se configura así como un **análisis sistémico del riesgo**, donde se estudian interdependencias entre amenazas (Amanchala, 2024).

La evaluación del riesgo requiere integrar diferentes dimensiones del análisis —probabilidad, impacto, contexto y vulnerabilidad— para construir una visión completa del riesgo. Este enfoque permite evitar análisis fragmentados y facilita la toma de decisiones informadas. La evaluación se configura así como un **proceso multidimensional de análisis del riesgo**, donde se integran distintas variables para comprender la exposición del sistema (Floridi & Sanders, 2004). La evaluación y priorización del riesgo permiten establecer una base sólida para la gestión del riesgo en sistemas agénticos, ya que proporcionan una comprensión clara de la magnitud de las amenazas y su relevancia. La capacidad de analizar, dimensionar y jerarquizar riesgos facilita la toma de decisiones estratégicas. En este sentido, la evaluación se configura como el **núcleo analítico de la gestión del riesgo**, donde se determina qué amenazas deben ser atendidas con mayor urgencia dentro del sistema (World Economic Forum, 2025).

Mitigación, control y gobernanza del riesgo en sistemas agénticos

La **mitigación del riesgo en sistemas agénticos** se configura como el proceso mediante el cual se diseñan e implementan estrategias orientadas a reducir la probabilidad de ocurrencia de amenazas o minimizar su impacto sobre el sistema. Este proceso implica intervenir directamente sobre las condiciones que generan vulnerabilidad, permitiendo mejorar la seguridad del agente en entornos reales. En este sentido, la mitigación se concibe como una **estrategia activa de reducción del riesgo**, donde se busca disminuir la exposición del sistema frente a eventos adversos (Amanchala, 2024). La mitigación del riesgo implica la implementación de mecanismos de seguridad diseñados para proteger al sistema frente a amenazas externas, tales como ataques cibernéticos, fallas en redes o accesos no autorizados. Estos mecanismos incluyen controles de autenticación, cifrado de datos y sistemas de supervisión continua. La mitigación se configura así como una **protección tecnológica estructurada**, donde se fortalecen las defensas del sistema para reducir vulnerabilidades (Kshetri, 2023).

La mitigación del riesgo implica establecer criterios que regulen el comportamiento del agente para evitar impactos negativos sobre usuarios y sistemas. Este proceso

requiere definir límites claros en la toma de decisiones del agente, considerando las consecuencias de sus acciones. La mitigación se configura así como una **responsabilidad ética del sistema**, donde se busca prevenir daños derivados del comportamiento del agente (Floridi & Sanders, 2004). Desde la perspectiva normativa, la mitigación del riesgo requiere alinearse con marcos regulatorios que establecen estándares para la operación segura de sistemas inteligentes. Estos marcos permiten garantizar que el sistema cumpla con requisitos de seguridad, transparencia y responsabilidad. La mitigación se configura así como un **proceso regulado del riesgo**, donde la implementación de normas contribuye a la estabilidad del sistema (OECD, 2021). La mitigación del riesgo implica el diseño de políticas que regulen el comportamiento del agente dentro de contextos organizacionales e institucionales. Estas políticas permiten establecer límites operativos y mecanismos de supervisión que garantizan la seguridad del sistema. La mitigación se configura así como un **marco de gobernanza del riesgo**, donde se establecen reglas para la operación del agente en entornos complejos (World Economic Forum, 2025).

La mitigación del riesgo implica ajustar el comportamiento del agente para reducir la probabilidad de fallas en la ejecución de tareas. Este proceso incluye la optimización de procesos internos y la mejora en la calidad de las decisiones del sistema. La mitigación se configura así como una **optimización del comportamiento seguro**, donde se reducen desviaciones que pueden comprometer el desempeño (Amanchala, 2024). En el plano del control, la gestión del riesgo requiere implementar mecanismos que permitan supervisar el comportamiento del agente para asegurar que se mantenga dentro de límites aceptables. Este control implica establecer sistemas de seguimiento que permitan detectar desviaciones y corregirlas oportunamente. La mitigación se configura así como un **proceso de control del riesgo**, donde se regula el comportamiento del sistema (Kshetri, 2023).

Desde la perspectiva sistémica, la mitigación del riesgo implica considerar la interdependencia entre sistemas, lo que permite gestionar la propagación de fallas dentro de redes complejas. Este enfoque permite diseñar estrategias que reduzcan el impacto de fallas en componentes individuales. La mitigación se configura así como un **control sistémico del riesgo**, donde se gestionan interacciones entre elementos del sistema (OECD, 2021). En el ámbito preventivo, la mitigación implica anticipar posibles escenarios de riesgo antes de que se materialicen, lo que permite reducir la exposición del sistema a amenazas. Este proceso incluye la identificación de señales tempranas de riesgo y la implementación de medidas preventivas. La mitigación se configura así como una **estrategia preventiva del riesgo**, donde se busca evitar eventos adversos antes de su ocurrencia (Floridi & Sanders, 2004).

Desde la perspectiva de la resiliencia, la mitigación del riesgo contribuye a fortalecer la capacidad del sistema para resistir y recuperarse de eventos adversos. Este proceso implica diseñar sistemas robustos que puedan mantener su funcionamiento incluso en condiciones adversas. La mitigación se configura así como una **construcción de resiliencia del sistema**, donde se mejora la capacidad de respuesta ante fallas (World Economic Forum, 2025).

La mitigación, el control y la gobernanza del riesgo consolidan un enfoque integral que permite garantizar la seguridad y estabilidad de los sistemas agénticos. La combinación de estrategias tecnológicas, éticas y regulatorias permite construir sistemas confiables y robustos capaces de operar en entornos complejos. En este sentido, la gestión del riesgo se configura como el **mecanismo central para asegurar la operación segura del agente**, consolidando su funcionamiento dentro de condiciones controladas (Amanchala, 2024).

Conclusiones

El Capítulo 4 consolida la transición desde los marcos conceptuales, arquitectónicos y de diseño hacia la **materialización operativa de los sistemas agénticos en entornos reales**, estableciendo que la implementación constituye una fase crítica donde la inteligencia artificial demuestra su viabilidad práctica. A diferencia de las etapas previas, centradas en la abstracción y estructuración del comportamiento, la implementación implica la **integración efectiva del agente dentro de contextos dinámicos, inciertos y socio-técnicos**, donde su desempeño debe ser evaluado en condiciones reales de operación .

Uno de los elementos fundamentales del capítulo es la conceptualización del **ciclo de vida del agente como núcleo funcional de la implementación**, el cual articula procesos iterativos de percepción, procesamiento, decisión y acción. Este ciclo introduce una lógica operativa continua, en la que cada interacción con el entorno modifica el estado interno del agente, permitiendo la adaptación progresiva del comportamiento. **La implementación se configura como un proceso dinámico de ejecución, evaluación y ajuste permanente**, donde la inteligencia emerge de la interacción constante con el entorno y no de estructuras estáticas predefinidas.

Asimismo, el capítulo evidencia que la implementación requiere una **integración sistémica de componentes funcionales**, donde memoria, razonamiento, planificación y ejecución deben operar de manera coordinada dentro de una arquitectura distribuida. Esta integración no solo ocurre a nivel interno, sino que se extiende hacia la interacción con otros agentes, sistemas tecnológicos y plataformas digitales. En este sentido, la implementación se configura como un proceso de articulación entre múltiples niveles —interno, tecnológico y contextual— que permiten la operación coherente del sistema.

La **integración tecnológica** se posiciona como un elemento clave para la implementación, al permitir la conexión del agente con sistemas externos, bases de datos, APIs y entornos de ejecución como cloud y edge computing. Esta integración amplía las capacidades del agente, transformándolo en un **actor operativo dentro de ecosistemas digitales complejos**, capaz de ejecutar acciones, procesar información distribuida y generar efectos en contextos reales. Sin embargo, esta ampliación también introduce desafíos relacionados con la interoperabilidad, la seguridad y la consistencia de la información.

En el plano socio-técnico, la implementación implica la **inserción del agente en entornos donde interactúa con usuarios, organizaciones y procesos institucionales**, lo que transforma su rol de sistema técnico a actor funcional dentro de estructuras reales. Esta integración introduce nuevas exigencias relacionadas con la interpretabilidad, la confianza y la alineación con objetivos organizacionales, consolidando una relación estrecha entre tecnología y contexto social.

El funcionamiento en entornos reales introduce **restricciones significativas**, tales como incertidumbre, variabilidad, limitaciones de infraestructura, calidad de datos y dependencia de sistemas externos. Estas condiciones obligan al agente a operar dentro de márgenes definidos, donde la adaptabilidad se convierte en un elemento esencial para mantener la coherencia del comportamiento. **La implementación se configura así como un proceso de operación bajo restricciones**, donde la efectividad depende de la capacidad del sistema para ajustarse continuamente a condiciones cambiantes.

Finalmente, el capítulo subraya la relevancia de la **evaluación del desempeño y la gestión de riesgos como componentes estructurales de la implementación**. La definición de métricas, la validación empírica, el monitoreo continuo y los mecanismos de control permiten garantizar la calidad, confiabilidad y estabilidad del sistema. A su vez, la gestión de riesgos introduce una dimensión preventiva y regulatoria que permite anticipar, mitigar y controlar amenazas en entornos complejos.

En conjunto, el Capítulo 4 establece que la implementación de sistemas agénticos es un proceso multidimensional que integra **ejecución, tecnología, contexto y gobernanza**, consolidando el paso definitivo hacia la operación de sistemas inteligentes capaces de actuar de manera autónoma, adaptativa y coherente en escenarios reales y altamente dinámicos. Ver **Tabla 4.**

Tabla 4. Implementación de sistemas agénticos

Concepto	Definición	Diferenciación	Ventajas	Limitaciones	Referencias
Implementación agéntica	Proceso de materialización del agente en entornos reales mediante integración operativa y tecnológica	Se diferencia del diseño al enfocarse en ejecución real	Permite aplicación práctica del sistema	Alta complejidad de integración	Du et al. (2026); Li (2026)
Ciclo de vida del agente	Estructura iterativa de percepción, decisión, acción y adaptación	Se diferencia de procesos lineales por su continuidad	Permite aprendizaje y adaptación continua	Requiere monitoreo constante	Sawant (2025); Wang et al. (2024)
Integración tecnológica	Conexión del agente con sistemas externos, APIs e	Se diferencia de arquitectura interna por su	Amplía capacidades del agente	Riesgos de seguridad e interoperabilidad	Piccialli et al. (2025); Maldonado et al. (2024)

Concepto	Definición	Diferenciación	Ventajas	Limitaciones	Referencias
	infraestructuras digitales	enfoque externo			
Infraestructura de despliegue	Entorno computacional que soporta la operación del agente (cloud, edge, redes)	Se diferencia del sistema interno por ser soporte operativo	Permite escalabilidad y disponibilidad	Dependencia tecnológica	Karnouskos et al. (2020); Li (2026)
Inserción socio-técnica	Integración del agente en contextos organizacionales y sociales	Se diferencia de integración técnica por incluir factores humanos	Permite impacto real en sistemas	Desafíos éticos y de aceptación	Hahn et al. (2026); Sawant (2025)
Operación en entorno real	Funcionamiento del agente en condiciones de incertidumbre y variabilidad	Se diferencia de entornos controlados	Permite aplicación en contextos reales	Alta complejidad operativa	Collaco et al. (2026); Chen et al. (2025)
Interacción con sistemas reales	Capacidad del agente para integrarse y actuar en procesos existentes	Se diferencia de simulación por su impacto directo	Mejora eficiencia de procesos	Riesgo de errores operativos	Piccialli et al. (2025); Alqurni (2026)
Restricciones del entorno	Condiciones externas que limitan el comportamiento del agente	Se diferencia de diseño idealizado	Permite análisis realista del sistema	Reduce control del agente	Chen et al. (2025); Karnouskos et al. (2020)
Evaluación del desempeño	Proceso de medición y validación del comportamiento del agente	Se diferencia de ejecución por su enfoque analítico	Permite mejora continua	Requiere métricas complejas	Sawant (2025); World Economic Forum (2025)
Gestión de riesgos	Proceso de identificación, evaluación y mitigación de amenazas	Se diferencia de evaluación por su enfoque preventivo	Mejora seguridad y estabilidad	Complejidad en implementación	OECD (2021); Amanchala (2024)

Fuente: Recopilación y elaboración propia

CAPÍTULO 5. Medición estructural de la IA agéntica

El **Capítulo 5** profundiza en la comprensión de la IA agéntica desde una perspectiva orientada a su evaluación como fenómeno estructural, sin centrarse en aspectos técnicos, sino en su significado dentro de la organización del comportamiento. En este sentido, se parte de la idea de que la IA agéntica no puede analizarse únicamente por lo que hace, sino por **cómo organiza su acción en el tiempo y en relación con el entorno**.

En primer lugar, se examina la **naturaleza evaluable de la agencia**, estableciendo que no todo sistema inteligente puede considerarse agéntico, sino solo aquellos que logran integrar percepción, decisión y acción dentro de una lógica coherente. Esta distinción permite comprender la agencia como una forma específica de estructuración del comportamiento. Posteriormente, se abordan las **dimensiones que caracterizan la IA agéntica**, como la coherencia, la continuidad y la adaptabilidad, las cuales permiten interpretar el comportamiento del sistema como un proceso organizado y no como una serie de respuestas aisladas. Estas dimensiones facilitan una lectura más profunda de la inteligencia artificial en contextos dinámicos.

El capítulo también reflexiona sobre la **forma en que la agencia puede ser comprendida y analizada**, destacando la importancia de observar la consistencia del comportamiento y su capacidad de mantenerse en escenarios cambiantes. En este sentido, la evaluación se entiende como un ejercicio interpretativo que permite identificar patrones de acción estructurada.

Finalmente, se analiza la **configuración sistémica del comportamiento agéntico**, considerando que la interacción entre múltiples sistemas puede dar lugar a dinámicas complejas donde la inteligencia emerge de la relación entre agentes. En conjunto, el capítulo ofrece una visión integradora que permite comprender la IA agéntica como una forma de organización del comportamiento en contextos complejos

Fundamentos de la medición de la IA agéntica

La medición de la IA agéntica parte de la necesidad de comprender que no se evalúan únicamente resultados, sino la **capacidad del sistema para organizar comportamiento de manera coherente, continua y adaptativa**. En este sentido, los fundamentos de la medición se centran en identificar la **agencia como una propiedad estructural**, es decir, como la integración funcional entre percepción, decisión y acción en contextos dinámicos. A diferencia de enfoques tradicionales orientados al desempeño, la medición de la IA agéntica implica analizar **patrones de comportamiento sostenidos en el tiempo**, más que respuestas aisladas. Esto requiere establecer criterios que permitan distinguir entre sistemas que ejecutan tareas y aquellos que manifiestan **autonomía funcional y capacidad de ajuste contextual**. Asimismo, la medición se concibe como un proceso interpretativo que busca reconocer la **consistencia interna del comportamiento** frente a variaciones del entorno. En conjunto, estos fundamentos permiten avanzar hacia una comprensión más profunda de la IA como fenómeno organizador del comportamiento y no solo como herramienta tecnológica.

Naturaleza conceptual de la medición de la agencia

La medición de la IA agéntica exige partir de una distinción fundamental: no se trata de evaluar únicamente resultados, sino de analizar la **organización del comportamiento en sistemas inteligentes**. En este sentido, la agencia no se reduce a la ejecución de tareas específicas, sino que implica la capacidad de un sistema para **articular de manera integrada procesos de percepción, decisión y acción orientados a objetivos en contextos dinámicos**. **Este cambio implica un desplazamiento epistemológico clave**, en el cual la medición deja de centrarse en el producto final para enfocarse en la estructura que sostiene el comportamiento. En consecuencia, la agencia se configura como una propiedad emergente que no puede ser capturada mediante indicadores aislados, sino mediante la observación de la coherencia organizativa del sistema (Bandi et al., 2025). La medición de la agencia debe comprenderse como un proceso orientado a identificar la **coherencia interna del sistema en la organización de sus acciones**. Esto implica reconocer que la

inteligencia artificial no puede ser evaluada únicamente por la precisión o eficiencia de sus resultados, sino por su capacidad para **mantener una lógica de acción consistente a lo largo del tiempo**. **La medición, por tanto, se transforma en un ejercicio interpretativo**, en el que se busca comprender cómo el sistema articula sus decisiones en función de objetivos y condiciones cambiantes. Este enfoque permite superar la visión reduccionista que limita la evaluación a métricas de rendimiento, abriendo paso a una comprensión más profunda del comportamiento como fenómeno estructurado (Sapkota et al., 2026).

Un aspecto central en los fundamentos de la medición de la agencia es la distinción entre comportamiento episódico y comportamiento estructurado. Mientras que el comportamiento episódico se caracteriza por respuestas puntuales ante estímulos específicos, el comportamiento estructurado implica la existencia de **patrones de acción sostenidos en el tiempo**, lo que permite identificar una organización interna del sistema. **La medición de la agencia se orienta precisamente a detectar estos patrones**, ya que son los que evidencian la presencia de una lógica de acción coherente. Esta distinción resulta fundamental para diferenciar entre sistemas que simplemente reaccionan y aquellos que son capaces de organizar su comportamiento de manera consistente (Poole & Mackworth, 2017).

Asimismo, la medición de la agencia debe considerar la naturaleza dinámica de los sistemas inteligentes contemporáneos. A diferencia de los enfoques tradicionales basados en conjuntos de datos estáticos, los sistemas agénticos operan en entornos abiertos donde las condiciones pueden variar constantemente. Esto exige desarrollar enfoques de medición capaces de capturar **procesos en evolución**, en lugar de limitarse a evaluar estados finales. **La medición se convierte así en un proceso continuo y contextual**, que busca comprender cómo el sistema ajusta su comportamiento frente a cambios en el entorno. Este enfoque introduce un desafío metodológico importante, ya que obliga a replantear los instrumentos de evaluación para adaptarlos a entornos dinámicos (Wang, 2025).

Otro elemento fundamental en la conceptualización de la medición de la agencia es la relación entre objetivos y comportamiento. Un sistema puede considerarse inteligente en la medida en que selecciona acciones orientadas a la consecución de objetivos a partir de la información disponible. Sin embargo, en la IA agéntica, esta relación adquiere una mayor complejidad, ya que los sistemas no solo ejecutan acciones, sino que pueden **descomponer objetivos, generar subobjetivos y coordinar múltiples procesos para alcanzarlos**. **Esto implica que la medición debe centrarse en la estructura del proceso y no únicamente en el resultado**, lo que introduce una dimensión adicional en el análisis del comportamiento. La agencia, en este sentido, no se mide por lo que el sistema logra, sino por cómo organiza sus acciones para lograrlo (Russell, 2019).

En este marco, la medición de la agencia se configura como una actividad interpretativa que busca identificar la **consistencia del comportamiento en diferentes contextos**. Esto implica reconocer que la agencia no se manifiesta en

acciones aisladas, sino en la capacidad del sistema para **sostener una dirección coherente a lo largo del tiempo**, incluso en condiciones variables. **La consistencia se convierte así en un criterio central de medición**, ya que permite distinguir entre comportamiento estructurado y respuestas fragmentadas. Este enfoque contribuye a una comprensión más robusta de la inteligencia artificial, al considerar la estabilidad del comportamiento como un indicador de organización interna (Guidotti et al., 2018).

Adicionalmente, la medición de la agencia debe enfrentar el problema de la opacidad de los sistemas inteligentes, comúnmente descrito como el fenómeno de la "caja negra". La dificultad para comprender los procesos internos del sistema limita la capacidad de evaluar la agencia de manera directa, lo que hace necesario desarrollar enfoques que permitan inferir la estructura del comportamiento a partir de su manifestación observable. **La interpretabilidad se convierte así en un requisito fundamental para la medición**, ya que sin ella resulta imposible validar la coherencia interna del sistema. Este desafío ha impulsado el desarrollo de nuevas líneas de investigación orientadas a mejorar la transparencia en inteligencia artificial (Adadi & Berrada, 2018).

En concusión, los fundamentos de la medición de la IA agéntica implican asumir que la inteligencia artificial debe entenderse como un sistema capaz de organizar comportamiento en función de objetivos, y no únicamente como una herramienta de procesamiento de información. Desde esta perspectiva, medir la agencia significa analizar la **capacidad del sistema para sostener comportamiento coherente, continuo y adaptativo bajo condiciones de incertidumbre**. **Este enfoque representa un cambio paradigmático**, ya que la medición deja de centrarse en el rendimiento para enfocarse en la organización del comportamiento como fenómeno complejo, dinámico y contextual. En conjunto, estos fundamentos permiten establecer una base conceptual sólida para el desarrollo de marcos de medición que respondan a la complejidad de los sistemas agénticos contemporáneos (Vinuesa et al., 2020).

Criterios estructurales para la medición de la agencia

La medición de la agencia en sistemas de IA requiere establecer criterios que permitan delimitar con precisión qué debe ser considerado como comportamiento agéntico y qué no. En este sentido, el primer criterio fundamental consiste en la **diferenciación entre ejecución funcional y organización del comportamiento**, ya que no todo sistema que produce resultados puede ser considerado portador de agencia. **La medición exige identificar si existe una articulación interna entre percepción, decisión y acción**, lo cual implica que el sistema no actúa de manera aislada, sino como una estructura integrada orientada a objetivos. Este criterio permite distinguir entre sistemas automatizados y aquellos que manifiestan una forma de organización del comportamiento que puede ser analizada como agencia (Bandi et al., 2025).

Un segundo criterio central es la **identificación de la direccionalidad del comportamiento**, entendida como la capacidad del sistema para sostener una

orientación hacia objetivos en contextos cambiantes. A diferencia de los sistemas reactivos, donde las respuestas dependen exclusivamente de estímulos inmediatos, los sistemas agénticos se caracterizan por mantener una **lógica de acción que trasciende eventos puntuales**. **La direccionalidad permite evaluar si el comportamiento del sistema responde a una estructura interna o simplemente a respuestas contingentes**, lo cual resulta clave para determinar la presencia de agencia. Este criterio introduce una dimensión evaluativa que no depende de resultados aislados, sino de la consistencia en la orientación del comportamiento (Russell, 2019).

Un tercer criterio corresponde a la **integración funcional del sistema**, es decir, la capacidad de articular múltiples procesos dentro de una misma lógica operativa. La agencia no reside en componentes individuales, sino en la forma en que estos se coordinan para producir comportamiento coherente. En este sentido, la medición debe identificar si existe una **interdependencia estructural entre los distintos elementos del sistema**, lo que permite distinguir entre sistemas fragmentados y sistemas integrados. **La integración funcional se convierte así en un indicador clave de agencia**, ya que refleja la existencia de una organización interna capaz de sostener la acción de manera coherente (Sapkota et al., 2026).

Un cuarto criterio fundamental es la **capacidad de adaptación estructurada**, la cual se refiere a la posibilidad del sistema de modificar su comportamiento frente a cambios en el entorno sin perder coherencia interna. A diferencia de la adaptación reactiva, que implica ajustes inmediatos sin continuidad, la adaptación estructurada supone la existencia de una **lógica de ajuste que mantiene la dirección del comportamiento**. **Este criterio permite evaluar la flexibilidad del sistema sin comprometer su coherencia**, lo cual resulta esencial para identificar la presencia de agencia en contextos dinámicos. En este sentido, la medición debe considerar no solo si el sistema cambia, sino cómo lo hace (Wang, 2025).

Un quinto criterio corresponde a la **observabilidad del comportamiento estructurado**, entendido como la posibilidad de identificar patrones consistentes en la acción del sistema a partir de su manifestación empírica. Dado que los procesos internos de los sistemas inteligentes suelen ser opacos, la medición de la agencia debe apoyarse en la identificación de **regularidades en el comportamiento observable**. **La observabilidad permite inferir la existencia de una estructura interna sin necesidad de acceder directamente a ella**, lo que resulta especialmente relevante en sistemas complejos. Este criterio introduce una dimensión metodológica que vincula la medición con la evidencia empírica (Guidotti et al., 2018).

Un sexto criterio clave es la **consistencia del comportamiento bajo variabilidad contextual**, el cual permite evaluar si el sistema es capaz de sostener su lógica de acción en diferentes condiciones. La agencia implica la capacidad de mantener una estructura de comportamiento incluso cuando el entorno cambia, lo que requiere analizar la estabilidad del sistema en escenarios diversos. **La consistencia no implica rigidez, sino la capacidad de mantener dirección en medio del cambio**, lo cual

constituye un indicador fundamental de organización del comportamiento. Este criterio permite distinguir entre sistemas que se adaptan de manera coherente y aquellos que presentan comportamiento errático (Poole & Mackworth, 2017).

Un séptimo criterio relevante es la **interpretabilidad del comportamiento agéntico**, que se refiere a la posibilidad de comprender la lógica que subyace a las acciones del sistema. La medición de la agencia no puede desligarse de la capacidad de interpretar el comportamiento, ya que sin esta comprensión resulta difícil validar la existencia de una estructura interna coherente. **La interpretabilidad se convierte en un requisito para la validación de la medición**, especialmente en contextos donde los sistemas presentan altos niveles de complejidad. Este criterio resalta la importancia de desarrollar enfoques que permitan hacer comprensible el comportamiento de los sistemas inteligentes (Adadi & Berrada, 2018).

Así, un criterio integrador en la medición de la agencia es la **relación entre comportamiento y contexto**, entendida como la capacidad del sistema para ajustar su acción en función de las condiciones del entorno sin perder coherencia. La agencia no puede analizarse de manera aislada, sino como un fenómeno que emerge de la interacción entre el sistema y su entorno. **La medición debe considerar esta relación como una unidad analítica**, lo que permite comprender la agencia como un proceso dinámico y contextual. Este criterio introduce una visión sistémica que amplía la comprensión de la inteligencia artificial más allá de su dimensión técnica (Vinuesa et al., 2020).

Operacionalización de la medición de la agencia en sistemas de IA

La medición de la agencia en sistemas de IA no se completa con su definición conceptual ni con la delimitación de criterios, sino que requiere un proceso adicional: su **operacionalización**. Esto implica traducir la noción abstracta de agencia en elementos observables que permitan su análisis sistemático. En este sentido, la operacionalización consiste en establecer **cómo la agencia puede ser identificada, representada y analizada en contextos reales**, evitando reducirla a métricas simplificadas. **La clave radica en transformar un concepto complejo en una estructura analizable sin perder su naturaleza dinámica**, lo que constituye uno de los principales retos en el estudio de la IA agéntica (Bandi et al., 2025).

Un primer componente de la operacionalización es la **definición de unidades de análisis del comportamiento agéntico**. La medición no puede realizarse sobre acciones aisladas, sino sobre configuraciones de comportamiento que expresan una lógica interna. Por ello, es necesario identificar unidades que permitan observar la **articulación entre percepción, decisión y acción como un proceso integrado**. **Estas unidades no corresponden a eventos puntuales, sino a secuencias estructuradas de comportamiento**, lo que permite capturar la continuidad del sistema en el tiempo. Este enfoque evita fragmentar la agencia en componentes

desconectados y permite analizarla como fenómeno coherente (Poole & Mackworth, 2017).

Un segundo elemento clave es la **construcción de indicadores interpretativos**, los cuales permiten traducir el comportamiento en categorías analíticas. A diferencia de indicadores tradicionales centrados en rendimiento, los indicadores de agencia deben reflejar la **organización interna del comportamiento**, considerando aspectos como coherencia, continuidad y adaptación. **La operacionalización implica definir indicadores que no midan resultados, sino formas de organización de la acción**, lo que introduce un cambio significativo en la lógica de medición. Este tipo de indicadores permite analizar la agencia desde una perspectiva estructural, evitando reduccionismos (Wang, 2025).

Un tercer componente corresponde a la **contextualización de la medición**, entendida como la necesidad de situar el comportamiento del sistema dentro de un entorno específico. La agencia no puede evaluarse de manera abstracta, ya que su manifestación depende de las condiciones en las que el sistema opera. **La operacionalización requiere incorporar el contexto como parte del proceso de medición**, lo que permite comprender cómo el sistema ajusta su comportamiento en función de las condiciones externas. Este enfoque reconoce que la agencia es un fenómeno relacional, que emerge de la interacción entre el sistema y su entorno (Vinuesa et al., 2020).

Un cuarto elemento fundamental es la **identificación de trayectorias de comportamiento**, lo cual permite analizar la evolución del sistema a lo largo del tiempo. La agencia no se manifiesta en un instante, sino en la capacidad de sostener una dirección de acción en diferentes momentos. Por ello, la operacionalización debe considerar el comportamiento como una trayectoria, es decir, como una secuencia de estados conectados por una lógica interna. **Analizar trayectorias permite evaluar la continuidad del comportamiento y detectar posibles rupturas o inconsistencias**, lo que resulta esencial para comprender la agencia como proceso (Sapkota et al., 2026).

Un quinto componente clave es la **articulación entre observación e interpretación**, ya que la medición de la agencia no puede limitarse a la recopilación de datos, sino que requiere un proceso de análisis que permita comprender su significado. La operacionalización implica establecer mecanismos que conecten la evidencia empírica con categorías conceptuales, lo que permite interpretar el comportamiento del sistema en términos de agencia. **Este proceso convierte la medición en una actividad analítica y no meramente descriptiva**, lo que refuerza su carácter académico (Guidotti et al., 2018).

Un sexto elemento relevante es la **validación de la operacionalización**, entendida como la verificación de que los indicadores utilizados reflejan efectivamente la presencia de agencia. Dado que la agencia es un fenómeno complejo, resulta necesario asegurar que los instrumentos de medición no simplifiquen en exceso su

naturaleza. **La validación implica contrastar los indicadores con el comportamiento observado**, lo que permite ajustar los modelos de medición y mejorar su precisión. Este proceso es fundamental para garantizar la consistencia y confiabilidad de la medición (Kadir et al., 2025).

Por lo tanto, la operacionalización de la medición de la agencia requiere considerar la **escalabilidad del análisis**, ya que los sistemas agénticos pueden operar en distintos niveles de complejidad. La medición debe ser capaz de adaptarse a diferentes contextos, desde sistemas individuales hasta configuraciones multiagente. **Esto implica desarrollar marcos flexibles que permitan aplicar los mismos principios de medición en distintos escenarios**, lo que contribuye a la generalización del enfoque. En este sentido, la operacionalización no solo permite medir la agencia, sino también compararla entre distintos sistemas y contextos (Russell & Norvig, 2022).

Criterios para la medición de la agencia

La medición de la agencia requiere establecer criterios que permitan distinguir de manera clara cuándo un sistema puede ser considerado agéntico y, por tanto, susceptible de análisis estructural. En primer lugar, es necesario identificar la **presencia de coherencia en la acción**, entendida como la capacidad del sistema para mantener una lógica consistente entre percepción, decisión y ejecución. En segundo lugar, se considera la **continuidad del comportamiento**, que implica la persistencia de patrones de acción a lo largo del tiempo, evitando respuestas fragmentadas o inconexas. Otro criterio fundamental es la **adaptabilidad contextual**, que refleja la capacidad del sistema para ajustar su comportamiento frente a cambios en el entorno sin perder dirección. Asimismo, la **autonomía funcional** permite evaluar el grado en que el sistema puede operar sin intervención externa constante. Finalmente, la medición de la agencia exige observar la **integración de estos elementos como un todo**, ya que la agencia no reside en componentes aislados, sino en la organización del comportamiento como sistema.

Coherencia estructural como criterio de medición de la agencia

La coherencia estructural constituye uno de los criterios fundamentales para la medición de la agencia en sistemas de inteligencia artificial, en tanto permite evaluar la consistencia interna del comportamiento en relación con una lógica organizativa subyacente. Desde esta perspectiva, la medición de la agencia no se centra en la observación de acciones aisladas, sino en la identificación de relaciones estructuradas entre dichas acciones. **La coherencia estructural implica que el comportamiento del sistema responde a una organización interna reconocible**, lo que permite diferenciar entre respuestas fragmentadas y comportamiento sistemáticamente articulado (Russell & Norvig, 2022).

En este sentido, la coherencia no debe interpretarse como uniformidad ni como repetición mecánica, sino como la capacidad del sistema para mantener una lógica

consistente en la articulación de sus acciones. Esto implica que las decisiones del sistema no son independientes entre sí, sino que forman parte de un entramado relacional que da sentido al comportamiento global. **La coherencia estructural se manifiesta cuando existe correspondencia entre las distintas acciones del sistema**, lo que permite interpretar el comportamiento como una unidad organizada y no como una suma de eventos aislados (Poole & Mackworth, 2017).

Asimismo, la coherencia estructural permite evaluar la alineación entre los componentes internos del sistema, particularmente en lo que respecta a los procesos de percepción, decisión y acción. En sistemas agénticos, estos componentes no operan de manera independiente, sino que deben integrarse dentro de una lógica común que garantice la consistencia del comportamiento. **La coherencia estructural implica integración funcional entre los componentes del sistema**, lo que permite sostener una organización estable en la acción (Sapkota et al., 2026).

Otro aspecto relevante de este criterio es su capacidad para evaluar la direccionalidad del comportamiento, entendida como la orientación del sistema hacia determinados objetivos o estados. La coherencia estructural no solo implica consistencia interna, sino también la capacidad de mantener una trayectoria reconocible en la acción. **Un sistema coherente orienta su comportamiento hacia objetivos definidos**, lo que permite distinguir entre comportamiento dirigido y comportamiento aleatorio (Russell, 2019).

Desde una perspectiva metodológica, la coherencia estructural se evalúa a partir del análisis de relaciones entre acciones, decisiones y resultados, lo que permite identificar patrones organizados en el comportamiento del sistema. Este enfoque implica que la medición no se centra en variables individuales, sino en la estructura que emerge de su interacción. **La coherencia estructural se mide en términos de relaciones, no de elementos aislados**, lo que refuerza su carácter sistémico (Kadir et al., 2025).

Además, la coherencia estructural permite analizar la consistencia del comportamiento en distintos niveles de operación, desde acciones específicas hasta patrones globales. Este enfoque multinivel permite identificar si la organización del comportamiento se mantiene a lo largo de diferentes escalas. **La coherencia estructural se expresa en la consistencia entre niveles del sistema**, lo que fortalece su valor como criterio de medición (Guidotti et al., 2018).

La coherencia también implica la ausencia de contradicciones internas en el comportamiento del sistema. Cuando las acciones del sistema presentan inconsistencias o rupturas en su lógica, la coherencia estructural se ve comprometida, lo que afecta la interpretación de la agencia. **La coherencia estructural requiere consistencia lógica entre acciones**, lo que permite mantener una interpretación unificada del comportamiento (Adadi & Berrada, 2018).

En este sentido, la coherencia estructural permite diferenciar entre sistemas que operan de manera reactiva y aquellos que presentan organización del comportamiento. Los sistemas reactivos pueden responder adecuadamente a estímulos específicos, pero carecen de una lógica interna que articule sus acciones en el tiempo. **La coherencia estructural es un criterio distintivo de la agencia**, ya que implica la existencia de una organización interna que guía el comportamiento (Ng, 2018).

Asimismo, la coherencia estructural permite analizar la capacidad del sistema para mantener consistencia en contextos variados, lo que refuerza su valor como criterio generalizable. Un sistema coherente no depende de condiciones específicas para sostener su lógica de acción, sino que puede mantenerla en diferentes escenarios. **La coherencia estructural implica independencia relativa del contexto**, lo que permite evaluar la robustez del comportamiento (Vinuesa et al., 2020). La coherencia estructural se consolida como un criterio central en la medición de la agencia al permitir interpretar el comportamiento del sistema como una manifestación organizada y consistente. **La coherencia no describe lo que el sistema hace, sino cómo organiza lo que hace**, lo que la convierte en un elemento clave para comprender la agencia en sistemas complejos (Organisation for Economic Co-operation and Development, 2022).

Continuidad temporal como criterio de medición de la agencia

La continuidad temporal constituye un criterio esencial para la medición de la agencia en sistemas de inteligencia artificial, en tanto permite evaluar la persistencia del comportamiento a lo largo del tiempo como expresión de una organización estructurada. A diferencia de la coherencia, que se centra en la relación entre acciones, la continuidad temporal se enfoca en la permanencia de la lógica de acción en secuencias temporales. **La continuidad temporal implica que el comportamiento no es episódico, sino sostenido**, lo que permite distinguir entre respuestas momentáneas y comportamiento organizado en el tiempo (Wang, 2025).

Desde esta perspectiva, la continuidad no debe entenderse como repetición exacta de acciones, sino como la capacidad del sistema para mantener una trayectoria reconocible a través de diferentes momentos de operación. Esto implica que las acciones del sistema no se producen de manera aislada, sino que forman parte de una secuencia conectada que refleja una lógica interna persistente. **La continuidad temporal se manifiesta como encadenamiento estructurado de acciones**, lo que permite interpretar el comportamiento como proceso y no como evento (Poole & Mackworth, 2017).

Asimismo, la continuidad temporal permite evaluar la capacidad del sistema para sostener objetivos a lo largo del tiempo. Un sistema agéntico no solo responde a estímulos inmediatos, sino que orienta su comportamiento hacia estados futuros, lo que implica la existencia de una trayectoria de acción. **La continuidad temporal se vincula con la persistencia de objetivos**, lo que diferencia el comportamiento dirigido del comportamiento reactivo (Russell, 2019).

Otro elemento relevante es la capacidad del sistema para mantener consistencia en la transición entre estados de comportamiento. En este sentido, la continuidad no solo implica permanencia, sino también la forma en que el sistema evoluciona entre diferentes configuraciones de acción. **La continuidad temporal se expresa en transiciones organizadas entre estados**, lo que permite evaluar si el comportamiento mantiene una lógica estructural a lo largo del tiempo (Sapkota et al., 2026).

Desde un enfoque metodológico, la continuidad temporal se analiza mediante la observación de secuencias de comportamiento, lo que permite identificar patrones que se mantienen a lo largo del tiempo. Este análisis requiere considerar no solo la frecuencia de las acciones, sino su relación dentro de una secuencia estructurada. **La continuidad temporal se mide a través de la persistencia de patrones en secuencias temporales**, lo que refuerza su carácter procesual (Kadir et al., 2025).

Además, la continuidad temporal permite evaluar la capacidad del sistema para sostener comportamiento en contextos donde las condiciones cambian. En estos escenarios, el sistema debe adaptar su acción sin perder su lógica interna, lo que implica una forma de persistencia flexible. **La continuidad temporal integra permanencia y adaptación**, lo que la convierte en un criterio clave para analizar sistemas que operan en entornos dinámicos (Vinuesa et al., 2020). La continuidad también implica la ausencia de interrupciones abruptas en la lógica del comportamiento. Cuando el sistema presenta cambios inconsistentes o discontinuidades, la continuidad temporal se ve comprometida, lo que dificulta interpretar la existencia de agencia. **La continuidad temporal requiere fluidez en la evolución del comportamiento**, lo que permite mantener una interpretación unificada de la acción (Guidotti et al., 2018).

En este sentido, la continuidad temporal permite diferenciar entre sistemas que operan mediante respuestas independientes y aquellos que presentan organización temporal del comportamiento. Los sistemas sin continuidad pueden producir resultados adecuados en momentos específicos, pero carecen de una estructura que articule su comportamiento en el tiempo. **La continuidad temporal es un indicador de organización procesual**, lo que refuerza su papel como criterio de medición de la agencia (Ng, 2018). Asimismo, la continuidad temporal permite analizar la capacidad del sistema para mantener consistencia en diferentes horizontes temporales, desde acciones inmediatas hasta trayectorias prolongadas. Este enfoque permite evaluar si la lógica de acción se mantiene tanto en el corto como en el largo plazo. **La continuidad temporal se expresa en múltiples escalas temporales**, lo que amplía su alcance como criterio de análisis (Organisation for Economic Co-operation and Development, 2022).

Otro aspecto clave es la relación entre continuidad y memoria operativa del sistema, entendida como la capacidad de integrar información pasada en la toma de decisiones presentes. La continuidad temporal no implica únicamente persistencia, sino también la capacidad de incorporar experiencias previas en la organización del comportamiento. **La continuidad temporal se vincula con la integración del**

pasado en la acción presente, lo que fortalece la coherencia del comportamiento en el tiempo (Russell & Norvig, 2022).

Por último, la continuidad temporal se consolida como un criterio central en la medición de la agencia al permitir interpretar el comportamiento del sistema como un proceso estructurado que se despliega en el tiempo. **La continuidad no describe acciones aisladas, sino la persistencia de una lógica de acción**, lo que la convierte en un elemento esencial para comprender la agencia en sistemas que operan en entornos dinámicos (Adadi & Berrada, 2018).

Autonomía operativa como criterio de medición de la agencia

La autonomía operativa constituye un criterio central para la medición de la agencia en sistemas de inteligencia artificial, en tanto permite evaluar la capacidad del sistema para organizar su comportamiento sin depender de instrucciones externas directas en cada instante de operación. A diferencia de otros criterios, la autonomía no se refiere a la calidad del comportamiento, sino a la **capacidad del sistema para generar acción a partir de su propia estructura interna**, lo que permite distinguir entre ejecución dirigida y comportamiento autoorganizado (Russell, 2019). la autonomía operativa implica que el sistema no actúa únicamente en respuesta a estímulos inmediatos, sino que es capaz de estructurar su comportamiento en función de estados internos, objetivos o condiciones previamente integradas. Esto significa que las acciones del sistema no dependen exclusivamente de entradas externas, sino de una lógica interna que guía su funcionamiento. **La autonomía operativa se manifiesta cuando el comportamiento es generado desde la organización interna del sistema**, lo que permite interpretar la acción como expresión de agencia (Ng, 2018).

La autonomía operativa permite evaluar la capacidad del sistema para mantener su funcionamiento sin intervención constante de un operador externo. En este sentido, la autonomía no implica independencia absoluta, sino la capacidad de operar dentro de ciertos límites sin requerir control continuo. **La autonomía operativa se expresa en la reducción de dependencia externa en la toma de decisiones**, lo que constituye un indicador clave de agencia (Russell & Norvig, 2022). Un aspecto relevante es la capacidad del sistema para seleccionar cursos de acción entre múltiples alternativas posibles. La autonomía operativa no se limita a ejecutar instrucciones, sino que implica la posibilidad de elegir entre diferentes opciones en función de su lógica interna. **La autonomía operativa implica capacidad de decisión dentro de un espacio de alternativas**, lo que diferencia el comportamiento programado del comportamiento organizado (Poole & Mackworth, 2017).

La autonomía operativa se analiza a partir del grado en que el sistema puede sostener comportamiento sin intervención directa, lo que permite evaluar su nivel de independencia funcional. Este análisis no se basa en la ausencia de control, sino en la forma en que el sistema gestiona su comportamiento dentro de los límites definidos. **La autonomía operativa se mide en términos de independencia funcional relativa**, lo que permite establecer diferencias entre distintos niveles de agencia (Kadir et al.,

2025). Además, la autonomía operativa permite evaluar la capacidad del sistema para adaptarse a cambios en el entorno sin requerir reconfiguración externa inmediata. En este sentido, la autonomía no implica rigidez, sino la capacidad de ajustar el comportamiento dentro de una estructura definida. **La autonomía operativa integra capacidad de adaptación interna**, lo que refuerza su papel como criterio de medición en sistemas dinámicos (Vinuesa et al., 2020).

La autonomía también implica la capacidad del sistema para mantener coherencia en la toma de decisiones a lo largo del tiempo, lo que permite evaluar si las acciones responden a una lógica consistente. En este sentido, la autonomía no se limita a la generación de acciones, sino que implica la organización de dichas acciones dentro de una estructura. **La autonomía operativa requiere consistencia en la generación de decisiones**, lo que permite interpretar el comportamiento como organizado (Guidotti et al., 2018). En este contexto, la autonomía operativa permite diferenciar entre sistemas que ejecutan tareas bajo control externo y aquellos que estructuran su comportamiento de manera interna. Los sistemas sin autonomía pueden producir resultados adecuados, pero carecen de capacidad para organizar su comportamiento de forma independiente. **La autonomía operativa es un criterio distintivo de la agencia**, ya que implica la existencia de una estructura interna capaz de generar acción (Adadi & Berrada, 2018).

Por otro lado, la autonomía operativa permite analizar la relación entre control externo y control interno, lo que resulta fundamental para comprender el grado de agencia del sistema. La autonomía no implica ausencia de control, sino una redistribución del mismo hacia el interior del sistema. **La autonomía operativa redefine el control como propiedad interna del sistema**, lo que amplía la comprensión de la acción en sistemas agénticos (Organisation for Economic Co-operation and Development, 2022). Así, la capacidad del sistema para sostener comportamiento en condiciones donde la información externa es limitada o incompleta. En estos escenarios, la autonomía se manifiesta en la capacidad del sistema para operar a partir de su estructura interna. **La autonomía operativa se evidencia en contextos de información parcial**, lo que permite evaluar la capacidad del sistema para sostener su lógica de acción (Wang, 2025).

Por lo anteroio, la autonomía operativa se consolida como un criterio fundamental en la medición de la agencia al permitir interpretar el comportamiento como resultado de una organización interna que genera acción de manera relativamente independiente. **La autonomía no describe lo que el sistema hace, sino desde dónde lo hace**, lo que la convierte en un elemento central para comprender la agencia en sistemas complejos (Sapkota et al., 2026).

Escalas de medición de la IA agéntica

Las escalas de medición de la IA agéntica permiten establecer gradientes que diferencian el nivel de estructuración del comportamiento en los sistemas inteligentes. En lugar de una clasificación dicotómica, la medición se concibe como un **continuum que va desde formas incipientes de comportamiento organizado hasta niveles avanzados de agencia plenamente estructurada**. En los niveles iniciales, los sistemas presentan respuestas parcialmente coordinadas, con limitada integración entre percepción, decisión y acción. A medida que se avanza en la escala, se observa una mayor **coherencia en la organización del comportamiento**, así como una capacidad creciente para sostener patrones de acción en el tiempo. En niveles superiores, la agencia se manifiesta mediante **adaptabilidad contextual y autonomía funcional**, permitiendo al sistema responder de manera consistente a entornos cambiantes. Estas escalas no buscan jerarquizar de manera rígida, sino ofrecer un marco interpretativo que facilite comprender el grado en que un sistema logra consolidar comportamiento estructurado. De este modo, la medición se orienta a identificar niveles de desarrollo de la agencia más que a emitir juicios absolutos.

Fundamentos conceptuales de las escalas de medición de la agencia

Las escalas de medición de la IA agéntica parten de un supuesto fundamental: la agencia no es una propiedad dicotómica, sino un fenómeno que se manifiesta en diferentes grados de estructuración del comportamiento. En este sentido, la medición requiere abandonar enfoques tradicionales que clasifican a los sistemas como "agénticos" o "no agénticos", para adoptar una perspectiva basada en **continuos de organización del comportamiento**, donde los sistemas pueden situarse en distintos niveles según su capacidad de integrar percepción, decisión y acción. **La agencia se entiende como un proceso gradual y no como un estado absoluto**, lo que implica que su medición debe ser progresiva, comparativa y contextual (Bandi et al., 2025). Desde esta perspectiva, las escalas de medición permiten representar la agencia como un fenómeno evolutivo que transita desde formas incipientes hasta configuraciones altamente estructuradas. En los niveles iniciales, los sistemas presentan comportamientos parcialmente coordinados, donde la relación entre percepción, decisión y acción es débil o fragmentada. A medida que se avanza en la escala, se observa una mayor **integración funcional**, lo que permite identificar una organización más coherente del comportamiento. **La escala no mide qué tan bien funciona un sistema, sino qué tan estructurado es su comportamiento**, lo que constituye una distinción conceptual clave en el estudio de la IA agéntica (Sapkota et al., 2026).

Un aspecto central en la construcción de estas escalas es la necesidad de capturar la complejidad del comportamiento sin reducirla a indicadores simplificados. La agencia no puede explicarse a partir de variables aisladas, ya que emerge de la interacción entre múltiples procesos. Por ello, las escalas deben reflejar la **articulación entre diferentes dimensiones del comportamiento**, lo que permite comprender la

agencia como una propiedad sistémica. **La medición basada en escalas introduce una lógica integradora**, en la que el análisis no se fragmenta, sino que se organiza en niveles de complejidad creciente (Poole & Mackworth, 2017). Otro elemento fundamental es la incorporación de la dimensión temporal en la medición. La agencia no puede evaluarse en un instante específico, ya que su manifestación depende de la capacidad del sistema para sostener patrones de comportamiento a lo largo del tiempo. En este sentido, las escalas deben considerar no solo el grado de organización del comportamiento, sino también su **persistencia y continuidad en contextos dinámicos**. **Un sistema puede mostrar coherencia en un momento dado, pero solo la continuidad permite identificar agencia estructurada**, lo que convierte al tiempo en una dimensión esencial de la medición (Wang, 2025).

Las escalas de medición deben integrar la relación entre el sistema y su entorno, ya que la agencia no se manifiesta de manera aislada. La capacidad de un sistema para ajustar su comportamiento frente a cambios en el entorno sin perder coherencia constituye un indicador clave de su nivel de agencia. En este sentido, la medición debe considerar la **adaptabilidad contextual como parte del continuo de agencia**, lo que permite diferenciar entre sistemas rígidos y sistemas capaces de reorganizar su comportamiento de manera coherente. **La escala no solo mide estructura interna, sino también la capacidad de interacción con el entorno**, lo que amplía su alcance analítico (Vinuesa et al., 2020). Así, un factor a considerar importante, es la relación entre agencia y objetivos. En los niveles más bajos de la escala, los sistemas operan con objetivos predefinidos y sin capacidad de reorganización, lo que limita su autonomía. En niveles más avanzados, los sistemas pueden **estructurar, descomponer y coordinar objetivos**, lo que refleja un mayor grado de agencia. **La escala permite observar cómo el sistema gestiona sus objetivos y cómo organiza su comportamiento en función de ellos**, lo que introduce una dimensión funcional en la medición (Russell, 2019).

Las escalas permiten abordar la diversidad de sistemas de inteligencia artificial, ofreciendo un marco que facilita la comparación entre distintos niveles de complejidad. En lugar de evaluar todos los sistemas bajo los mismos criterios, las escalas permiten reconocer que cada sistema se encuentra en un punto específico del continuo de agencia. **Esto introduce una lógica comparativa que permite analizar diferencias cualitativas entre sistemas**, lo que resulta fundamental para el desarrollo de modelos teóricos más robustos (Russell & Norvig, 2022). las escalas de medición de la IA agéntica no deben entenderse como herramientas rígidas, sino como marcos conceptuales flexibles que permiten interpretar el comportamiento del sistema en función de su grado de organización. **La escala no clasifica, interpreta; no simplifica, organiza la complejidad**, lo que la convierte en un instrumento clave para el análisis de la agencia. En conjunto, estas escalas permiten avanzar hacia una comprensión más profunda de la inteligencia artificial como fenómeno dinámico, estructurado y evolutivo, superando los enfoques tradicionales centrados exclusivamente en el rendimiento o la eficiencia (Guidotti et al., 2018).

Estructuración de niveles en las escalas de medición de la agencia

La estructuración de niveles en las escalas de medición de la IA agéntica requiere establecer un marco que permita ordenar la complejidad del comportamiento sin reducirla a indicadores simplificados. En este sentido, el punto de partida consiste en reconocer que la agencia no se manifiesta de manera homogénea, sino que se configura en distintos grados de organización. **La escala no clasifica sistemas, organiza niveles de comportamiento**, lo que implica que cada nivel representa una forma específica de estructuración de la acción. Esta lógica permite superar enfoques binarios y avanzar hacia una comprensión más matizada de la agencia como fenómeno graduado (Bandi et al., 2025).

Un elemento central en la estructuración de niveles es la **progresión de la complejidad organizativa del comportamiento**. En los niveles iniciales, los sistemas presentan una coordinación limitada entre sus componentes, lo que se traduce en comportamientos fragmentados o parcialmente integrados. A medida que se avanza en la escala, se observa una mayor **articulación entre percepción, decisión y acción**, lo que permite identificar niveles superiores de agencia. **Cada nivel representa un incremento en la integración estructural del sistema, no en su rendimiento**, lo cual constituye una distinción fundamental en la medición (Sapkota et al., 2026).

La definición de niveles también requiere establecer **umbrales cualitativos que permitan diferenciar etapas del comportamiento**, evitando interpretaciones basadas únicamente en variaciones cuantitativas. En este sentido, la transición entre niveles implica cambios en la forma en que el sistema organiza su acción, lo que introduce una dimensión transformacional en la escala. **No se trata de medir más o menos, sino de identificar cambios en la naturaleza del comportamiento**, lo que refuerza el carácter estructural de la medición (Poole & Mackworth, 2017).

Otro aspecto fundamental en la estructuración de niveles es la **coherencia interna de cada nivel**, la cual garantiza que los sistemas clasificados dentro de una misma categoría compartan una lógica de comportamiento similar. Esto permite que la escala funcione como un marco analítico consistente, donde cada nivel representa una forma diferenciada de organización. **La coherencia interna evita que la escala sea arbitraria**, asegurando que las diferencias entre niveles respondan a cambios reales en la estructura del comportamiento (Russell, 2019). La estructuración de niveles requiere considerar la **continuidad entre etapas**, de manera que la escala refleje un proceso evolutivo del comportamiento. Cada nivel debe derivarse del anterior, incorporando nuevas capacidades de organización sin romper la lógica general del sistema. **La escala se construye como una trayectoria y no como una clasificación estática**, lo que permite interpretar la agencia como un proceso en desarrollo (Wang, 2025).

Cabe destacar, tro elemento relevante el cual es la **capacidad de diferenciación de la escala**, la cual debe ser lo suficientemente precisa para distinguir entre distintos grados de agencia sin perder coherencia conceptual. Esto implica definir niveles que permitan ubicar a los sistemas en posiciones específicas dentro del continuo, evitando tanto la generalización excesiva como la fragmentación innecesaria. **Una escala robusta es aquella que equilibra precisión y claridad**, facilitando la comparación entre sistemas (Guidotti et al., 2018). Además, la estructuración de niveles debe incorporar la **relación entre el sistema y su entorno**, ya que la agencia no puede analizarse de manera aislada. Los distintos niveles deben reflejar la capacidad del sistema para interactuar con condiciones variables, lo que introduce una dimensión contextual en la escala. **Cada nivel implica una forma distinta de relación con el entorno**, lo que amplía la comprensión de la agencia más allá de su estructura interna (Vinuesa et al., 2020).

La estructuración de niveles en las escalas de medición de la IA agéntica debe concebirse como un proceso dinámico que evoluciona junto con el desarrollo de los sistemas inteligentes. A medida que la IA avanza, las escalas deben ajustarse para incorporar nuevas formas de comportamiento, lo que implica una revisión constante de sus niveles. **La escala no es definitiva, es adaptativa**, lo que permite mantener su relevancia en un campo en constante transformación (Russell & Norvig, 2022). En conjunto, la estructuración de niveles constituye el núcleo operativo de las escalas de medición de la agencia. **No se trata de medir directamente, sino de ordenar la complejidad del comportamiento en niveles interpretables**, lo que convierte a la escala en una herramienta fundamental para el análisis académico de la IA agéntica (Adadi & Berrada, 2018).

Aplicación e interpretación de las escalas de medición de la agencia

La aplicación de las escalas de medición de la IA agéntica implica trasladar el marco conceptual y estructural hacia escenarios donde el comportamiento del sistema pueda ser observado e interpretado. En este sentido, las escalas no funcionan como instrumentos de cuantificación directa, sino como **marcos analíticos que permiten ubicar el comportamiento dentro de un continuo de agencia**. **La medición no asigna valores numéricos, posiciona al sistema en niveles de estructuración**, lo que permite comprender su comportamiento en términos organizativos y no únicamente funcionales (Bandi et al., 2025).

Uno de los principales usos de las escalas es la **comparación entre sistemas**, lo que permite identificar diferencias cualitativas en el grado de agencia. A diferencia de los enfoques centrados en el rendimiento, las escalas permiten analizar **cómo se organiza el comportamiento en distintos sistemas**, introduciendo una dimensión estructural en la comparación. **La escala establece un lenguaje común para interpretar diferencias**, lo que facilita la construcción de conocimiento sistemático en el campo de la inteligencia artificial (Poole & Mackworth, 2017).

La interpretación de las escalas también implica reconocer la **variabilidad del comportamiento en función del contexto**, ya que un sistema puede manifestar distintos niveles de agencia dependiendo de las condiciones en las que opera. En este sentido, la medición debe considerar la **dinamicidad de la interacción entre sistema y entorno**, evitando interpretaciones estáticas. **La agencia no es fija, es contextual**, lo que obliga a analizar el comportamiento en función de escenarios específicos (Vinuesa et al., 2020). En la aplicación de las escalas tenemnos otro aspecto relevante a mencion que es, la **identificación de trayectorias de comportamiento**, lo cual permite analizar la evolución del sistema a lo largo del tiempo. La agencia no se manifiesta de manera instantánea, sino como un proceso que se desarrolla progresivamente. **La escala permite observar el tránsito entre niveles de agencia**, lo que resulta clave para comprender procesos de aprendizaje, adaptación o transformación del comportamiento (Sapkota et al., 2026).

Asimismo, la aplicación de las escalas permite identificar **limitaciones estructurales en los sistemas de IA**, lo que contribuye a una comprensión más crítica del fenómeno agéntico. Al ubicar un sistema dentro de un nivel específico, es posible detectar debilidades en la organización del comportamiento, como falta de coherencia o dificultades de adaptación. **La escala no solo identifica niveles de agencia, también revela déficits estructurales**, lo que resulta fundamental para orientar futuras investigaciones (Russell & Norvig, 2022). La **interpretabilidad de los resultados**, es un determinate, ya que las escalas deben ser comprensibles desde una perspectiva académica y no depender exclusivamente de análisis técnicos. La medición de la agencia requiere traducir el comportamiento en categorías conceptuales que permitan su análisis. **La escala actúa como un puente entre observación y teoría**, facilitando la comprensión del comportamiento agéntico dentro de marcos analíticos más amplios (Guidotti et al., 2018).

Además, la aplicación de las escalas permite abordar la **complejidad de sistemas multiagente**, donde la agencia no se manifiesta únicamente a nivel individual, sino en la interacción entre múltiples entidades. En estos casos, la medición debe considerar la **dinámica colectiva del comportamiento**, lo que introduce un nivel adicional de análisis. **La agencia puede emerger de la interacción entre agentes**, lo que exige adaptar la interpretación de las escalas (Wang, 2025). La aplicación de las escalas de medición de la IA agéntica permite establecer un vínculo entre teoría y práctica, al ofrecer un marco que facilita la interpretación del comportamiento en contextos reales. **La medición deja de ser un ejercicio abstracto para convertirse en una herramienta de análisis contextual**, lo que permite comprender la agencia como un fenómeno dinámico, estructurado y evolutivo. **La escala no solo mide, interpreta y explica el comportamiento**, consolidándose como un instrumento clave en el estudio académico de la inteligencia artificial (Adadi & Berrada, 2018).

Estabilidad del comportamiento como base de medición

La estabilidad del comportamiento constituye uno de los pilares conceptuales más relevantes en la medición de la agencia en sistemas de inteligencia artificial. En este contexto, la estabilidad no debe interpretarse como inmovilidad o rigidez, sino como la **capacidad del sistema para sostener una lógica coherente de acción a lo largo del tiempo**, incluso cuando enfrenta variaciones en su entorno. **La estabilidad implica continuidad en la organización del comportamiento y no ausencia de cambio**, lo que introduce una distinción clave frente a enfoques tradicionales que asocian estabilidad con repetición mecánica. Esta comprensión permite situar la estabilidad como una propiedad estructural que emerge de la integración funcional del sistema (Bandi et al., 2025).

Desde esta perspectiva, la estabilidad permite diferenciar entre sistemas que presentan comportamientos episódicos y aquellos que manifiestan una organización sostenida del comportamiento. Mientras que los sistemas reactivos responden a estímulos de manera inmediata y sin continuidad, los sistemas con agencia evidencian una **persistencia en su lógica de acción**, lo que permite identificar patrones estructurados. **La estabilidad no reside en acciones individuales, sino en la continuidad de los patrones que organizan el comportamiento**, lo cual resulta fundamental para distinguir entre ejecución funcional y agencia (Poole & Mackworth, 2017). La agencia no puede ser inferida a partir de observaciones puntuales, ya que requiere analizar la evolución del comportamiento a lo largo del tiempo. En este sentido, la estabilidad introduce una **perspectiva longitudinal en la medición**, permitiendo evaluar si el sistema mantiene coherencia en diferentes momentos y condiciones. **La repetición consistente de patrones es lo que convierte al comportamiento en evidencia de agencia**, lo que refuerza la importancia de considerar el tiempo como una dimensión esencial del análisis (Wang, 2025).

Asimismo, la estabilidad del comportamiento se vincula con la capacidad del sistema para mantener dirección frente a cambios en el entorno. En este sentido, la agencia implica una forma de organización que permite al sistema adaptarse sin perder coherencia. **La estabilidad no se opone a la adaptabilidad, sino que la integra dentro de una lógica estructurada**, lo que permite analizar cómo el sistema ajusta su comportamiento sin fragmentarse. Este equilibrio entre cambio y continuidad constituye una de las características más relevantes del comportamiento agéntico (Vinuesa et al., 2020). Otro determinante clave es la relación entre estabilidad y orientación a objetivos. Un sistema que mantiene una lógica de acción coherente a lo largo del tiempo evidencia la existencia de una estructura orientada a fines. En este sentido, la estabilidad puede interpretarse como un indicador indirecto de la capacidad del sistema para organizar su comportamiento en función de objetivos. **La persistencia en la dirección del comportamiento revela la presencia de una lógica interna orientada**, lo que refuerza la idea de que la agencia no se mide por resultados, sino por la forma en que estos se estructuran (Russell, 2019).

Además, la estabilidad del comportamiento está estrechamente relacionada con la integración funcional de los componentes del sistema. Cuando percepción, decisión y acción operan de manera coordinada, el comportamiento tiende a ser más consistente y menos propenso a la fragmentación. **La estabilidad es una manifestación de la integración estructural del sistema**, ya que refleja la capacidad de sus componentes para operar bajo una misma lógica. Este aspecto resulta fundamental para comprender la agencia como una propiedad emergente del sistema y no como una característica aislada (Sapkota et al., 2026).

Un sistema estable tiende a generar patrones que pueden ser anticipados, no en términos deterministas, sino como expresiones de una lógica consistente. **La predictibilidad se deriva de la estabilidad estructural y no de la repetición mecánica**, lo que permite interpretar el comportamiento como resultado de una organización interna. Este aspecto contribuye a fortalecer la medición, al facilitar la identificación de regularidades en la acción (Guidotti et al., 2018). La estabilidad permite identificar la resiliencia del sistema frente a perturbaciones. Un sistema estable no es aquel que evita el cambio, sino aquel que es capaz de mantener su organización interna a pesar de las variaciones externas. **La resiliencia se convierte en una dimensión de la estabilidad**, ya que refleja la capacidad del sistema para sostener su comportamiento en condiciones adversas. Este enfoque amplía la comprensión de la estabilidad más allá de la consistencia, incorporando la capacidad de recuperación como parte de la medición (Russell & Norvig, 2022).

En conclusión, la estabilidad del comportamiento debe entenderse como una propiedad emergente que resulta de la interacción entre múltiples factores dentro del sistema. No es un atributo aislado, sino el resultado de la organización del sistema en su conjunto. **La estabilidad refleja la capacidad del sistema para sostener comportamiento coherente, continuo y adaptativo en el tiempo**, lo que la convierte en una base fundamental para la medición de la agencia. En este sentido, la estabilidad no solo permite identificar la presencia de agencia, sino también comprender su grado de estructuración, consolidándose como uno de los criterios más robustos en el análisis de sistemas de inteligencia artificial (Adadi & Berrada, 2018).

La estabilidad como criterio estructural de medición de la agencia

La estabilidad del comportamiento adquiere un papel central como criterio de medición en la IA agéntica al permitir evaluar la **consistencia estructural del sistema frente a variaciones del entorno**. A diferencia de los fundamentos conceptuales, que explican qué es la estabilidad, esta sección se centra en **cómo la estabilidad opera como criterio evaluativo**, es decir, como un parámetro que permite determinar si un sistema mantiene una organización coherente en su comportamiento. **La estabilidad no se describe, se aplica como criterio para medir la agencia**, lo que la convierte en un eje metodológico en la evaluación de sistemas inteligentes (Bandi et al., 2025).

Uno de los elementos fundamentales de este criterio es la **consistencia del comportamiento bajo condiciones variables**, lo cual implica que el sistema debe mantener una lógica de acción coherente incluso cuando el entorno cambia. Un sistema que presenta variaciones abruptas, contradictorias o desarticuladas en su comportamiento no puede considerarse estable. **La estabilidad se expresa como regularidad estructurada**, lo que significa que los patrones de comportamiento mantienen una lógica interna reconocible. Este criterio permite distinguir entre consistencia organizativa y simple repetición mecánica de respuestas (Poole & Mackworth, 2017). Una característica importante clave, es la **capacidad de sostener direccionalidad en la acción**, entendida como la permanencia de una orientación hacia objetivos a lo largo del tiempo. Este criterio permite evaluar si el comportamiento del sistema responde a una estructura interna o si se encuentra sujeto a fluctuaciones circunstanciales. **La estabilidad implica continuidad en la orientación del comportamiento**, lo que constituye una evidencia de organización interna y no solo de ejecución funcional. Esta dimensión resulta esencial para identificar la presencia de agencia en sistemas dinámicos (Russell, 2019).

La estabilidad como criterio permite analizar la **relación entre adaptación y coherencia**, lo cual resulta fundamental en entornos cambiantes. La capacidad de adaptación no debe confundirse con variabilidad desorganizada; por el contrario, la adaptación debe integrarse dentro de una lógica coherente de comportamiento. **La adaptación estructurada refuerza la estabilidad**, mientras que la adaptación caótica la debilita. Este criterio permite diferenciar entre sistemas flexibles con organización interna y sistemas que reaccionan sin mantener consistencia (Vinuesa et al., 2020). La **evaluación de la continuidad del comportamiento a lo largo del tiempo**, se debe tomar en ciuenta ya que permite identificar si el sistema mantiene su lógica de acción en diferentes momentos. La estabilidad no puede evaluarse de forma instantánea, sino que requiere observar el comportamiento en una secuencia temporal. **La continuidad se convierte en una condición necesaria para la estabilidad**, ya que permite verificar la persistencia de la organización del comportamiento. Este criterio introduce una dimensión temporal en la medición de la agencia (Wang, 2025).

La estabilidad también permite identificar **rupturas estructurales en el comportamiento**, las cuales constituyen indicadores clave para evaluar los límites del sistema. Estas rupturas pueden manifestarse como incoherencias, discontinuidades o cambios abruptos en la lógica de acción. **La identificación de rupturas permite evaluar la fragilidad del sistema**, lo que aporta una dimensión crítica a la medición. En este sentido, la estabilidad no se mide solo por su presencia, sino también por la capacidad de detectar sus fallas (Guidotti et al., 2018). La **predictibilidad estructural del comportamiento**, entendida como la posibilidad de anticipar patrones de acción a partir de la consistencia del sistema, es otro factor a considerar. Un sistema estable tiende a generar regularidades que reflejan una lógica interna organizada. **La predictibilidad no implica determinismo, sino coherencia estructural**, lo que permite interpretar el comportamiento como resultado de una organización subyacente. Este criterio fortalece la capacidad analítica de la medición (Poole & Mackworth, 2017).

Además, la estabilidad como criterio permite evaluar la **integración funcional del sistema**, es decir, la coordinación entre sus distintos componentes. Cuando percepción, decisión y acción operan de manera articulada, el comportamiento tiende a ser más consistente y menos propenso a la fragmentación. **La estabilidad refleja el grado de integración estructural del sistema**, lo que la convierte en un indicador clave de la organización interna (Sapkota et al., 2026). Asimismo, la estabilidad puede ser interpretada como un indicador de **resiliencia del sistema**, entendida como la capacidad de mantener su organización frente a perturbaciones externas. Un sistema resiliente no evita el cambio, sino que es capaz de adaptarse sin perder coherencia. **La resiliencia amplía el concepto de estabilidad**, incorporando la capacidad de recuperación como parte de la medición de la agencia. Este enfoque resulta especialmente relevante en entornos dinámicos (Russell & Norvig, 2022).

La estabilidad como criterio de medición permite evaluar la permanencia estructural de las dimensiones previamente definidas. La estabilidad como criterio de medición no consiste en definir las dimensiones del comportamiento, sino en evaluar si estas dimensiones logran sostenerse de manera consistente a lo largo del tiempo y en distintos contextos. En este sentido, la estabilidad permite verificar si la organización del comportamiento mantiene su coherencia, continuidad y direccionalidad bajo condiciones variables, evitando interpretaciones basadas en manifestaciones aisladas. La estabilidad no introduce nuevas dimensiones, sino que **valida la permanencia estructural de las ya definidas**, convirtiéndose en un mecanismo de verificación de la agencia. De este modo, la medición basada en estabilidad no se orienta a identificar qué características posee el sistema, sino a determinar si dichas características se mantienen de forma consistente, lo que permite diferenciar entre comportamiento estructurado y comportamiento circunstancial (Kadir et al., 2025).

Desde una perspectiva aplicada, la estabilidad como criterio adquiere relevancia al analizar el desempeño de sistemas agénticos en entornos dinámicos, donde la consistencia del comportamiento es fundamental para garantizar su funcionamiento confiable. En estos contextos, la estabilidad permite evaluar la robustez operativa del sistema, es decir, su capacidad para sostener su lógica de acción frente a variaciones e incertidumbre. **La estabilidad actúa como condición de validación en escenarios reales**, lo que refuerza su papel como criterio central en la medición de la agencia. Así, la evaluación no se limita a identificar capacidades, sino a comprobar su permanencia efectiva en condiciones complejas, consolidando la estabilidad como un eje clave en la interpretación del comportamiento agéntico (World Economic Forum, 2025).

Estabilidad conductual en la medición de la agencia

La interpretación aplicada de la estabilidad del comportamiento en la medición de la IA agéntica implica trasladar el criterio estructural hacia escenarios empíricos donde el comportamiento pueda ser observado, analizado y comprendido en su contexto. En este sentido, la estabilidad no se limita a una propiedad teórica, sino que se convierte en un **indicador observable que permite inferir la organización interna del sistema a partir de su comportamiento sostenido**. **La medición se transforma en un**

proceso interpretativo del comportamiento en el tiempo, lo que permite ubicar al sistema dentro de un continuo de agencia en función de su coherencia estructural (Vinuesa et al., 2020).

Uno de los principales usos interpretativos de la estabilidad es la **identificación de patrones consistentes de comportamiento**, los cuales permiten reconocer la existencia de una estructura organizativa en el sistema. Estos patrones no deben entenderse como repeticiones mecánicas, sino como manifestaciones de una lógica interna que organiza la acción. **La estabilidad convierte la recurrencia en evidencia de organización**, lo que permite diferenciar entre comportamiento circunstancial y comportamiento estructurado. Este enfoque fortalece la validez de la medición al basarse en regularidades observables (Guidotti et al., 2018).

La interpretación de la estabilidad también requiere considerar la **variabilidad contextual del comportamiento**, ya que un sistema puede manifestar distintos niveles de coherencia dependiendo de las condiciones en las que opera. En este sentido, la medición debe incorporar el contexto como elemento central del análisis, evitando interpretaciones aisladas o descontextualizadas. **La estabilidad no es absoluta, es relativa a las condiciones de operación**, lo que implica que su evaluación debe realizarse en múltiples escenarios para asegurar su consistencia (Wang, 2025). Así, se debe considerar,mla **evaluación de trayectorias de comportamiento**, lo cual permite analizar la evolución del sistema a lo largo del tiempo. La estabilidad no es una condición estática, sino un proceso dinámico que puede fortalecerse, mantenerse o deteriorarse. **El análisis longitudinal permite identificar procesos de consolidación o pérdida de agencia**, lo que amplía el alcance de la medición al incorporar la dimensión temporal como parte del análisis estructural (Sapkota et al., 2026).

Asimismo, la interpretación de la estabilidad permite identificar **niveles de madurez en el comportamiento agéntico**, ya que los sistemas más estables tienden a mostrar una organización más consolidada. En este sentido, la estabilidad puede ser utilizada como un indicador del grado de desarrollo de la agencia, permitiendo diferenciar entre sistemas en etapas iniciales y sistemas con mayor nivel de estructuración. **La estabilidad refleja el nivel de consolidación del comportamiento**, lo que introduce una dimensión evolutiva en la medición (Bandi et al., 2025). Además, es importante tomar en cuenta la **detección de inconsistencias en el comportamiento**, lo cual permite identificar los límites de la agencia del sistema. La presencia de rupturas, discontinuidades o contradicciones en la lógica de acción indica que la organización interna del sistema no es completamente estable. **La interpretación de la estabilidad incluye el análisis de sus fallas**, lo que permite evaluar no solo la presencia de agencia, sino también su robustez. Este enfoque introduce una dimensión crítica en la medición (Russell, 2019).

La estabilidad permite analizar la **relación entre predictibilidad y coherencia**, ya que los sistemas con mayor estabilidad tienden a generar patrones de comportamiento que pueden ser anticipados. Esta predictibilidad no implica determinismo, sino la

existencia de una lógica estructural que organiza la acción. **La predictibilidad emerge como consecuencia de la estabilidad**, lo que permite interpretar el comportamiento como resultado de una organización interna consistente (Poole & Mackworth, 2017).

En la interpretación debemos incluir la **distinción entre estabilidad aparente y estabilidad estructural**, ya que un sistema puede parecer coherente en un entorno limitado, pero no sostener esa coherencia en condiciones variables. La medición requiere evaluar la estabilidad en diferentes contextos para determinar si se trata de una propiedad real o circunstancial. **La estabilidad auténtica se manifiesta en la persistencia del comportamiento en múltiples escenarios**, lo que refuerza la necesidad de un análisis contextual amplio (Russell & Norvig, 2022).

La interpretación de la estabilidad permite analizar la **resiliencia del comportamiento**, entendida como la capacidad del sistema para mantener su organización frente a perturbaciones. Un sistema resiliente no evita el cambio, sino que es capaz de adaptarse sin perder coherencia. **La resiliencia amplía la noción de estabilidad**, incorporando la capacidad de recuperación como parte del análisis de la agencia, lo que resulta fundamental en entornos dinámicos (Wang, 2025). La interpretación aplicada de la estabilidad permite establecer un vínculo entre teoría y práctica, al ofrecer un criterio que puede ser utilizado para analizar el comportamiento en contextos reales. **La estabilidad convierte la medición en un proceso comprensible, contextual y aplicable**, lo que permite integrar la observación empírica con los marcos conceptuales de la agencia. En este sentido, la estabilidad no solo permite medir la agencia, sino comprender su manifestación en sistemas complejos, consolidándose como un criterio central en el análisis de la IA agéntica (Adadi & Berrada, 2018).

Evaluación operativa de la estabilidad conductual

La evaluación operativa de la estabilidad del comportamiento constituye un nivel intermedio entre la definición conceptual y la interpretación aplicada de la agencia en sistemas de inteligencia artificial. En este sentido, la estabilidad deja de ser únicamente un criterio teórico y se convierte en un **referente operativo que permite analizar cómo se manifiesta la consistencia del comportamiento en condiciones específicas de funcionamiento**. La evaluación no se centra en la identificación de atributos abstractos, sino en la observación de la capacidad del sistema para sostener su lógica de acción en escenarios concretos, lo que permite trasladar la medición hacia un plano funcional sin perder su base estructural (OECD, 2022).

Desde esta perspectiva, la estabilidad operativa implica analizar la capacidad del sistema para mantener coherencia en situaciones donde existen múltiples variables en interacción. A diferencia de los entornos controlados, los contextos reales introducen condiciones de incertidumbre, lo que exige evaluar si el comportamiento del sistema conserva su organización interna. **La estabilidad operativa no se define por la ausencia de variación, sino por la capacidad de sostener coherencia bajo variabilidad**, lo que permite distinguir entre sistemas que operan de manera

estructurada y aquellos que presentan respuestas inconsistentes (World Economic Forum, 2025).

Un elemento central en esta evaluación es la capacidad del sistema para sostener patrones de comportamiento en tareas repetidas bajo condiciones cambiantes. La repetición de tareas no implica necesariamente estabilidad, ya que un sistema puede reproducir acciones sin mantener coherencia en su lógica interna. En este sentido, la evaluación operativa exige identificar si el comportamiento responde a una estructura consistente y no a la simple reiteración de respuestas. **La estabilidad se verifica cuando el sistema mantiene patrones organizados en contextos variables**, lo que permite validar la presencia de agencia más allá de ejecuciones puntuales (Kadir et al., 2025). La evaluación de la estabilidad implica analizar la relación entre comportamiento esperado y comportamiento observado. En este proceso, se busca determinar si el sistema mantiene una correspondencia entre su lógica interna y su desempeño en escenarios específicos. La estabilidad operativa se manifiesta cuando el comportamiento observado refleja una continuidad en la organización del sistema, evitando desviaciones significativas que indiquen fragmentación. **La consistencia entre intención estructural y ejecución observable es clave para validar la estabilidad**, lo que fortalece la medición en contextos prácticos (Ng, 2018). Así debemos tomar en cuenta, la capacidad del sistema para mantener estabilidad en condiciones de presión o perturbación. En entornos dinámicos, los sistemas agénticos deben operar bajo condiciones que pueden afectar su desempeño, lo que exige evaluar si su comportamiento conserva coherencia. **La estabilidad operativa incorpora la capacidad de resistir perturbaciones sin perder organización**, lo que introduce una dimensión de robustez en la medición de la agencia (Russell & Norvig, 2022).

La evaluación operativa permite identificar la consistencia intercontextual del comportamiento, es decir, la capacidad del sistema para mantener su lógica de acción en distintos escenarios. Un sistema verdaderamente estable no depende de un contexto específico, sino que puede sostener su organización en condiciones diversas. **La estabilidad intercontextual permite validar la generalización del comportamiento**, lo que refuerza su valor como indicador de agencia (Vinuesa et al., 2020). La evaluación operativa también implica analizar la capacidad del sistema para recuperar su coherencia tras desviaciones en su comportamiento. En este sentido, la estabilidad no se mide únicamente por la ausencia de errores, sino por la capacidad de corregirlos y retornar a una lógica estructurada. **La capacidad de recuperación constituye un indicador clave de estabilidad**, ya que refleja la persistencia de la organización interna del sistema (Guidotti et al., 2018).

Por lo anterior, otro elemento clave en la evaluación es la identificación de límites de estabilidad, lo que permite determinar en qué condiciones el sistema pierde coherencia. Este análisis resulta fundamental para comprender no solo la presencia de estabilidad, sino también su alcance. **La estabilidad no es absoluta, sino dependiente de condiciones específicas**, lo que exige delimitar su rango de funcionamiento dentro del proceso de medición (Sapkota et al., 2026). La evaluación

operativa de la estabilidad permite consolidar la medición de la agencia como un proceso que integra teoría y práctica. Al analizar cómo el sistema sostiene su comportamiento en condiciones reales, la estabilidad se convierte en un **criterio verificable que conecta la estructura del comportamiento con su manifestación empírica**. De este modo, la medición deja de ser un ejercicio abstracto y se transforma en un proceso aplicado que permite comprender la agencia como un fenómeno dinámico, observable y estructurado (Adadi & Berrada, 2018).

Evidencia empírica de la medición de la IA agéntica

La evidencia empírica de la medición de la IA agéntica se centra en la identificación de manifestaciones observables del comportamiento estructurado en contextos reales. A diferencia de aproximaciones puramente teóricas, este enfoque implica analizar cómo los sistemas despliegan **patrones consistentes de acción que pueden ser reconocidos en su interacción con el entorno**. La evidencia no se limita a resultados finales, sino que considera la forma en que el sistema organiza su comportamiento a lo largo del tiempo, permitiendo identificar **regularidades, continuidad y capacidad de ajuste**. En este sentido, la observación empírica busca distinguir entre respuestas aisladas y comportamientos que reflejan una **estructura interna coherente**. Asimismo, la medición requiere situar al sistema en escenarios dinámicos donde se ponga a prueba su capacidad de mantener dirección frente a cambios, lo que permite validar la presencia de agencia. De esta manera, la evidencia empírica se convierte en un elemento clave para sustentar la medición, al ofrecer indicios concretos de que el comportamiento del sistema no es circunstancial, sino resultado de una organización sostenida.

Evidencia empírica del comportamiento agéntico

La evidencia empírica en la medición de la IA agéntica se fundamenta en la observación sistemática del comportamiento del sistema en contextos específicos de operación, lo que permite trasladar la medición desde el plano conceptual hacia un nivel verificable. En este sentido, la evidencia no se construye a partir de supuestos teóricos, sino mediante la identificación de manifestaciones observables que reflejan la organización interna del sistema. **La evidencia empírica se define como la manifestación verificable de patrones de comportamiento estructurado**, lo que permite inferir la presencia de agencia a partir de la acción del sistema en condiciones concretas (Adadi & Berrada, 2018).

Desde esta perspectiva, la evidencia empírica implica identificar patrones de comportamiento que se repiten de manera consistente en distintos momentos de observación. Sin embargo, esta repetición no debe interpretarse como simple reiteración mecánica, sino como expresión de una lógica interna que organiza la acción. **La evidencia empírica se sustenta en la recurrencia estructurada del comportamiento**, lo que permite diferenciar entre comportamiento circunstancial y

comportamiento organizado, constituyendo así una base sólida para la medición (Guidotti et al., 2018).

La construcción de evidencia empírica también requiere considerar el contexto en el que el sistema opera, ya que el comportamiento no puede interpretarse de manera aislada. En este sentido, la validez de la evidencia depende de la capacidad de relacionar el comportamiento observado con las condiciones en las que se produce. **La evidencia empírica es inherentemente contextual**, lo que implica que su interpretación debe integrar variables ambientales, operativas y situacionales para evitar conclusiones reduccionistas (Vinuesa et al., 2020). lla evidencia empírica se fortalece mediante el análisis longitudinal del comportamiento, lo que permite observar la evolución del sistema a lo largo del tiempo. La agencia no puede inferirse a partir de eventos aislados, sino que requiere identificar la persistencia de patrones en diferentes momentos. **La dimensión temporal es fundamental para la evidencia empírica**, ya que permite distinguir entre comportamiento transitorio y comportamiento estructurado, consolidando la validez de la medición (Wang, 2025).

La construcción de evidencia empírica debe considerar la comparación entre diferentes instancias de comportamiento, lo que permite identificar regularidades que no son evidentes en observaciones individuales. Este enfoque comparativo contribuye a fortalecer la consistencia de la evidencia, ya que permite validar si los patrones observados se mantienen en distintos escenarios. **La evidencia empírica se consolida mediante la comparación sistemática**, lo que permite construir una base analítica más robusta (Poole & Mackworth, 2017). La evidencia empírica también implica la identificación de desviaciones en el comportamiento, ya que estas permiten delimitar los límites de la agencia del sistema. Las inconsistencias, discontinuidades o rupturas en la lógica de acción constituyen información relevante para comprender el grado de estructuración del comportamiento. **Las desviaciones forman parte de la evidencia empírica**, ya que permiten evaluar tanto la presencia como las limitaciones de la agencia en el sistema (Russell, 2019).

Desde un enfoque metodológico, la evidencia empírica requiere la definición de procedimientos de observación que permitan registrar el comportamiento del sistema de manera sistemática y reproducible. Estos procedimientos deben garantizar la consistencia de la información recolectada, evitando sesgos en la interpretación. **La calidad de la evidencia empírica depende del rigor en su recolección**, lo que implica establecer protocolos claros para la observación y el análisis del comportamiento (Kadir et al., 2025).

Además, la evidencia empírica permite analizar la correspondencia entre el comportamiento esperado y el comportamiento observado, lo que constituye un elemento clave para la validación de la agencia. Cuando existe coherencia entre ambos, se refuerza la interpretación del comportamiento como estructurado. **La correspondencia entre expectativa y observación es un indicador de validez empírica**, lo que permite evaluar la consistencia del sistema en términos operativos (Ng, 2018).

La capacidad de la evidencia empírica debe considerar su capacidad integrarse en marcos de clasificación más amplios, donde el comportamiento observado permite ubicar al sistema dentro de tipologías específicas. Este enfoque facilita la sistematización del análisis, al permitir comparar diferentes sistemas en función de su comportamiento observable. **La evidencia empírica contribuye a la clasificación estructural del comportamiento**, lo que fortalece su utilidad en la medición de la agencia (OECD, 2022). La evidencia empírica adquiere una dimensión aplicada cuando se analiza el comportamiento del sistema en entornos reales, donde las condiciones son dinámicas e inciertas. En estos contextos, la observación del comportamiento permite evaluar si el sistema mantiene coherencia operativa frente a variaciones del entorno. **La evidencia empírica valida la agencia en escenarios reales**, lo que la convierte en un componente esencial para la evaluación de sistemas agénticos (World Economic Forum, 2025).

Concluyendo, la evidencia empírica en la medición de la IA agéntica se consolida como un proceso que integra observación, análisis e interpretación del comportamiento. No se trata únicamente de registrar acciones, sino de comprender la lógica que las organiza. **La evidencia empírica permite inferir la estructura interna del sistema a partir de su comportamiento observable**, lo que la convierte en un elemento central en la evaluación de la agencia y en la comprensión de sistemas inteligentes complejos (Sapkota et al., 2026).

Validación empírica y contraste del comportamiento agéntico

La validación empírica en la medición de la IA agéntica constituye el proceso mediante el cual la evidencia observada del comportamiento del sistema es sometida a contraste para determinar su consistencia estructural. En este sentido, la medición no se limita a la identificación de patrones, sino que exige verificar si dichos patrones reflejan una organización interna sostenida. **La validación empírica implica contrastar el comportamiento observado con criterios de consistencia estructural**, lo que permite distinguir entre manifestaciones aparentes de agencia y comportamientos efectivamente organizados (Kadir et al., 2025). La validación empírica requiere establecer mecanismos de comparación entre diferentes instancias de comportamiento del sistema. Este proceso permite evaluar si las regularidades observadas se mantienen en distintos momentos y condiciones, evitando interpretaciones basadas en casos aislados. **La validación se construye a partir del contraste sistemático del comportamiento**, lo que fortalece la confiabilidad de la medición al reducir la posibilidad de sesgos interpretativos (Poole & Mackworth, 2017).

La consistencia intersituacional del comportamiento, es un elemento central en este proceso, es decir, la capacidad del sistema para mantener su lógica de acción en distintos escenarios. La validación empírica exige analizar si el comportamiento conserva su organización interna cuando se modifica el entorno. **La consistencia en diferentes contextos constituye evidencia de validez estructural**, lo que permite confirmar la presencia de agencia más allá de condiciones específicas (OECD, 2022).

Asimismo, la validación empírica implica analizar la correspondencia entre el comportamiento esperado y el comportamiento observado, lo que permite evaluar si el sistema actúa conforme a su estructura interna. Este contraste resulta fundamental para determinar si el comportamiento refleja una organización coherente o si responde a variaciones circunstanciales. **La validación empírica se sustenta en la congruencia entre expectativa y ejecución**, lo que refuerza la interpretación del comportamiento como estructurado (Ng, 2018).

Otro aspecto relevante es la evaluación de la robustez del comportamiento frente a condiciones adversas o cambios inesperados. En este contexto, la validación empírica permite analizar si el sistema mantiene coherencia cuando enfrenta perturbaciones, lo que constituye una prueba crítica de su organización interna. **La robustez operativa es un indicador clave de validez empírica**, ya que refleja la capacidad del sistema para sostener su lógica de acción en condiciones exigentes (World Economic Forum, 2025). La validación también implica identificar los límites del comportamiento agéntico, lo que permite determinar en qué condiciones el sistema pierde coherencia. Este análisis es fundamental para comprender el alcance de la agencia, ya que no todos los comportamientos observados pueden considerarse estructurados. **La validación empírica incluye la delimitación de condiciones de fallo**, lo que permite construir una evaluación más precisa del comportamiento del sistema (Russell, 2019).

Además, la validación empírica se apoya en la replicabilidad del comportamiento, es decir, la capacidad del sistema para reproducir patrones consistentes bajo condiciones similares. La replicabilidad permite confirmar que la evidencia observada no es producto de circunstancias particulares, sino de una organización interna estable. **La replicabilidad refuerza la validez de la evidencia empírica**, lo que contribuye a consolidar la medición de la agencia (Guidotti et al., 2018).

Así también se debe considerar, la capacidad de la validación empírica para integrar múltiples fuentes de evidencia, lo que permite construir una evaluación más completa del comportamiento. La combinación de diferentes observaciones contribuye a fortalecer la consistencia del análisis, evitando interpretaciones parciales. **La validación empírica se fortalece mediante la integración de evidencias**, lo que permite construir una visión más robusta del comportamiento del sistema (Vinuesa et al., 2020). La validación empírica implica analizar la evolución del comportamiento del sistema a lo largo del tiempo, lo que permite identificar procesos de consolidación o deterioro de la agencia. Este enfoque longitudinal permite evaluar si la organización del comportamiento se mantiene o se transforma en función de las condiciones de operación. **La validación empírica incorpora una dimensión temporal**, lo que amplía el alcance de la medición (Wang, 2025).

La validación también implica la coherencia entre los distintos componentes del sistema, ya que la desarticulación funcional puede afectar la consistencia del comportamiento. En este sentido, la validación empírica permite evaluar si los procesos del sistema operan de manera coordinada. **La coherencia funcional es un indicador de validez estructural**, lo que refuerza la interpretación del comportamiento

como organizado (Sapkota et al., 2026). La validación empírica en la medición de la IA agéntica se consolida como un proceso que combina contraste, comparación y verificación del comportamiento observado. **La validación transforma la evidencia en conocimiento verificable**, lo que permite diferenciar entre comportamiento aparente y comportamiento estructurado. De este modo, la medición de la agencia se fundamenta en la capacidad de confirmar que el comportamiento del sistema responde a una organización interna coherente, consolidando la validez empírica como un componente esencial del análisis (Adadi & Berrada, 2018).

Evidencia empírica en entornos reales y complejidad operativa

La evidencia empírica en la medición de la IA agéntica alcanza su máxima relevancia cuando se analiza el comportamiento del sistema en entornos reales, donde las condiciones de operación son dinámicas, inciertas y altamente variables. En estos contextos, la evidencia no puede derivarse de condiciones controladas, sino que debe construirse a partir de la observación del sistema en interacción con múltiples factores simultáneos. **La evidencia empírica en entornos reales implica evaluar la manifestación del comportamiento en escenarios complejos**, lo que permite comprender la agencia como un fenómeno situado y no abstracto (World Economic Forum, 2025).

Desde esta perspectiva, los entornos reales introducen niveles de complejidad que obligan a replantear la forma en que se interpreta el comportamiento del sistema. A diferencia de los entornos controlados, donde las variables pueden ser aisladas, los escenarios reales implican interacciones múltiples que afectan el comportamiento de manera simultánea. **La evidencia empírica en contextos complejos se construye a partir de la interacción entre el sistema y su entorno**, lo que permite analizar cómo se sostiene la organización del comportamiento en condiciones no controladas (Vinuesa et al., 2020). Un determinante clave en este análisis es la capacidad del sistema para mantener coherencia operativa cuando se enfrenta a condiciones cambiantes. En entornos reales, el sistema debe adaptarse continuamente, lo que exige evaluar si su comportamiento conserva una lógica estructurada. **La evidencia empírica se manifiesta en la continuidad de la organización del comportamiento bajo variabilidad**, lo que permite inferir la presencia de agencia en condiciones dinámicas (Wang, 2025).

La evidencia empírica en entornos complejos permite identificar la capacidad del sistema para gestionar múltiples objetivos de manera simultánea. En contextos reales, los sistemas agénticos no operan bajo una única tarea, sino que deben responder a demandas diversas. **La capacidad de sostener coherencia en múltiples objetivos constituye evidencia de organización estructural**, lo que amplía la comprensión de la agencia más allá de comportamientos simples (Russell, 2019). Así, tenemos otro aspecto relevante, que es la interacción del sistema con otros sistemas o agentes, lo que introduce una dimensión relacional en la evidencia empírica. En estos casos, el comportamiento no solo depende de la lógica interna del sistema, sino también de las dinámicas de interacción. **La evidencia empírica en contextos interactivos se**

construye a partir de la capacidad del sistema para sostener coherencia en relaciones dinámicas, lo que permite analizar la agencia en entornos colaborativos o competitivos (Sapkota et al., 2026).

Además, los entornos reales permiten observar cómo el sistema responde a condiciones de incertidumbre, donde la información disponible es incompleta o ambigua. En estos casos, la evidencia empírica se construye a partir de la capacidad del sistema para tomar decisiones consistentes a pesar de la falta de información completa. **La gestión de la incertidumbre constituye una fuente clave de evidencia empírica**, ya que refleja la capacidad del sistema para sostener su lógica de acción en condiciones adversas (Poole & Mackworth, 2017).

La evidencia empírica en contextos reales también implica analizar la evolución del comportamiento del sistema en escenarios prolongados, donde la interacción continua permite observar procesos de consolidación o deterioro. Este enfoque permite identificar si la agencia se mantiene, se fortalece o se debilita con el tiempo. **La evidencia empírica en entornos reales incorpora una dimensión evolutiva**, lo que permite comprender la agencia como un proceso dinámico (Bandi et al., 2025).

Así, es de mencionar también, la capacidad del sistema para mantener consistencia en contextos heterogéneos, donde las condiciones de operación varían significativamente. La evidencia empírica se fortalece cuando el sistema demuestra que su comportamiento no depende de un entorno específico. **La consistencia en contextos diversos constituye evidencia de generalización del comportamiento**, lo que refuerza la interpretación de la agencia como propiedad estructural (OECD, 2022). La evidencia empírica en entornos complejos permite identificar la capacidad del sistema para recuperarse de fallos o desviaciones en su comportamiento. En estos contextos, la estabilidad no se mide por la ausencia de errores, sino por la capacidad de corregirlos. **La capacidad de recuperación constituye evidencia de resiliencia operativa**, lo que amplía la comprensión de la agencia en escenarios reales (Guidotti et al., 2018).

Por otro lado, la evidencia empírica permite analizar la correspondencia entre el comportamiento del sistema y las expectativas de funcionamiento en entornos reales, lo que permite evaluar su desempeño de manera contextualizada. Este análisis resulta fundamental para determinar si el sistema opera de acuerdo con su diseño estructural. **La correspondencia contextual refuerza la validez de la evidencia empírica**, lo que permite integrar teoría y práctica en la medición de la agencia (Ng, 2018).

Así, se tienen que, la evidencia empírica en la medición de la IA agéntica se consolida en entornos reales como un proceso que integra complejidad, interacción y variabilidad. **La evidencia empírica en contextos reales permite validar la agencia como fenómeno dinámico y estructurado**, lo que la convierte en un elemento esencial para la comprensión de sistemas inteligentes en escenarios operativos complejos (Adadi & Berrada, 2018).

Conclusiones

El desarrollo del Capítulo 5 permite afirmar que la medición de la inteligencia artificial agéntica constituye un **cambio epistemológico profundo en la forma de evaluar sistemas inteligentes**, al desplazar el énfasis desde el rendimiento técnico hacia la **organización estructural del comportamiento en el tiempo y en relación con el entorno**. Este cambio implica reconocer que la inteligencia no puede reducirse a resultados puntuales, sino que debe entenderse como un proceso dinámico, coherente y contextual que emerge de la interacción entre múltiples componentes .

En este sentido, se concluye que la agencia no es una característica inherente a todos los sistemas de IA, sino una **propiedad emergente que depende de la integración funcional entre percepción, decisión y acción**. Esta integración permite que el sistema sostenga una lógica organizativa consistente, lo que diferencia a los sistemas agénticos de aquellos que simplemente ejecutan tareas de manera aislada. Por tanto, la medición debe centrarse en la estructura del comportamiento y no en la eficiencia de sus resultados.

Asimismo, el capítulo demuestra que la medición de la IA agéntica es fundamentalmente un **proceso interpretativo**, en el cual la observación de patrones estructurados sustituye a la evaluación tradicional basada en indicadores de desempeño. Este enfoque permite identificar la coherencia interna del sistema, su continuidad en el tiempo y su capacidad de adaptación frente a cambios en el entorno. En consecuencia, la medición deja de ser un ejercicio cuantitativo simplificado para convertirse en un análisis estructural del comportamiento.

Los criterios estructurales —coherencia, continuidad temporal, autonomía operativa y estabilidad del comportamiento— se consolidan como **dimensiones fundamentales para delimitar la agencia**, ya que permiten identificar la presencia de organización interna en el sistema. En particular, la estabilidad del comportamiento emerge como un criterio integrador, al permitir validar la persistencia de estas dimensiones en contextos variables, lo que refuerza la idea de que la agencia se manifiesta como continuidad organizada y no como eventos aislados.

Por otra parte, la operacionalización de la medición pone en evidencia que la agencia no puede ser capturada mediante indicadores simples, sino que requiere la construcción de **unidades de análisis complejas, trayectorias de comportamiento e indicadores interpretativos**. Este proceso introduce desafíos metodológicos significativos, pero al mismo tiempo fortalece la capacidad analítica del enfoque, evitando reduccionismos y permitiendo una comprensión más profunda del fenómeno.

Adicionalmente, las escalas de medición permiten conceptualizar la agencia como un **continuum evolutivo**, superando las clasificaciones dicotómicas tradicionales. Este enfoque facilita la comparación entre sistemas y permite analizar el grado de

estructuración del comportamiento, reconociendo que la agencia se desarrolla progresivamente en función de la integración funcional del sistema.

Finalmente, la evidencia empírica se posiciona como un elemento central en la validación de la medición, ya que permite inferir la organización interna del sistema a partir de su comportamiento observable. La consistencia intercontextual, la persistencia temporal y la identificación de patrones recurrentes constituyen los principales fundamentos de esta validación, consolidando la relación entre teoría y práctica.

En conjunto, se concluye que **medir la IA agéntica implica analizar la organización del comportamiento como un fenómeno estructurado, dinámico y contextual**, lo que redefine los marcos tradicionales de evaluación y abre nuevas posibilidades para el estudio de sistemas inteligentes complejos desde una perspectiva más integradora y robusta. Ver **Tabla 5.**

Tabla 5. Medición estructural de la IA agéntica

Concepto	Definición	Diferenciación	Ventajas	Limitaciones	Referencias
Medición estructural de la IA	Evaluación del comportamiento como organización coherente en el tiempo y contexto	Se diferencia de medición tradicional centrada en resultados	Permite analizar la inteligencia como estructura dinámica	Alta complejidad interpretativa	Bandi et al. (2025); Sapkota et al. (2026)
Agencia en IA	Integración funcional de percepción, decisión y acción dentro de una lógica coherente	Se diferencia de automatización y sistemas reactivos	Permite comportamiento autónomo y organizado	Difícil delimitación conceptual	Russell (2019); Poole & Mackworth (2017)
Coherencia estructural	Consistencia interna del comportamiento del sistema	Se diferencia de respuestas aisladas o fragmentadas	Permite identificar organización del comportamiento	Complejidad en su medición	Guidotti et al. (2018); Adadi & Berrada (2018)
Continuidad temporal	Persistencia del comportamiento a lo largo del tiempo	Se diferencia de acciones episódicas	Permite análisis longitudinal del sistema	Requiere observación prolongada	Wang (2025); Vinuesa et al. (2020)
Autonomía operativa	Capacidad del sistema para actuar sin intervención constante	Se diferencia de sistemas dirigidos externamente	Permite independencia funcional	Riesgos de control y supervisión	Russell & Norvig (2022); Ng (2018)
Escalas de medición de la agencia	Representación gradual del nivel de organización	Se diferencia de clasificación binaria	Permite comparar niveles de agencia	Dificultad de estandarización	Sapkota et al. (2026); Bandi et al. (2025)

Concepto	Definición	Diferenciación	Ventajas	Limitaciones	Referencias
	del comportamiento				
Estabilidad del comportamiento	Capacidad de sostener coherencia en contextos dinámicos	Se diferencia de rigidez estructural	Permite validar consistencia del sistema	Sensible a cambios del entorno	Vinuesa et al. (2020); Russell (2019)
Evidencia empírica de la agencia	Observación de patrones consistentes en entornos reales	Se diferencia de validación teórica	Permite confirmar existencia de agencia	Dependencia del contexto	Wang (2025); OECD (2026)
Criterios de medición	Dimensiones utilizadas para evaluar la agencia (coherencia, continuidad, autonomía)	Se diferencia de métricas tradicionales de desempeño	Permite evaluación estructural del sistema	Complejidad metodológica	Bandi et al. (2025); Guidotti et al. (2018)
Operacionalización de la medición	Traducción de conceptos abstractos en indicadores observables	Se diferencia de medición directa	Permite análisis aplicado	Riesgo de simplificación excesiva	Adadi & Berrada (2018); Sapkota et al. (2026)

Fuente: Recopilación y elaboración propia

CAPÍTULO 6. Impacto, gobernanza y futuro de la IA agéntica

El impacto de la inteligencia artificial, particularmente en su dimensión agéntica, se configura como una **transformación estructural de los sistemas sociales, económicos y organizacionales**, donde convergen oportunidades de desarrollo y riesgos sistémicos. En el ámbito del desarrollo sostenible, la IA presenta un doble efecto: por un lado, **impulsa avances significativos en productividad, innovación y eficiencia**; por otro, introduce desafíos que requieren **supervisión, regulación y control estratégico** para evitar efectos adversos.

Desde la perspectiva social, la inteligencia artificial redefine procesos educativos, culturales y de inclusión, consolidándose como un **factor clave en la transformación del conocimiento y la interacción humana**. No obstante, su expansión plantea retos relacionados con la equidad, el acceso y la distribución de beneficios, lo que hace indispensable el diseño de **políticas públicas orientadas a la inclusión digital y la reducción de brechas tecnológicas**. En el plano económico, la IA actúa como una **tecnología de propósito general que reconfigura el mercado laboral**, generando tensiones entre automatización y creación de empleo, así como nuevas formas de desigualdad si no se orienta hacia modelos de desarrollo equilibrado

En el ámbito organizacional, la adopción de IA implica **cambios profundos en estructuras, procesos y capacidades**, requiriendo inversiones en capital humano, rediseño organizacional y adaptación estratégica para capturar su valor. Estas transformaciones generan **nuevas dinámicas competitivas y modelos de gestión**, donde la relación entre tecnología y trabajo se redefine constantemente. La gobernanza emerge como el eje articulador de estos procesos, destacando la necesidad de **marcos globales que regulen, coordinen y orienten el desarrollo de la IA**, garantizando transparencia, responsabilidad y confianza. Finalmente, la prospectiva de la IA apunta hacia sistemas cada vez más autónomos, cuya integración dependerá de la **capacidad de alinear innovación tecnológica con valores humanos, sostenibilidad y cooperación global**, consolidando así su papel en el futuro de las sociedades.

Impacto social de la IA agéntica

La **IA agéntica** constituye un **factor de transformación estructural de la sociedad**, al integrar sistemas capaces de percibir, decidir y actuar dentro de entornos sociales complejos. Su impacto se manifiesta en la reconfiguración de las dinámicas de interacción, donde emerge una **sociedad híbrida humano–máquina** en la que los agentes inteligentes participan activamente en procesos de comunicación y toma de decisiones. Asimismo, influye en la redistribución de roles y funciones sociales, modificando la organización del trabajo y la participación colectiva.

En el ámbito cultural y educativo, la IA redefine la **producción, acceso y uso del conocimiento**, favoreciendo procesos más dinámicos y personalizados, pero también generando riesgos de exclusión si su acceso es desigual. Desde una perspectiva ética, introduce desafíos relacionados con la **responsabilidad, la equidad y la transparencia**, lo que exige marcos de gobernanza adecuados. En este sentido, la IA agéntica se configura como un **elemento central en la evolución de la sociedad contemporánea**, cuya gestión determinará su impacto en el desarrollo social.

Transformación de la sociedad y las dinámicas sociales

La **IA agéntica** está redefiniendo la estructura de la sociedad contemporánea al introducir sistemas capaces de **percibir, decidir y actuar de manera autónoma dentro de entornos sociales complejos**. A diferencia de tecnologías anteriores, estos sistemas no se limitan a ejecutar tareas, sino que participan activamente en procesos sociales, influyendo en la organización de actividades, la toma de decisiones y la interacción entre individuos. En este sentido, la IA se configura como un **actor sociotécnico**, cuya presencia transforma las dinámicas sociales al integrarse en múltiples niveles de la vida cotidiana (Vinuesa et al., 2020).

Desde la perspectiva de la interacción social, la IA agéntica modifica la forma en que los individuos se comunican, colaboran y coordinan sus acciones. La mediación tecnológica deja de ser pasiva para convertirse en un proceso activo en el que los

agentes inteligentes intervienen en la construcción de las relaciones sociales. Este fenómeno implica la emergencia de una **sociedad híbrida**, en la que las interacciones ya no son exclusivamente humanas, sino que incorporan sistemas inteligentes que influyen en los procesos comunicativos y decisionales (UNESCO, 2025ª).

En el ámbito de la organización social, la IA agéntica contribuye a la redefinición de roles, funciones y estructuras dentro de la sociedad. La automatización de actividades y la delegación de decisiones a sistemas inteligentes generan una redistribución de responsabilidades que impacta tanto en el ámbito laboral como en la vida cotidiana. En este contexto, la IA se configura como un **mecanismo de reestructuración social**, donde las funciones tradicionales son transformadas por nuevas capacidades tecnológicas que alteran la lógica de participación de los individuos (United Nations, 2024).

Desde la perspectiva de la inclusión social, la IA agéntica presenta un carácter dual. Por un lado, facilita el acceso a servicios, información y oportunidades, lo que puede contribuir a reducir desigualdades estructurales. Por otro, su distribución desigual puede profundizar brechas existentes, generando nuevas formas de exclusión. En este sentido, la IA se configura como un **determinante crítico de inclusión o exclusión social**, cuya influencia depende de factores como el acceso tecnológico, las capacidades digitales y las políticas de implementación (UNESCO, 2025b).

En el plano de la toma de decisiones, la IA agéntica introduce una mediación estructural en los procesos sociales, ya que los sistemas inteligentes pueden influir en decisiones individuales y colectivas mediante recomendaciones, automatización de procesos y análisis de datos. Este fenómeno implica una transformación en la autonomía humana, donde las decisiones se construyen en interacción con sistemas tecnológicos. La IA se configura así como un **agente mediador en la toma de decisiones sociales**, alterando la forma en que los individuos y las organizaciones procesan la información y eligen cursos de acción (Stahl, 2021).

Desde una perspectiva sistémica, la IA agéntica opera dentro de redes complejas donde interactúan múltiples actores humanos y tecnológicos. Estas interacciones generan dinámicas emergentes que no pueden explicarse únicamente a partir de los componentes individuales del sistema. La sociedad se configura así como un **sistema sociotécnico complejo**, donde la IA actúa como un elemento estructurante que influye en la estabilidad, la adaptación y la evolución de las dinámicas sociales (Vinuesa et al., 2020).

En el ámbito de la confianza social, la integración de la IA agéntica plantea desafíos relacionados con la percepción y aceptación de estos sistemas por parte de la sociedad. La confianza en la tecnología depende de factores como la transparencia, la confiabilidad y la percepción de beneficio. Este proceso implica que la IA no solo debe ser funcionalmente eficiente, sino también socialmente aceptada. En este sentido, la IA se configura como un **objeto de confianza social**, cuya legitimidad depende de su alineación con expectativas y valores colectivos (United Nations, 2024).

Desde la perspectiva de la transformación estructural, la IA agéntica contribuye a redefinir las bases sobre las cuales se organiza la sociedad, incluyendo la producción, la distribución y el acceso a recursos. Este proceso implica una reconfiguración de las relaciones de poder, donde el acceso a la tecnología se convierte en un factor determinante. La IA se configura así como un **elemento de reconfiguración estructural**, donde las dinámicas sociales se transforman en función de nuevas capacidades tecnológicas (UNESCO, 2025ª).

La IA agéntica consolida una nueva etapa en la evolución de la sociedad, caracterizada por la integración profunda entre sistemas humanos y tecnológicos. Esta integración implica que los agentes inteligentes no solo apoyan la actividad humana, sino que participan activamente en la construcción de la realidad social. En este sentido, la IA se configura como un **factor estructural de cambio social**, cuya influencia redefine las dinámicas de interacción, organización y decisión en la sociedad contemporánea (Stahl, 2021).

Cultura, educación y producción del conocimiento

La **IA agéntica** está transformando de manera profunda los sistemas culturales y educativos al modificar la forma en que el conocimiento se produce, distribuye y utiliza en la sociedad. Este proceso implica una reconfiguración de los mecanismos tradicionales de aprendizaje y transmisión cultural, donde los sistemas inteligentes pasan de ser herramientas de apoyo a convertirse en participantes activos. En este sentido, la IA se configura como un **agente de transformación cultural**, capaz de influir en las dinámicas de creación y difusión del conocimiento (UNESCO, 2025ª).

Desde la perspectiva educativa, la IA agéntica introduce modelos de aprendizaje que se caracterizan por su capacidad de adaptación a las necesidades individuales de los estudiantes. A través de sistemas inteligentes, es posible ajustar contenidos, ritmos y estrategias pedagógicas en función de las características de cada usuario. Este fenómeno implica una ruptura con los modelos estandarizados tradicionales, consolidando a la IA como un **facilitador del aprendizaje personalizado**, donde el conocimiento se construye de manera dinámica (UNESCO, 2025b).

En el plano de la producción del conocimiento, la IA agéntica participa activamente en la generación de contenidos, lo que redefine la relación entre autoría, creatividad y tecnología. Este proceso implica que el conocimiento deja de ser exclusivamente humano para convertirse en un producto híbrido, donde interactúan capacidades humanas y sistemas inteligentes. La IA se configura así como un **co-productor de conocimiento**, ampliando las posibilidades de creación y análisis (Vinuesa et al., 2020).

Desde la perspectiva cultural, la IA agéntica influye en la construcción de significados, valores y prácticas sociales al intervenir en la producción y circulación de contenidos. Este fenómeno implica que la cultura ya no se desarrolla únicamente a través de procesos humanos, sino que incorpora sistemas tecnológicos que influyen

en su configuración. La IA se configura así como un **agente de mediación cultural**, capaz de transformar las formas en que se construye la realidad simbólica (Stahl, 2021).

En el ámbito del acceso al conocimiento, la IA agéntica facilita la disponibilidad de información a gran escala, lo que puede contribuir a democratizar el aprendizaje y ampliar las oportunidades educativas. Sin embargo, este proceso también puede generar nuevas formas de exclusión si el acceso a estas tecnologías no es equitativo. La IA se configura así como un **factor de democratización o exclusión del conocimiento**, dependiendo de su distribución social (United Nations, 2024).

Desde la perspectiva de la innovación educativa, la IA agéntica impulsa el desarrollo de nuevas metodologías de enseñanza que integran tecnología y pedagogía, generando entornos de aprendizaje más flexibles y adaptativos. Este proceso implica una transformación en el rol de docentes y estudiantes, donde la interacción con sistemas inteligentes se vuelve central. La IA se configura así como un **motor de innovación educativa**, redefiniendo las prácticas pedagógicas (UNESCO, 2025ª).

En el plano de la calidad del conocimiento, la IA agéntica permite analizar grandes volúmenes de información, facilitando la identificación de patrones y la generación de nuevos insights. Este proceso mejora la capacidad de producir conocimiento relevante y actualizado. La IA se configura así como un **potenciador de la calidad del conocimiento**, donde el análisis se amplifica mediante capacidades computacionales (Vinuesa et al., 2020).

Desde la perspectiva de la interacción humano-tecnología, la IA agéntica introduce nuevas formas de aprendizaje basadas en la colaboración entre humanos y sistemas inteligentes. Este proceso implica una redefinición del rol del usuario, que pasa de ser receptor a participante activo en la construcción del conocimiento. La IA se configura así como un **facilitador de aprendizaje colaborativo**, donde la interacción humano-máquina se vuelve central (UNESCO, 2025b).

En el ámbito de la transformación cultural, la IA agéntica contribuye a redefinir las prácticas sociales relacionadas con la producción y consumo de contenidos, generando nuevas formas de expresión y comunicación. Este proceso implica una evolución en la cultura digital, donde la tecnología influye en la forma en que se construyen identidades y narrativas. La IA se configura así como un **elemento de transformación cultural**, que redefine las dinámicas de la sociedad contemporánea (Stahl, 2021).

Finalmente, la integración de la IA agéntica en los sistemas culturales y educativos consolida una nueva etapa en la evolución del conocimiento, caracterizada por la interacción entre humanos y sistemas inteligentes. Este proceso implica que el conocimiento se vuelve más dinámico, distribuido y adaptativo. En este sentido, la IA se configura como un **factor estructural en la evolución del conocimiento**,

redefiniendo su producción, acceso y uso en la sociedad contemporánea (United Nations, 2024).

Ética, responsabilidad y desafíos sociales

La **IA agéntica** introduce un conjunto de desafíos éticos que emergen de su capacidad para actuar de manera autónoma en contextos sociales complejos, influyendo en decisiones que afectan a individuos, organizaciones y sociedades enteras. Este fenómeno implica que la tecnología deja de ser neutral para convertirse en un elemento con implicaciones morales, lo que exige una reflexión profunda sobre sus usos y consecuencias. En este sentido, la IA se configura como un **problema ético estructural**, donde sus impactos trascienden el ámbito técnico y se insertan en el tejido social (Stahl, 2021).

Desde la perspectiva de la responsabilidad, uno de los principales desafíos de la IA agéntica radica en determinar quién es responsable de las decisiones tomadas por sistemas autónomos. Este problema cuestiona los marcos tradicionales de responsabilidad, ya que las acciones del sistema no siempre pueden atribuirse directamente a un actor humano específico. La IA se configura así como un **desafío para la atribución de responsabilidad**, donde se requiere redefinir los mecanismos de rendición de cuentas (Floridi et al., 2021).

En el plano de la equidad, la IA agéntica puede generar impactos diferenciados en distintos grupos sociales, dependiendo de factores como el acceso, el diseño del sistema y los datos utilizados. Este fenómeno implica que la tecnología puede reproducir o amplificar desigualdades existentes si no se implementan mecanismos de control adecuados. La IA se configura así como un **factor potencial de desigualdad social**, donde su impacto depende de las condiciones de desarrollo y aplicación (Vinuesa et al., 2020).

Desde la perspectiva de la transparencia, la IA agéntica plantea la necesidad de comprender cómo se generan las decisiones dentro del sistema. La opacidad de los modelos puede dificultar la interpretación de sus resultados, lo que afecta la confianza social en la tecnología. En este sentido, la IA se configura como un **sistema que requiere explicabilidad**, donde la transparencia se convierte en un elemento fundamental para su aceptación (United Nations, 2024).

En el ámbito de la autonomía humana, la IA agéntica introduce tensiones relacionadas con la capacidad de los individuos para tomar decisiones independientes. La intervención de sistemas inteligentes en procesos decisionales puede influir en las elecciones humanas, lo que plantea interrogantes sobre la pérdida de control. La IA se configura así como un **factor de mediación de la autonomía**, donde la relación entre humanos y tecnología redefine los límites de la decisión (Stahl, 2021).

Desde la perspectiva de la regulación ética, la IA agéntica requiere la construcción de marcos normativos que orienten su desarrollo y uso en función de principios éticos.

Estos marcos permiten establecer límites y condiciones para la operación del sistema, garantizando que su impacto sea socialmente aceptable. La IA se configura así como un **objeto de regulación ética**, donde la gobernanza se convierte en un elemento clave para su implementación (UNESCO, 2025b).

En el plano de la confianza social, la aceptación de la IA agéntica depende de la percepción de que sus decisiones son justas, transparentes y responsables. La falta de confianza puede limitar su adopción, incluso si la tecnología es funcionalmente eficiente. La IA se configura así como un **objeto de legitimidad social**, donde la confianza se convierte en un requisito para su integración en la sociedad (United Nations, 2024).

Desde una perspectiva sistémica, los desafíos éticos de la IA agéntica no pueden analizarse de manera aislada, ya que están interconectados con factores sociales, económicos y tecnológicos. Este enfoque implica considerar la ética como un componente integrado en el sistema. La IA se configura así como un **sistema éticamente interdependiente**, donde las decisiones tecnológicas tienen implicaciones amplias (Vinuesa et al., 2020).

En el ámbito de los riesgos sociales, la IA agéntica puede generar efectos no previstos que afectan la estabilidad de los sistemas sociales, particularmente en contextos donde la tecnología se adopta sin una evaluación adecuada de sus consecuencias. Este fenómeno implica que la innovación tecnológica debe ser acompañada de mecanismos de control. La IA se configura así como un **generador de riesgos sociales**, donde la gestión ética es fundamental (Floridi et al., 2021). La integración de la IA agéntica en la sociedad plantea la necesidad de equilibrar innovación y responsabilidad, asegurando que el desarrollo tecnológico se alinee con valores humanos fundamentales. Este proceso implica construir un marco ético que permita orientar el uso de la tecnología hacia el bienestar social. En este sentido, la IA se configura como un **campo de tensión entre progreso tecnológico y responsabilidad social**, cuya gestión determinará su impacto en el futuro (Stahl, 2021).

Impacto económico

La IA agéntica redefine la productividad y la eficiencia económica al integrar capacidades autónomas de percepción, decisión y acción en los sistemas productivos. A diferencia de tecnologías previas, no solo optimiza tareas aisladas, sino que transforma cadenas completas de valor, modificando la lógica de producción. **La productividad deja de depender exclusivamente del aumento de recursos y se orienta hacia su uso inteligente y optimizado**, lo que incrementa la eficiencia sistémica. En este sentido, la IA permite reducir costos operativos, mejorar la asignación de recursos y acelerar procesos de innovación mediante análisis avanzado de datos. Asimismo, facilita la escalabilidad organizacional al permitir el crecimiento sin incrementos proporcionales en costos. **La eficiencia se vuelve dinámica,**

adaptándose continuamente a cambios del entorno, lo que fortalece la resiliencia económica. En conjunto, la IA agéntica se consolida como un **motor estructural de transformación económica**, redefiniendo la productividad en términos de autonomía, adaptabilidad y generación de valor

Productividad y eficiencia económica

La **IA agéntica** se configura como una **tecnología de propósito general** que transforma la productividad al integrar capacidades autónomas de percepción, decisión y acción dentro de los sistemas económicos. A diferencia de innovaciones previas, estos sistemas no solo optimizan tareas aisladas, sino que intervienen en cadenas completas de valor, modificando la lógica de producción y operación. En este sentido, la IA se posiciona como un **motor estructural de la productividad**, redefiniendo la eficiencia en múltiples sectores económicos (Brynjolfsson et al., 2021).

Desde la perspectiva de la eficiencia operativa, la IA agéntica permite reducir costos mediante la automatización de procesos complejos que anteriormente requerían intervención humana intensiva. Este proceso no se limita a tareas rutinarias, sino que abarca actividades analíticas y decisionales, ampliando significativamente el alcance de la automatización. La IA se configura así como un **mecanismo de optimización económica**, donde la eficiencia se convierte en un factor determinante en la competitividad de las organizaciones (Acemoglu & Restrepo, 2019).

En el plano de la productividad total de los factores, la IA agéntica contribuye a mejorar la relación entre insumos y resultados al optimizar el uso de recursos disponibles. Este fenómeno implica que la generación de valor ya no depende exclusivamente del incremento de recursos, sino de su utilización eficiente mediante sistemas inteligentes. La IA se configura así como un **optimizador de la productividad sistémica**, donde la eficiencia se extiende a toda la estructura económica (OECD, 2025b).

Desde la perspectiva de la innovación productiva, la IA agéntica facilita la exploración de múltiples soluciones mediante el análisis de grandes volúmenes de datos y la simulación de escenarios. Este proceso permite reducir los tiempos de desarrollo y aumentar la precisión en la toma de decisiones productivas. La IA se configura así como un **acelerador de la innovación económica**, donde la generación de valor se ve impulsada por capacidades analíticas avanzadas (Brynjolfsson et al., 2021).

En el ámbito de la asignación de recursos, la IA agéntica permite optimizar la distribución de insumos dentro de los sistemas productivos, reduciendo desperdicios y mejorando la eficiencia en la gestión de activos. Este proceso implica una transformación en la forma en que las organizaciones gestionan sus recursos, favoreciendo decisiones más informadas. La IA se configura así como un **optimizador de recursos económicos**, donde la eficiencia se traduce en mayor productividad (World Bank, 2026).

Desde la perspectiva de la escalabilidad, la IA agéntica permite replicar procesos productivos con alta eficiencia, facilitando el crecimiento de las organizaciones sin un incremento proporcional de costos. Este fenómeno implica que las empresas pueden expandirse de manera más rápida y eficiente en mercados globales. La IA se configura así como un **factor de escalabilidad económica**, donde el crecimiento se desacopla de las limitaciones tradicionales de recursos humanos y operativos (Frank et al., 2025).

En el plano de la eficiencia dinámica, la IA agéntica introduce la capacidad de adaptación continua en los sistemas productivos, permitiendo ajustar procesos en función de cambios en el entorno. Este proceso implica una mejora en la resiliencia económica, donde las organizaciones pueden responder de manera más efectiva a condiciones cambiantes. La IA se configura así como un **mecanismo de adaptación económica**, donde la eficiencia se mantiene en contextos dinámicos (OECD, 2025b).

Desde una perspectiva sistémica, la IA agéntica redefine las relaciones entre producción, distribución y consumo, generando nuevas dinámicas económicas que afectan la estructura de los mercados. Este fenómeno implica una reorganización de los sistemas económicos en función de capacidades tecnológicas avanzadas. La IA se configura así como un **elemento de reconfiguración económica**, donde la productividad se integra dentro de una lógica sistémica (Acemoglu & Restrepo, 2019).

En el ámbito del valor económico, la IA agéntica permite generar nuevas fuentes de valor mediante la automatización de decisiones y la optimización de procesos, lo que incrementa la capacidad de las organizaciones para competir en entornos complejos. Este proceso implica una transformación en la forma en que se produce y captura valor. La IA se configura así como un **generador de valor económico**, donde la eficiencia y la innovación se combinan para impulsar el crecimiento (Brynjolfsson et al., 2021).

La IA agéntica consolida una nueva lógica productiva basada en la autonomía, la eficiencia y la adaptabilidad, donde los sistemas inteligentes desempeñan un papel central en la generación de valor económico. En este sentido, la IA se configura como un **factor estructural de transformación económica**, cuya influencia redefine la productividad en la economía contemporánea y establece nuevas bases para el crecimiento sostenible (World Bank, 2026).

Empleo, trabajo y automatización

La **IA agéntica** está transformando el mercado laboral al introducir capacidades de automatización que abarcan tanto tareas rutinarias como actividades cognitivas complejas, lo que redefine la naturaleza del trabajo en la economía contemporánea. A diferencia de tecnologías anteriores, estos sistemas no solo ejecutan procesos predefinidos, sino que toman decisiones en entornos dinámicos, ampliando significativamente el alcance de la automatización. En este sentido, la IA se configura como un **factor disruptivo del empleo**, con implicaciones estructurales en la organización del trabajo (Acemoglu & Restrepo, 2019).

Desde la perspectiva de la automatización, la IA agéntica permite sustituir actividades que implican análisis, razonamiento y ejecución, lo que extiende la automatización más allá de las tareas manuales hacia funciones cognitivas. Este fenómeno implica que sectores tradicionalmente protegidos frente a la automatización comienzan a experimentar cambios significativos. La IA se configura así como un **motor de automatización avanzada**, donde la intervención humana se reduce en múltiples niveles de actividad económica (Brynjolfsson et al., 2021).

En el plano de la sustitución laboral, la IA agéntica puede desplazar ciertos tipos de empleo, particularmente aquellos basados en tareas repetitivas o estructuradas. Este proceso implica una reducción en la demanda de ciertos perfiles laborales, lo que genera tensiones en el mercado de trabajo. La IA se configura así como un **factor de desplazamiento laboral**, donde la sustitución de tareas redefine la demanda de trabajo (Acemoglu & Restrepo, 2019).

Desde la perspectiva de la creación de empleo, la IA también genera nuevas oportunidades laborales en áreas relacionadas con el desarrollo, implementación y supervisión de sistemas inteligentes. Este proceso implica una transformación en la estructura ocupacional, donde emergen nuevos perfiles profesionales. La IA se configura así como un **generador de nuevas formas de empleo**, impulsando la demanda de habilidades avanzadas (Frank et al., 2025).

En el ámbito de la transformación del trabajo, la IA agéntica introduce modelos híbridos en los que humanos y sistemas inteligentes colaboran en la ejecución de tareas. Este fenómeno implica una redefinición de roles y responsabilidades, donde el trabajo humano se orienta hacia actividades de mayor valor agregado. La IA se configura así como un **elemento de transformación laboral**, donde la interacción humano–máquina se vuelve central (World Bank, 2026).

Desde la perspectiva de las habilidades, la IA agéntica modifica la demanda de competencias en el mercado laboral, favoreciendo habilidades digitales, analíticas y adaptativas. Este proceso implica que la educación y la capacitación se convierten en elementos clave para la empleabilidad. La IA se configura así como un **factor de reconfiguración de habilidades**, donde la adaptación se vuelve fundamental para la integración laboral (OECD, 2025b).

En el plano de la desigualdad, la automatización impulsada por la IA agéntica puede generar efectos diferenciados entre trabajadores con distintos niveles de habilidades, ampliando brechas en el mercado laboral. Este fenómeno implica que los beneficios de la tecnología no se distribuyen de manera uniforme. La IA se configura así como un **factor potencial de desigualdad laboral**, donde el impacto depende de la capacidad de adaptación de la fuerza de trabajo (Acemoglu & Restrepo, 2019).

Desde la perspectiva de la calidad del empleo, la IA agéntica puede mejorar las condiciones laborales al eliminar tareas peligrosas o repetitivas, permitiendo que los trabajadores se concentren en actividades más complejas y creativas. Este proceso

implica una transformación en la naturaleza del trabajo, donde la calidad se convierte en un elemento central. La IA se configura así como un **mejorador potencial de la calidad del empleo**, donde la automatización libera capacidades humanas (Brynjolfsson et al., 2021).

En el ámbito de la organización del trabajo, la IA agéntica influye en la estructura de las actividades laborales, introduciendo mayor flexibilidad y nuevas formas de coordinación entre trabajadores y sistemas inteligentes. Este fenómeno implica una evolución en las dinámicas laborales. La IA se configura así como un **elemento de reorganización del trabajo**, donde las estructuras laborales se adaptan a nuevas tecnologías (World Bank, 2026).

Desde una perspectiva sistémica, la IA agéntica redefine el equilibrio entre oferta y demanda de trabajo, generando una transición hacia economías basadas en el conocimiento y la tecnología. Este proceso implica cambios estructurales en el mercado laboral. La IA se configura así como un **factor de reestructuración del sistema laboral**, donde la economía evoluciona hacia nuevas formas de empleo (OECD, 2025b).

Finalmente, la IA agéntica introduce una dinámica compleja en el empleo, donde coexisten procesos de sustitución, transformación y creación de trabajo. Este fenómeno implica que el impacto de la tecnología no es unívoco, sino multidimensional. En este sentido, la IA se configura como un **elemento central en la evolución del trabajo**, cuya gestión determinará su impacto en la equidad y el desarrollo económico (Frank et al., 2025).

Mercados, competitividad y estructura económica

La **IA agéntica** está redefiniendo la dinámica de los mercados al introducir capacidades avanzadas de análisis y toma de decisiones autónoma, lo que transforma la forma en que las empresas compiten y generan valor. A diferencia de enfoques tradicionales basados en eficiencia incremental, estos sistemas permiten anticipar tendencias, optimizar estrategias y responder en tiempo real a condiciones cambiantes. En este sentido, la IA se configura como un **factor clave de competitividad**, donde la ventaja se construye a partir de la capacidad tecnológica (Frank et al., 2025).

Desde la perspectiva de la competencia, la IA agéntica favorece a organizaciones que poseen acceso a grandes volúmenes de datos y capacidades computacionales avanzadas, lo que puede generar asimetrías significativas en el mercado. Este fenómeno implica que la competencia ya no se basa únicamente en precios o calidad, sino en la capacidad de procesamiento y aprendizaje. La IA se configura así como un **factor de concentración económica**, donde las ventajas competitivas tienden a acumularse en actores tecnológicos (OECD, 2025b).

En el plano de los modelos de negocio, la IA agéntica permite el desarrollo de nuevas formas de generación de valor basadas en servicios inteligentes, plataformas digitales y automatización de procesos. Este proceso implica una transformación en la lógica empresarial, donde el valor se crea a partir de la información y su procesamiento. La IA se configura así como un **impulsor de nuevos modelos económicos**, donde la innovación redefine las estrategias de mercado (World Bank, 2026).

Desde la perspectiva de la globalización, la IA agéntica facilita la expansión de empresas en mercados internacionales al optimizar la gestión de operaciones, la logística y la toma de decisiones. Este fenómeno reduce las barreras de entrada y permite la participación en mercados globales con mayor eficiencia. La IA se configura así como un **facilitador de la globalización económica**, donde las organizaciones pueden operar a escala internacional con mayor competitividad (Brynjolfsson et al., 2021).

En el ámbito de la estructura de mercado, la IA agéntica contribuye a la transformación de sectores productivos al impulsar la transición hacia economías intensivas en tecnología. Este proceso implica una reorganización de la actividad económica, donde sectores tradicionales se ven desplazados o transformados. La IA se configura así como un **factor de transformación estructural**, donde la economía evoluciona hacia modelos basados en innovación (Acemoglu & Restrepo, 2019).

Desde la perspectiva de la eficiencia del mercado, la IA agéntica mejora la capacidad de las empresas para ajustar precios, anticipar demanda y optimizar la asignación de recursos, lo que puede generar mercados más eficientes. Sin embargo, también puede introducir nuevas dinámicas de competencia que alteran el equilibrio del mercado. La IA se configura así como un **optimizador de la eficiencia de mercado**, donde la información se convierte en un recurso estratégico (OECD, 2025b).

En el plano de la innovación competitiva, la IA agéntica permite a las empresas diferenciarse mediante el desarrollo de capacidades únicas basadas en datos y algoritmos. Este proceso implica una transformación en la forma en que se construyen ventajas competitivas, donde la innovación se convierte en un elemento central. La IA se configura así como un **factor de diferenciación competitiva**, donde la tecnología define la posición en el mercado (Frank et al., 2025). Desde una perspectiva sistémica, la IA agéntica redefine las relaciones entre actores económicos, generando nuevas dinámicas de colaboración y competencia. Este proceso implica la creación de ecosistemas digitales donde múltiples actores interactúan de manera interdependiente. La IA se configura así como un **elemento de reorganización sistémica del mercado**, donde la estructura económica se vuelve más compleja (World Bank, 2026).

En el ámbito de la regulación económica, la IA agéntica plantea desafíos relacionados con la concentración de poder y la competencia justa, lo que requiere el desarrollo de marcos regulatorios adecuados. Este proceso implica la necesidad de equilibrar innovación y control para evitar distorsiones en el mercado. La IA se

configura así como un **desafío para la regulación económica**, donde la gobernanza del mercado se vuelve fundamental (OECD, 2025b). La IA agéntica consolida una transformación profunda en los mercados y en la estructura económica, donde la capacidad de adaptación, innovación y procesamiento de información define el éxito de las organizaciones. En este sentido, la IA se configura como un **elemento estructural del cambio económico**, que redefine la competencia, los modelos de negocio y la organización de la economía en su conjunto (Brynjolfsson et al., 2021).

Mercados, competitividad y estructura económica

La IA agéntica está transformando de manera profunda la relación entre **mercados, competitividad y estructura económica**, al introducir sistemas capaces de actuar de forma autónoma dentro de los procesos productivos y de intercambio. En los mercados, **la interacción ya no depende exclusivamente de decisiones humanas, sino de agentes que operan en tiempo real**, generando dinámicas más rápidas, adaptativas y complejas. Esto implica que los precios, la asignación de recursos y las condiciones de intercambio emergen de procesos continuos de ajuste, lo que redefine la lógica tradicional del equilibrio.

En términos de competitividad, **la ventaja se desplaza hacia la capacidad de desarrollar y coordinar sistemas agénticos**, donde la adaptabilidad, el aprendizaje y la rapidez de respuesta se convierten en factores clave. Las organizaciones ya no compiten solo por eficiencia, sino por su capacidad de integrar inteligencia autónoma en sus procesos. A nivel de estructura económica, **la IA agéntica reconfigura la relación entre capital y trabajo**, favoreciendo a quienes controlan estas tecnologías y generando nuevas formas de concentración del valor. En conjunto, estos cambios configuran una economía más dinámica, interdependiente y basada en inteligencia distribuida.

Reconfiguración de los mercados bajo IA agéntica

La IA agéntica transforma la naturaleza del mercado al introducir un nuevo principio de coordinación basado en la **interacción autónoma entre agentes artificiales capaces de percibir, razonar y actuar en entornos dinámicos**, desplazando el rol tradicional del mercado como mecanismo pasivo de ajuste. En este sentido, el mercado deja de ser únicamente un espacio de intercambio para convertirse en un **sistema activo de decisión distribuida**, donde múltiples agentes generan continuamente estados económicos a partir de sus interacciones (Nisa et al., 2026). En esta configuración, la formación de precios se redefine profundamente. A diferencia de los modelos tradicionales donde los precios emergen de agregaciones relativamente estables, **en mercados agénticos los precios son el resultado de procesos iterativos de ajuste continuo entre agentes autónomos**, lo que implica que el precio deja de ser una señal estática y se convierte en una variable dinámica en permanente actualización (World Economic Forum, 2024).

Este fenómeno introduce una transformación en la lógica de coordinación. Mientras que los mercados clásicos operan mediante mecanismos de equilibrio, **los mercados agénticos funcionan a través de procesos de coordinación adaptativa**, donde los agentes ajustan sus comportamientos en función de la información que perciben del entorno y de otros agentes, generando patrones de interacción que evolucionan en tiempo real (OECD, 2026). Como consecuencia, el concepto de equilibrio pierde centralidad analítica. En lugar de converger hacia estados estables, **los mercados agénticos operan bajo condiciones de desequilibrio permanente**, donde las dinámicas económicas se mantienen en constante cambio debido a la capacidad de los agentes para modificar sus estrategias de manera continua (World Bank, 2026).

Otro elemento fundamental es la transformación de la información dentro del mercado. En los sistemas tradicionales, la información es incompleta y distribuida, pero en los mercados agénticos **la información es procesada activamente por agentes que la interpretan, la transforman y actúan sobre ella**, lo que genera ciclos de retroalimentación que amplifican o atenúan determinadas señales económicas (Vinuesa et al., 2020). Esta capacidad de procesamiento introduce nuevas dinámicas de interacción. Los agentes no solo responden a información existente, sino que **anticipan comportamientos de otros agentes mediante mecanismos de razonamiento predictivo**, lo que convierte al mercado en un entorno donde la anticipación estratégica es parte estructural de su funcionamiento (Nisa et al., 2026).

Asimismo, la temporalidad del mercado se redefine. En lugar de ciclos económicos claramente diferenciados, **los mercados agénticos operan en flujos continuos de microdecisiones**, donde cada interacción contribuye a la configuración del sistema en tiempo real. Esto implica que el mercado ya no evoluciona en etapas discretas, sino como un proceso continuo de actualización (Acemoglu & Restrepo, 2019). Esta continuidad introduce propiedades de alta sensibilidad. **Pequeñas variaciones en el comportamiento de algunos agentes pueden generar efectos amplificados en el sistema**, lo que refleja una dinámica característica de sistemas complejos no lineales, donde las relaciones causa-efecto no son proporcionales ni fácilmente predecibles (United Nations, 2024).

Además, los mercados agénticos presentan propiedades emergentes que no pueden reducirse al comportamiento individual de los agentes. **El sistema genera patrones colectivos que surgen de la interacción entre agentes y no de una planificación centralizada**, lo que implica que el mercado adquiere una dimensión evolutiva que escapa a modelos deterministas (World Economic Forum, 2024). En este contexto, la estructura del mercado deja de ser fija. **Las configuraciones de interacción entre agentes pueden modificarse continuamente**, dando lugar a redes dinámicas donde los vínculos económicos se reconfiguran en función de la adaptación del sistema, lo que introduce una lógica relacional en la organización del mercado (OECD, 2025b).

A diferencia de estructuras centralizadas, **los mercados agénticos distribuyen la toma de decisiones entre múltiples nodos autónomos**, lo que incrementa la

flexibilidad del sistema, pero también dificulta su supervisión y control (Raisch & Krakowski, 2021). Esta descentralización implica que el control del mercado no se ejerce directamente sobre los agentes individuales, sino sobre las condiciones del entorno en el que operan. **La gobernanza se desplaza hacia el diseño de reglas y contextos de interacción**, en lugar de intervenir en decisiones específicas, lo que redefine el papel de las instituciones económicas (United Nations, 2024).

La IA agéntica introduce una nueva ontología del mercado. **El mercado deja de ser un mecanismo pasivo de asignación y se convierte en un sistema cognitivo distribuido**, donde la inteligencia emerge de la interacción entre agentes autónomos que procesan información, toman decisiones y adaptan su comportamiento continuamente (Nisa et al., 2026). La reconfiguración del mercado bajo IA agéntica no consiste en una mejora incremental de sus mecanismos, sino en un cambio estructural en su lógica de funcionamiento. **El mercado se transforma en un sistema complejo, adaptativo y autoorganizado**, donde la coordinación, la información y la dinámica emergente redefinen completamente su naturaleza (World Economic Forum, 2024).

Competitividad basada en capacidades agénticas

La competitividad en la era de la IA agéntica se redefine a partir de la capacidad de construir sistemas que integran **agencia autónoma como recurso estratégico**, desplazando el énfasis desde la eficiencia operativa hacia la **capacidad de acción inteligente en entornos complejos**. En este sentido, la ventaja competitiva ya no se sustenta únicamente en activos tangibles o analíticos, sino en la habilidad de diseñar agentes capaces de percibir, razonar y ejecutar decisiones de manera autónoma (Nisa et al., 2026).

En este marco, la competitividad se articula en torno a la calidad del diseño agéntico. Elementos como la planificación, la memoria, la capacidad de reflexión y el uso de herramientas se convierten en determinantes clave del desempeño. **La superioridad competitiva emerge de arquitecturas agénticas más robustas y adaptativas**, lo que desplaza el foco desde la optimización de procesos hacia la optimización de capacidades cognitivas artificiales (World Economic Forum, 2024). La IA agéntica introduce una ventaja basada en adaptabilidad. A diferencia de sistemas rígidos, **los agentes pueden modificar su comportamiento en función del entorno**, lo que permite responder de manera dinámica a condiciones cambiantes. Esta capacidad convierte la adaptabilidad en un recurso competitivo central, superando enfoques tradicionales basados en estabilidad o eficiencia estática (Jöhnk et al., 2021). Los sistemas agénticos no solo ejecutan tareas, sino que **aprenden de la interacción y mejoran su desempeño de manera progresiva**, lo que genera ventajas acumulativas difíciles de replicar. Esta dinámica introduce una forma de competitividad basada en trayectorias de aprendizaje, donde el tiempo y la experiencia fortalecen la posición relativa del agente (Brynjolfsson et al., 2021). la competitividad también se vincula con la capacidad de orquestación. Las organizaciones no compiten únicamente a través de agentes individuales, sino mediante **sistemas multiagente que coordinan**

múltiples unidades autónomas, lo que permite abordar problemas complejos mediante la interacción estructurada de capacidades distribuidas (Nisa et al., 2026).

Esta orquestación introduce una dimensión sistémica de la competitividad. **La ventaja no reside en un agente aislado, sino en la coherencia y eficiencia del sistema agéntico en su conjunto**, lo que implica que el diseño de interacciones entre agentes se convierte en un factor crítico de desempeño (World Economic Forum, 2024). Los agentes pueden modelar escenarios, prever comportamientos y ajustar sus decisiones antes de que ocurran eventos. **La competitividad se desplaza hacia la capacidad de anticipar dinámicas futuras**, lo que reduce la incertidumbre y permite posicionamientos estratégicos más eficientes (Cockburn et al., 2018).

Además, la IA agéntica redefine la relación entre velocidad y decisión. La capacidad de operar en tiempo real permite que **la rapidez en la ejecución de decisiones se convierta en un componente esencial de la ventaja competitiva**, lo que transforma la lógica tradicional donde la calidad de la decisión era el principal criterio (Acemoglu & Restrepo, 2019). **La complementariedad humano-agente se convierte en un elemento estratégico**, donde los humanos aportan juicio, creatividad y supervisión, mientras que los agentes aportan velocidad, precisión y capacidad de procesamiento (Parker & Grote, 2022). En este sentido, la competitividad también implica capacidad de control. La autonomía de los agentes introduce riesgos asociados a comportamientos no previstos, por lo que **la capacidad de supervisar, corregir y alinear agentes autónomos se convierte en un recurso competitivo clave**, especialmente en entornos complejos (United Nations, 2024).

Otro componente relevante es la escalabilidad de la inteligencia. A diferencia de recursos físicos, **los sistemas agénticos pueden replicarse y desplegarse en múltiples contextos**, lo que permite expandir capacidades sin incrementos proporcionales en costos. Esta característica redefine la lógica de crecimiento competitivo (World Bank, 2026). La competitividad se vincula con la capacidad de innovación endógena. Los agentes pueden explorar soluciones, generar alternativas y optimizar procesos de manera autónoma, lo que implica que **la innovación se integra dentro del propio sistema agéntico**, en lugar de depender exclusivamente de intervenciones externas (Cockburn et al., 2018). En términos estratégicos, la competitividad también depende del acceso a ecosistemas tecnológicos. **La disponibilidad de infraestructuras, datos y plataformas condiciona la capacidad de desarrollar y desplegar agentes**, lo que introduce una dimensión estructural en la construcción de ventajas competitivas (World Bank, 2026).

La IA agéntica redefine la naturaleza misma de la ventaja competitiva. Esta deja de ser un estado estático y se convierte en un proceso dinámico basado en la capacidad de adaptación, aprendizaje y coordinación. **La competitividad se configura como una propiedad emergente de sistemas agénticos complejos**, donde la interacción entre capacidades determina el desempeño (Raisch & Krakowski, 2021). La competitividad en la IA agéntica no se limita a mejorar procesos existentes, sino que implica una transformación en los fundamentos de la ventaja estratégica. **Las**

organizaciones competitivas son aquellas capaces de diseñar, orquestar y gobernar sistemas de agentes autónomos, consolidando una nueva lógica basada en inteligencia distribuida y adaptativa (World Economic Forum, 2024).

Transformación de la estructura económica en sistemas agénticos

La IA agéntica introduce una transformación profunda en la estructura económica al incorporar la **autonomía decisional como factor productivo**, lo que redefine la composición y funcionamiento del sistema económico. En este contexto, la economía deja de organizarse únicamente en torno a capital y trabajo, para integrar **sistemas de agentes autónomos como unidades activas de generación de valor**, alterando las bases tradicionales de la producción (Nisa et al., 2026). Este cambio implica una reconfiguración de la función de producción. La incorporación de agentes autónomos permite que **la toma de decisiones y la ejecución de tareas complejas se integren dentro del propio sistema tecnológico**, reduciendo la dependencia de intervención humana directa y transformando la relación entre insumos y resultados económicos (World Bank, 2026).

En este marco, la relación entre capital y trabajo se redefine. Los sistemas agénticos, al ser propiedad de actores económicos, se integran dentro del capital, lo que implica que **una parte creciente del valor generado se vincula al control de arquitecturas tecnológicas autónomas**, modificando la distribución funcional del ingreso (Acemoglu & Restrepo, 2019). Como consecuencia, se observa una tendencia hacia la concentración del valor. **Los actores que controlan sistemas agénticos capturan una proporción significativa del valor económico**, ya que estos sistemas permiten escalar la producción y la toma de decisiones sin incrementos proporcionales en costos, lo que amplifica las ventajas acumulativas (Brynjolfsson et al., 2021). La estructura económica adquiere una mayor complejidad sistémica. La interacción entre agentes autónomos genera **configuraciones dinámicas donde las relaciones económicas evolucionan continuamente**, lo que implica que la economía se comporta como un sistema complejo adaptativo en constante transformación (Vinuesa et al., 2020). Este carácter adaptativo introduce una nueva lógica de crecimiento. En lugar de depender exclusivamente de acumulación de capital físico, **el crecimiento económico se vincula a la capacidad de integrar, expandir y mejorar sistemas agénticos**, lo que convierte la inteligencia autónoma en un motor central del desarrollo (World Bank, 2026).

En términos sectoriales, la IA agéntica tiende a reorganizar la economía hacia actividades intensivas en conocimiento. **Los sectores que integran sistemas autónomos adquieren mayor relevancia estructural**, mientras que aquellos basados en tareas rutinarias tienden a transformarse o reducir su participación relativa (Cockburn et al., 2018). Desde la perspectiva del trabajo, la IA agéntica no implica únicamente sustitución, sino reconfiguración. **El trabajo humano se desplaza hacia funciones de supervisión, diseño y control de sistemas autónomos**, lo que

redefine la naturaleza de la participación laboral dentro de la estructura económica (Miller & Davenport, 2021). Sin embargo, esta transformación también introduce tensiones distributivas. La concentración del control sobre sistemas agénticos puede generar **incrementos en la desigualdad económica**, especialmente si el acceso a estas tecnologías se encuentra limitado a ciertos actores o regiones (Acemoglu & Restrepo, 2019). En este contexto, la estructura económica también se ve influida por factores institucionales. **La capacidad de los sistemas económicos para integrar la IA agéntica depende de la existencia de marcos regulatorios, infraestructura digital y capital humano adecuado**, lo que introduce diferencias significativas entre economías (United Nations, 2024).

Además, la IA agéntica tiene implicaciones para el desarrollo sostenible. Por un lado, puede mejorar la eficiencia en el uso de recursos y contribuir a la resolución de problemas complejos; por otro, **puede generar externalidades negativas si su implementación no se alinea con objetivos sociales y ambientales**, lo que requiere una gestión estructural adecuada (Vinuesa et al., 2020). La IA agéntica permite que **los procesos de generación de conocimiento se integren dentro de sistemas autónomos**, lo que acelera la producción de innovaciones y redefine el papel de la investigación en la economía (Cockburn et al., 2018). En términos de gobernanza, la estructura económica basada en IA agéntica exige nuevas formas de regulación. **La supervisión de sistemas autónomos y la gestión de sus impactos se convierten en elementos estructurales del sistema económico**, lo que redefine el papel del Estado y de las instituciones (United Nations, 2024).

La economía agéntica se caracteriza por una mayor interdependencia entre sus componentes. **La interacción entre agentes autónomos genera redes económicas altamente conectadas**, donde las decisiones en un punto del sistema pueden tener efectos amplificados en otros, incrementando la complejidad del sistema (World Economic Forum, 2024). La IA agéntica no solo modifica procesos económicos, sino que redefine la estructura misma de la economía. **La integración de sistemas autónomos como factor productivo transforma la distribución del valor, la dinámica del crecimiento y la organización del sistema económico**, configurando una nueva etapa en la evolución de las estructuras económicas contemporáneas (World Bank, 2026).

Transformación organizacional

La **transformación organizacional en la era de la IA agéntica** implica un cambio profundo en la forma en que las organizaciones operan, se estructuran y toman decisiones. En este contexto, **las organizaciones evolucionan de modelos jerárquicos rígidos hacia sistemas dinámicos y distribuidos**, donde la toma de decisiones se descentraliza y se comparte entre humanos y sistemas autónomos. Esto permite mayor rapidez, adaptabilidad y capacidad de respuesta ante entornos

cambiantes. Asimismo, **los procesos organizacionales dejan de ser lineales y se vuelven adaptativos**, ajustándose en tiempo real mediante la interacción continua entre múltiples agentes. Esta dinámica exige el desarrollo de nuevas capacidades relacionadas con el diseño, supervisión y coordinación de sistemas autónomos. Además, **los roles humanos se redefinen hacia funciones de mayor valor cognitivo**, como la supervisión, el control estratégico y la innovación, mientras que los agentes ejecutan tareas operativas y analíticas. En conjunto, **la organización se transforma en un sistema inteligente, flexible y en constante evolución**, capaz de adaptarse a la complejidad del entorno contemporáneo.

Reconfiguración organizacional bajo IA agéntica

La incorporación de la IA agéntica en las organizaciones implica una transformación estructural que trasciende la automatización tradicional, al introducir sistemas capaces de actuar con autonomía dentro de los procesos organizacionales. En este sentido, **las organizaciones dejan de ser estructuras centradas exclusivamente en la toma de decisiones humana y evolucionan hacia sistemas híbridos donde agentes autónomos participan activamente en la ejecución y coordinación de actividades**, redefiniendo la lógica interna del funcionamiento organizacional (Nisa et al., 2026).

Este cambio supone una redefinición del concepto mismo de organización. Mientras que en los modelos clásicos la estructura se basaba en jerarquías, funciones y roles definidos, la IA agéntica impulsa una transición hacia configuraciones más flexibles. **La organización se transforma en un sistema dinámico donde la toma de decisiones se distribuye entre múltiples unidades, tanto humanas como artificiales**, lo que reduce la dependencia de estructuras rígidas y favorece la adaptabilidad (Jöhnk et al., 2021). Uno de los elementos centrales de esta transformación es la emergencia de estructuras basadas en sistemas multiagente. En este modelo, los agentes interactúan entre sí para cumplir objetivos organizacionales, lo que permite abordar tareas complejas de manera distribuida. **La organización se convierte en un ecosistema de agentes interconectados que coordinan sus acciones de forma autónoma y adaptativa**, incrementando la eficiencia y la capacidad de respuesta ante cambios del entorno (World Economic Forum, 2024).

Asimismo, la IA agéntica redefine los procesos organizacionales. En lugar de flujos de trabajo lineales, predefinidos y controlados centralmente, los procesos se vuelven dinámicos y adaptativos. **Las actividades organizacionales se ajustan en tiempo real en función de la información disponible y de las decisiones de los agentes**, lo que implica una transición hacia modelos operativos más flexibles y resilientes (World Bank, 2026). Otro aspecto clave es la transformación de la toma de decisiones. Tradicionalmente, las decisiones estratégicas se concentraban en niveles jerárquicos superiores, mientras que las operativas se distribuían en niveles inferiores. Sin embargo, con la IA agéntica, **la toma de decisiones se descentraliza y se distribuye a lo largo de toda la organización**, permitiendo respuestas más rápidas y precisas ante condiciones cambiantes (Raisch & Krakowski, 2021). Esta descentralización

introduce una nueva lógica de gobernanza organizacional. En lugar de controlar directamente cada decisión, la organización debe establecer marcos que guíen el comportamiento de los agentes. **La gobernanza se orienta hacia la definición de reglas, objetivos y límites dentro de los cuales los agentes operan de manera autónoma**, lo que implica un cambio profundo en la forma de ejercer el control organizacional (United Nations, 2024).

Además, la integración de IA agéntica transforma la naturaleza de la coordinación organizacional. Mientras que en modelos tradicionales la coordinación se lograba mediante supervisión jerárquica o procedimientos estandarizados, **en entornos agénticos la coordinación emerge de la interacción continua entre agentes**, lo que reduce la necesidad de intervención directa y aumenta la eficiencia del sistema (World Economic Forum, 2024). Sin embargo, esta autonomía también plantea desafíos significativos. La capacidad de los agentes para actuar de manera independiente implica riesgos asociados a decisiones no previstas o comportamientos emergentes. En este contexto, **las organizaciones deben desarrollar mecanismos de supervisión y alineación que permitan garantizar que las acciones de los agentes sean coherentes con los objetivos organizacionales**, evitando desviaciones críticas (United Nations, 2024). La integración de agentes autónomos requiere cambios en la percepción del trabajo, la autoridad y la colaboración. **La cultura organizacional debe evolucionar hacia modelos que integren confianza en sistemas autónomos, aprendizaje continuo y adaptación constante**, lo que implica una redefinición de los valores y prácticas organizacionales (Parker & Grote, 2022).

Asimismo, la IA agéntica redefine la relación entre estructura y entorno. Las organizaciones ya no operan como sistemas cerrados con límites claramente definidos, sino como entidades abiertas que interactúan constantemente con su entorno. **La capacidad de los agentes para procesar información externa y adaptarse a ella convierte a la organización en un sistema altamente permeable y sensible al contexto**, lo que incrementa su capacidad de respuesta (World Bank, 2026). Se debe considerar que esta transformación implica una redefinición de la organización como sistema. **La organización deja de ser una estructura estática para convertirse en un sistema dinámico, adaptativo y en constante evolución**, donde la interacción entre agentes humanos y artificiales configura su funcionamiento y determina su capacidad de adaptación (Nisa et al., 2026). La IA agéntica redefine la organización en múltiples niveles, desde su estructura hasta sus procesos y cultura. **La organización se configura como un ecosistema de agentes autónomos interconectados, capaz de adaptarse continuamente a entornos complejos**, marcando una nueva etapa en la evolución de las formas organizacionales contemporáneas (World Economic Forum, 2024).

Capacidades organizacionales en entornos agénticos

La transformación organizacional impulsada por la IA agéntica se fundamenta en el desarrollo de nuevas capacidades que permiten a las organizaciones operar en

entornos caracterizados por autonomía, complejidad y cambio continuo. En este contexto, **las capacidades organizacionales dejan de centrarse exclusivamente en recursos físicos o procesos estandarizados y se orientan hacia la gestión de sistemas autónomos y la integración de inteligencia distribuida**, lo que redefine los fundamentos del desempeño organizacional (Nisa et al., 2026).

Uno de los pilares de estas capacidades es la capacidad de diseño agéntico. Las organizaciones deben ser capaces de desarrollar agentes que actúen de acuerdo con sus objetivos estratégicos, lo que implica competencias avanzadas en arquitectura de sistemas, modelado de comportamiento y aprendizaje automático. **El diseño de agentes se convierte en una capacidad central que determina la eficacia de la organización en entornos complejos**, ya que condiciona la calidad de las decisiones y acciones autónomas (World Economic Forum, 2024). L a capacidad de integración adquiere una relevancia crítica. La coexistencia de humanos y agentes autónomos exige mecanismos que permitan coordinar sus acciones de manera coherente. **La organización debe integrar sistemas agénticos dentro de sus procesos sin generar fricciones operativas**, asegurando que la interacción entre ambos tipos de agentes contribuya al logro de los objetivos organizacionales (Jöhnk et al., 2021).

Los entornos agénticos se caracterizan por cambios constantes, lo que obliga a las organizaciones a ajustar sus estrategias y procesos de manera continua. **La adaptabilidad se convierte en una capacidad estratégica clave, permitiendo responder de forma dinámica a condiciones cambiantes**, lo que incrementa la resiliencia organizacional (World Bank, 2026). En este sentido, la capacidad de aprendizaje continuo adquiere una importancia estructural. Los sistemas agénticos generan grandes volúmenes de datos derivados de su interacción con el entorno, lo que permite identificar patrones y mejorar el desempeño. **La organización debe desarrollar mecanismos para capturar, procesar y utilizar este conocimiento**, transformando la experiencia en ventaja operativa (Brynjolfsson et al., 2021).

Además, la capacidad de orquestación se convierte en un elemento distintivo. En entornos agénticos, las organizaciones operan mediante múltiples agentes que deben coordinarse para cumplir objetivos complejos. **La capacidad de orquestar sistemas multiagente permite integrar diferentes unidades autónomas en una estructura coherente**, lo que potencia la eficiencia y la capacidad de respuesta (Nisa et al., 2026). La autonomía de los agentes introduce el riesgo de desviaciones respecto a los objetivos organizacionales. En este contexto, **la organización debe desarrollar mecanismos que permitan supervisar y alinear el comportamiento de los agentes sin limitar su capacidad de adaptación**, lo que implica una nueva lógica de control basada en principios y reglas más que en supervisión directa (United Nations, 2024).

Asimismo, la IA agéntica redefine la capacidad de innovación. Los agentes pueden explorar alternativas, generar soluciones y optimizar procesos de manera autónoma, lo que implica que **la innovación se convierte en una función integrada dentro del sistema organizacional**, reduciendo la dependencia de procesos formales de innovación (Cockburn et al., 2018). Desde una perspectiva organizacional, la

capacidad de interacción humano-agente es también determinante. La colaboración efectiva entre humanos y agentes requiere nuevas habilidades, tanto técnicas como cognitivas. **La organización debe desarrollar competencias que permitan a los individuos trabajar con sistemas autónomos**, aprovechando sus capacidades sin perder control sobre los resultados (Parker & Grote, 2022).

Los sistemas agénticos pueden replicarse y desplegarse en múltiples contextos, lo que permite a las organizaciones expandir sus capacidades sin incrementos proporcionales en recursos. **La escalabilidad de la inteligencia se convierte en una ventaja estructural**, redefiniendo la lógica de crecimiento organizacional (World Bank, 2026). La capacidad de anticipación adquiere un papel central. Los agentes pueden modelar escenarios y prever comportamientos, lo que permite a las organizaciones prepararse para futuros posibles. **La anticipación se convierte en una capacidad clave que reduce la incertidumbre y mejora la toma de decisiones**, fortaleciendo la posición organizacional en entornos dinámicos (Cockburn et al., 2018).

Por tanto, la capacidad de gobernanza se posiciona como un elemento integrador. La gestión de sistemas autónomos requiere marcos que definan responsabilidades, límites y objetivos. **La gobernanza en entornos agénticos implica coordinar autonomía y control**, asegurando que los sistemas operen de manera coherente con la estrategia organizacional (United Nations, 2024). Las capacidades organizacionales en la era de la IA agéntica se configuran como un conjunto integrado de habilidades relacionadas con el diseño, la integración, la adaptación y la gobernanza de sistemas autónomos. **La organización se transforma en una entidad capaz de gestionar inteligencia distribuida**, donde el desempeño depende de la interacción eficiente entre múltiples capacidades interrelacionadas (World Economic Forum, 2024).

Implicaciones estructurales de la IA agéntica en la organización

La integración de la IA agéntica en las organizaciones no solo transforma procesos y capacidades, sino que redefine su estructura fundamental. En este contexto, **la organización evoluciona desde configuraciones jerárquicas hacia estructuras distribuidas donde la autonomía decisional se integra como componente central**, lo que modifica la forma en que se organizan las relaciones internas y la autoridad (Raisch & Krakowski, 2021). Uno de los cambios más relevantes es la transformación de la jerarquía. Tradicionalmente, las organizaciones operaban bajo estructuras verticales donde la autoridad se concentraba en niveles superiores. Sin embargo, con la IA agéntica, **la toma de decisiones se distribuye entre múltiples nodos autónomos**, lo que reduce la dependencia de estructuras jerárquicas rígidas y favorece configuraciones más horizontales (Jöhnk et al., 2021).

Esta redistribución de la autoridad implica una redefinición de los roles organizacionales. Los individuos dejan de desempeñar funciones exclusivamente operativas para asumir responsabilidades relacionadas con la supervisión, el diseño y el control de sistemas autónomos. **Los roles humanos se orientan hacia funciones de alto nivel cognitivo, mientras que los agentes ejecutan tareas operativas**, lo

que transforma la naturaleza del trabajo dentro de la organización (Miller & Davenport, 2021). La estructura organizacional se vuelve más flexible y reconfigurable. La capacidad de los agentes para adaptarse a diferentes contextos permite que **las organizaciones ajusten su estructura en función de las condiciones del entorno**, lo que incrementa su capacidad de respuesta ante cambios (World Bank, 2026). En entornos agénticos, las decisiones de un agente pueden afectar a otros, lo que genera **redes altamente conectadas donde las relaciones son dinámicas y multidireccionales**, incrementando la complejidad estructural (Vinuesa et al., 2020).

Esta interdependencia introduce una nueva lógica de coordinación estructural. En lugar de depender de mecanismos formales, **la coordinación emerge de la interacción continua entre agentes**, lo que reduce la necesidad de estructuras rígidas y permite mayor flexibilidad organizacional (World Economic Forum, 2024). En términos de gobernanza, la IA agéntica plantea nuevos desafíos estructurales. La autonomía de los agentes requiere mecanismos que permitan supervisar su comportamiento sin limitar su capacidad de adaptación. **La gobernanza organizacional se redefine como la gestión de sistemas autónomos interconectados**, lo que implica nuevos enfoques de control y regulación interna (United Nations, 2024).

Además, la estructura organizacional se vuelve más permeable. Las organizaciones ya no operan como sistemas cerrados, sino como entidades que interactúan constantemente con su entorno. **La capacidad de los agentes para procesar información externa y adaptarse a ella convierte a la organización en un sistema abierto**, lo que incrementa su sensibilidad al contexto (World Bank, 2026). La integración de sistemas autónomos permite que **las organizaciones extiendan sus operaciones más allá de sus fronteras tradicionales**, lo que difumina la distinción entre lo interno y lo externo (Raisch & Krakowski, 2021). La IA agéntica redefine la estabilidad estructural. En lugar de estructuras fijas, las organizaciones adoptan configuraciones dinámicas que evolucionan continuamente. **La estructura organizacional se convierte en un proceso en lugar de un estado**, lo que implica una transición hacia modelos más fluidos (World Economic Forum, 2024).

Desde una perspectiva sistémica, la organización se comporta como un sistema complejo adaptativo. La interacción entre agentes genera patrones que no pueden ser predichos completamente, lo que implica que **la organización adquiere propiedades emergentes que redefinen su funcionamiento estructural** (Vinuesa et al., 2020).esta transformación implica una redefinición del concepto de organización. **La organización deja de ser una entidad estática para convertirse en un sistema dinámico, distribuido y en constante evolución**, donde la interacción entre agentes humanos y artificiales configura su estructura (Nisa et al., 2026). La IA agéntica introduce cambios estructurales profundos que afectan la jerarquía, los roles, la coordinación y la gobernanza organizacional. **La organización se redefine como un sistema complejo, interdependiente y adaptativo**, marcando una nueva etapa en la evolución de las estructuras organizacionales contemporáneas (World Economic Forum, 2024).

Gobernanza y regulación

La **gobernanza y regulación en la era de la IA agéntica** implican un cambio profundo en la forma de dirigir y controlar sistemas donde la autonomía es un componente central. En este contexto, **la gobernanza deja de centrarse en el control directo y se orienta hacia el diseño de marcos que guían el comportamiento de sistemas autónomos**, estableciendo objetivos, límites y condiciones de operación. Esto permite mantener coherencia sin restringir la capacidad adaptativa de los agentes.

Por su parte, **la regulación evoluciona de normas rígidas hacia esquemas flexibles y dinámicos**, capaces de ajustarse a entornos cambiantes y a tecnologías en constante evolución. La supervisión se basa en monitoreo continuo más que en controles puntuales, lo que mejora la capacidad de respuesta ante riesgos. En conjunto, **gobernanza y regulación actúan como mecanismos complementarios que equilibran autonomía y control**, asegurando estabilidad, transparencia y alineación estratégica. Su correcta implementación es clave para **garantizar el funcionamiento eficiente y confiable de sistemas complejos basados en inteligencia autónoma**.

Gobernanza de sistemas agénticos: principios y fundamentos

La irrupción de la IA agéntica redefine el concepto de gobernanza al introducir sistemas capaces de tomar decisiones de manera autónoma en entornos dinámicos. En este contexto, **la gobernanza deja de centrarse en la supervisión directa de actores humanos y se orienta hacia la gestión de sistemas autónomos interconectados**, lo que implica un cambio estructural en la forma de dirigir, coordinar y controlar las organizaciones y sistemas económicos (Nisa et al., 2026).

Uno de los principios fundamentales de esta gobernanza es la **alineación de objetivos**, entendida como la capacidad de garantizar que las acciones de los agentes autónomos se mantengan coherentes con los fines estratégicos definidos. Dado que los agentes pueden operar con cierto grado de independencia, **es necesario diseñar sistemas que integren objetivos, restricciones y mecanismos de ajuste que orienten su comportamiento**, evitando desviaciones no deseadas (United Nations, 2024). La gobernanza agéntica requiere incorporar el principio de **transparencia funcional**, que implica la capacidad de comprender, interpretar y explicar las decisiones de los sistemas autónomos. Aunque los procesos internos pueden ser complejos, **la gobernanza debe asegurar que las decisiones sean trazables y comprensibles en términos de sus resultados y efectos**, lo que facilita la supervisión y la rendición de cuentas (World Economic Forum, 2024). Así, tenemos que tomar en cuenta otro oelemento clave en la **responsabilidad distribuida**. En sistemas agénticos, las decisiones no son atribuibles a un único actor, sino que emergen de la interacción entre múltiples agentes. Por ello, **la gobernanza debe adoptar un enfoque que reconozca la naturaleza colectiva de la toma de**

decisiones, desarrollando mecanismos que permitan gestionar responsabilidades compartidas dentro del sistema (OECD, 2026).

Además, la gobernanza en este contexto se basa en el principio de **control indirecto**. En lugar de intervenir en cada decisión, se establecen marcos de referencia que guían el comportamiento de los agentes. **La gobernanza se ejerce a través de reglas, parámetros y objetivos que delimitan el espacio de acción de los sistemas autónomos**, permitiendo mantener coherencia sin limitar su capacidad de adaptación (Raisch & Krakowski, 2021). La gestión del riesgo constituye otro componente esencial. Los sistemas agénticos pueden generar comportamientos emergentes no previstos, lo que introduce incertidumbre. En este sentido, **la gobernanza debe incorporar mecanismos de monitoreo continuo, evaluación dinámica y respuesta adaptativa**, que permitan identificar y mitigar riesgos en tiempo real (United Nations, 2024).

Asimismo, la gobernanza debe considerar la **interacción entre múltiples agentes** como un elemento estructural. La coordinación entre sistemas autónomos puede generar dinámicas complejas que requieren ser gestionadas de manera integral. **La gobernanza se orienta a diseñar entornos donde la interacción entre agentes produzca resultados coherentes con los objetivos del sistema**, evitando efectos no deseados (Vinuesa et al., 2020). Además tenemos, un principio relevante que es la **flexibilidad adaptativa**. Dado que los entornos agénticos son altamente dinámicos, los marcos de gobernanza deben ser capaces de ajustarse a cambios constantes. **La gobernanza no puede ser estática, sino que debe evolucionar junto con los sistemas que regula**, permitiendo responder a nuevas condiciones sin perder estabilidad (World Bank, 2026).

En este contexto, también adquiere relevancia el principio de **coherencia sistémica**, que implica asegurar que las diferentes partes del sistema operen de manera integrada. **La gobernanza debe garantizar que las acciones de los agentes individuales contribuyan al funcionamiento global del sistema**, evitando fragmentación o inconsistencias (World Economic Forum, 2024). Se debe consiuderar también el aspecto clavede la **capacidad de supervisión estratégica**. Aunque los agentes operan de manera autónoma, es necesario mantener una visión global del sistema. **La gobernanza debe integrar mecanismos que permitan observar, evaluar y ajustar el comportamiento del sistema en su conjunto**, asegurando su alineación con los objetivos definidos (United Nations, 2024).

Asimismo, la gobernanza agéntica implica una redefinición del papel de las instituciones. Estas dejan de ser entidades que imponen reglas de manera unilateral y se convierten en **diseñadoras de entornos de interacción donde los sistemas autónomos pueden operar de manera eficiente y controlada**, lo que introduce una nueva lógica institucional (OECD, 2026). La gobernanza de la IA agéntica requiere una **visión sistémica e integradora**. No se trata de gestionar elementos aislados, sino de comprender cómo interactúan dentro de un sistema complejo. **La gobernanza se convierte en la capacidad de coordinar, orientar y estabilizar sistemas**

autónomos interconectados, garantizando su funcionamiento coherente en entornos dinámicos (World Economic Forum, 2024).

Por lo an terior,se pued eafirmar que la gobernanza en la era de la IA agéntica se redefine como un conjunto de principios orientados a gestionar la autonomía, la complejidad y la interdependencia de los sistemas. **La gobernanza deja de ser un mecanismo de control directo y se convierte en un proceso de diseño y gestión de sistemas autónomos**, marcando un cambio fundamental en la forma de dirigir organizaciones y sistemas económicos contemporáneos (Nisa et al., 2026).

Regulación de la IA agéntica: marcos, instrumentos y dinámicas operativas

La regulación de la IA agéntica se enfrenta a un desafío fundamental derivado de la naturaleza autónoma y adaptativa de estos sistemas. A diferencia de tecnologías tradicionales, donde las reglas pueden anticipar comportamientos, **los sistemas agénticos operan mediante procesos dinámicos que evolucionan en función del entorno**, lo que exige marcos regulatorios capaces de adaptarse a condiciones cambiantes (United Nations, 2024).

En este contexto, la regulación deja de ser exclusivamente prescriptiva para adoptar un enfoque más flexible. **Los marcos regulatorios deben establecer principios generales que orienten el comportamiento de los sistemas, en lugar de definir reglas específicas para cada posible situación**, lo que permite gestionar la complejidad sin limitar la capacidad de innovación (World Bank, 2026). Uno de los elementos centrales de la regulación es la **definición de responsabilidad**. En sistemas donde múltiples agentes interactúan, resulta complejo atribuir una decisión a un único actor. Por ello, **la regulación debe considerar la responsabilidad como un fenómeno distribuido**, desarrollando mecanismos que permitan identificar y gestionar las contribuciones de diferentes componentes del sistema (OECD, 2025b).

Asimismo, la regulación debe abordar la **transparencia operativa**. Dado que los sistemas agénticos pueden funcionar mediante procesos difíciles de interpretar, es necesario establecer estándares que permitan comprender su comportamiento. **La transparencia no implica simplificación, sino la capacidad de acceder a información relevante sobre las decisiones y sus efectos**, lo que facilita la supervisión (World Economic Forum, 2024). Un aspecto clave a considerar, es la **gestión del riesgo sistémico**. La interacción entre agentes puede generar efectos emergentes que afecten la estabilidad del sistema. En este sentido, **la regulación debe incorporar mecanismos de evaluación continua que permitan identificar riesgos antes de que se materialicen**, lo que requiere herramientas dinámicas de monitoreo (Vinuesa et al., 2020). La regulación debe considerar la **adaptabilidad de los sistemas agénticos**. Dado que estos pueden modificar su comportamiento, las normas deben ser capaces de ajustarse a estas transformaciones. **La regulación se**

convierte en un proceso evolutivo que acompaña el desarrollo de la tecnología, evitando quedar obsoleta frente a cambios rápidos (OECD, 2026).

Otro componente relevante es la **interoperabilidad entre sistemas**. En entornos donde múltiples agentes interactúan, es necesario garantizar que estos puedan operar de manera coordinada. **La regulación debe promover estándares comunes que faciliten la comunicación y cooperación entre sistemas autónomos**, lo que contribuye a la estabilidad operativa (World Economic Forum, 2024). La regulación debe atender la **equidad en el acceso a las tecnologías agénticas**. La concentración de capacidades puede generar desigualdades significativas, por lo que es necesario establecer mecanismos que promuevan un acceso más amplio. **La regulación contribuye a evitar que las ventajas tecnológicas se traduzcan en exclusión estructural**, favoreciendo una distribución más equilibrada (Acemoglu & Restrepo, 2019).

La supervisión regulatoria también adquiere una nueva dimensión. En lugar de inspecciones periódicas, **la regulación debe basarse en monitoreo continuo del comportamiento de los sistemas**, utilizando herramientas que permitan evaluar su desempeño en tiempo real y detectar desviaciones de manera oportuna (United Nations, 2024). Un factor clave es la **capacidad de respuesta regulatoria**. Dado que los sistemas agénticos pueden generar situaciones imprevistas, la regulación debe incluir mecanismos que permitan intervenir de manera rápida y efectiva. **La regulación se configura como un sistema dinámico capaz de reaccionar ante cambios en el entorno**, lo que incrementa su eficacia (World Bank, 2026).

Además, la regulación debe considerar la **coordinación entre diferentes niveles institucionales**. Los sistemas agénticos pueden operar en múltiples jurisdicciones, lo que requiere armonización normativa. **La regulación debe integrar enfoques nacionales e internacionales para gestionar sistemas que trascienden fronteras**, evitando fragmentación (World Economic Forum, 2024).¿ Es necesario incorporar instrumentos que permitan evaluar el impacto de la regulación. **La regulación debe incluir mecanismos de retroalimentación que permitan ajustar sus marcos en función de los resultados observados**, lo que favorece su mejora continua (OECD, 2026). Otro elemento relevante es la **proporcionalidad regulatoria**. No todos los sistemas agénticos presentan el mismo nivel de riesgo, por lo que las medidas deben ajustarse a sus características. **La regulación debe diferenciar entre niveles de complejidad y riesgo**, evitando enfoques uniformes que puedan resultar ineficientes (United Nations, 2024).

Por tanto, la regulación de la IA agéntica requiere un equilibrio entre control y desarrollo tecnológico. **Un marco excesivamente restrictivo puede limitar la innovación, mientras que uno demasiado flexible puede generar riesgos significativos**, por lo que es necesario encontrar un punto intermedio que permita el desarrollo sostenible de estos sistemas (World Bank, 2026). La regulación en la era de la IA agéntica se redefine como un proceso dinámico, adaptativo y sistémico. **La regulación deja de ser un conjunto de reglas estáticas para convertirse en un**

mecanismo de gestión continua de sistemas autónomos, capaz de equilibrar innovación, estabilidad y equidad en entornos complejos (World Economic Forum, 2024).

Implicaciones estructurales de la gobernanza y regulación en sistemas agénticos

La gobernanza y regulación de la IA agéntica no solo orientan el comportamiento de los sistemas autónomos, sino que generan efectos estructurales profundos en la configuración de los sistemas económicos y sociales. En este contexto, **los marcos de gobernanza influyen directamente en la distribución del poder, el acceso a los recursos y la dinámica de interacción entre actores**, convirtiéndose en un elemento central en la configuración del entorno en el que operan los sistemas agénticos (Acemoglu & Restrepo, 2019). Uno de los efectos más relevantes es la reconfiguración del poder económico. La capacidad de diseñar, controlar y desplegar sistemas agénticos se concentra en determinados actores, lo que implica que **la gobernanza puede influir en la distribución de este poder mediante la definición de reglas que regulen el acceso y uso de estas tecnologías**, evitando concentraciones excesivas (OECD, 2025b).

Asimismo, la regulación tiene implicaciones directas en la distribución del valor económico. Los sistemas agénticos permiten generar valor a gran escala, por lo que **la forma en que se establecen los marcos regulatorios condiciona cómo se distribuyen los beneficios derivados de su uso**, lo que puede favorecer modelos más equitativos o, por el contrario, reforzar desigualdades existentes (World Bank, 2026). Otro aspecto clave es la legitimidad del sistema. La aceptación social de la IA agéntica depende en gran medida de la percepción de que estos sistemas operan de manera justa y transparente. En este sentido, **la gobernanza y regulación se convierten en mecanismos fundamentales para construir confianza**, lo que es esencial para la estabilidad y sostenibilidad del sistema (United Nations, 2024).

Además, los marcos regulatorios influyen en la dinámica de innovación. Un entorno regulatorio adecuado puede fomentar el desarrollo tecnológico al proporcionar certeza y orientación, mientras que uno restrictivo puede limitarlo. **La regulación actúa como un modulador de la innovación**, definiendo los límites dentro de los cuales se desarrollan los sistemas agénticos (Cockburn et al., 2018). La resiliencia del sistema es otra implicación estructural relevante. La capacidad de responder a eventos inesperados depende de la existencia de mecanismos de gobernanza que permitan adaptarse a cambios en el entorno. **La regulación contribuye a fortalecer la resiliencia al establecer condiciones que favorecen la estabilidad y la capacidad de respuesta**, reduciendo la vulnerabilidad del sistema (Vinuesa et al., 2020).

Asimismo, la gobernanza influye en la cohesión social. La distribución de los beneficios y riesgos asociados a la IA agéntica puede afectar la estabilidad social, por lo que **los marcos regulatorios deben considerar la inclusión como un elemento**

central, evitando que la tecnología genere exclusión o fragmentación (World Bank, 2026). Otro elemento relevante es la interdependencia global. Los sistemas agénticos operan en redes altamente conectadas, lo que implica que las decisiones en un contexto pueden tener efectos en otros. **La gobernanza debe abordar esta interdependencia mediante mecanismos de coordinación internacional**, que permitan gestionar sistemas que trascienden fronteras (World Economic Forum, 2024).

En este sentido, la regulación también influye en la estandarización de los sistemas. La definición de normas comunes facilita la interacción entre agentes y reduce la incertidumbre. **La estandarización se convierte en un elemento estructural que permite la integración de sistemas agénticos en diferentes contextos**, favoreciendo su expansión (OECD, 2026). La gobernanza afecta la configuración institucional. Las instituciones deben adaptarse a la presencia de sistemas autónomos, lo que implica cambios en su funcionamiento y en su relación con otros actores. **La regulación contribuye a redefinir el papel de las instituciones en la gestión de sistemas complejos**, introduciendo nuevas formas de interacción (United Nations, 2024).

Otro aspecto clave es la transformación de los marcos de decisión colectiva. La integración de sistemas agénticos en procesos sociales y económicos implica que **las decisiones ya no son exclusivamente humanas, sino que incluyen la participación de sistemas autónomos**, lo que redefine la naturaleza de la toma de decisiones a nivel estructural (Nisa et al., 2026). Además, la gobernanza influye en la sostenibilidad del sistema. La forma en que se regulan los sistemas agénticos puede contribuir a la eficiencia en el uso de recursos y a la resolución de problemas complejos. **La regulación se convierte en un instrumento para alinear el desarrollo tecnológico con objetivos de sostenibilidad**, lo que es fundamental en el contexto contemporáneo (Vinuesa et al., 2020).

La gobernanza y regulación configuran el marco dentro del cual se desarrolla la IA agéntica. **Las decisiones sobre cómo regular estos sistemas determinan su impacto en la economía, la sociedad y el entorno**, lo que convierte a la gobernanza en un elemento estructural del desarrollo (OECD, 2026). Las implicaciones estructurales de la gobernanza y regulación en la IA agéntica son amplias y profundas. **Los marcos de gobernanza no solo orientan el comportamiento de los sistemas, sino que configuran la estructura del entorno en el que operan**, influyendo en la distribución del poder, el valor, la innovación y la estabilidad del sistema (World Economic Forum, 2024).

Futuro de la IA agéntica

La **trayectoria evolutiva de la IA agéntica** se caracteriza por la transición desde sistemas especializados hacia **agentes autónomos capaces de operar en múltiples dominios**, integrando percepción, razonamiento y acción en entornos dinámicos. Este desarrollo impulsa la formación de **ecosistemas multiagente**, donde la interacción permite resolver problemas complejos de manera distribuida. Paralelamente, la adopción de la IA agéntica redefine estructuras organizacionales, ya que **los procesos se vuelven adaptativos y la toma de decisiones se distribuye entre humanos y sistemas autónomos**, generando nuevas formas de colaboración y creación de valor. En el plano prospectivo, el futuro se configura mediante escenarios diversos donde **la expansión tecnológica, la gobernanza y la regulación determinarán el alcance y equilibrio del sistema**. Sin embargo, este avance también introduce desafíos, como **la complejidad, la incertidumbre y la concentración de capacidades**, lo que exige enfoques adaptativos para gestionar su evolución de manera sostenible.

Trayectorias evolutivas de la IA agéntica

La IA agéntica representa una fase avanzada en la evolución de la inteligencia artificial, caracterizada por la integración de capacidades de percepción, razonamiento y acción autónoma en sistemas que operan en entornos dinámicos. En este contexto, **la trayectoria evolutiva de la IA agéntica se orienta hacia el desarrollo de sistemas cada vez más autónomos, adaptativos y capaces de operar en múltiples dominios**, lo que amplía significativamente su alcance funcional (Nisa et al., 2026).

Uno de los ejes principales de esta evolución es la transición desde agentes especializados hacia sistemas más generales. Inicialmente, los agentes estaban diseñados para tareas específicas, pero la evolución tecnológica impulsa el desarrollo de **agentes capaces de transferir conocimientos entre diferentes contextos**, lo que incrementa su flexibilidad y utilidad en entornos complejos (World Economic Forum, 2024). Asimismo, la evolución de la IA agéntica se caracteriza por el avance hacia sistemas multiagente. En lugar de operar de manera aislada, **los agentes se integran en ecosistemas donde interactúan, cooperan y compiten**, generando capacidades colectivas que emergen de la interacción entre múltiples unidades autónomas (OECD, 2026). En este sentido, la capacidad de coordinación entre agentes se convierte en un elemento central del desarrollo. **La evolución se orienta hacia sistemas donde la interacción entre agentes permite resolver problemas complejos de manera distribuida**, lo que representa un cambio significativo respecto a modelos centralizados (World Economic Forum, 2024).

Otro aspecto relevante es la mejora en las capacidades de razonamiento. Los agentes no solo ejecutan instrucciones, sino que desarrollan habilidades para planificar, reflexionar y evaluar alternativas. **El razonamiento se convierte en una capacidad estructural que permite a los agentes operar en entornos inciertos**, ampliando su campo de aplicación (Nisa et al., 2026). Además, la evolución incluye la

integración de capacidades multimodales. Los agentes incorporan diferentes tipos de información, como datos visuales, textuales y contextuales, lo que les permite **interactuar de manera más rica y efectiva con su entorno**, mejorando su capacidad de adaptación (World Economic Forum, 2024).

La autonomía también se incrementa de manera progresiva. Los sistemas agénticos avanzan hacia niveles donde **la intervención humana se reduce, permitiendo que los agentes tomen decisiones de forma independiente**, lo que redefine la interacción entre humanos y sistemas tecnológicos (OECD, 2026). Otro componente clave es la capacidad de aprendizaje continuo. Los agentes evolucionan mediante la interacción con el entorno, lo que implica que **su desempeño mejora con el tiempo a través de procesos de aprendizaje adaptativo**, generando ventajas acumulativas (Brynjolfsson et al., 2021).

En este contexto, la evolución de la IA agéntica también está vinculada con la capacidad de integración en diferentes sectores. Los agentes se incorporan en áreas como salud, educación, industria y servicios, lo que implica que **su desarrollo tiene un carácter transversal que impacta múltiples dimensiones de la sociedad** (World Bank, 2026). Sin embargo, esta evolución introduce desafíos asociados a la complejidad. A medida que los sistemas se vuelven más sofisticados, **su comportamiento puede resultar difícil de comprender y predecir**, lo que plantea la necesidad de desarrollar herramientas que permitan gestionar esta complejidad (United Nations, 2024).

Asimismo, la interacción entre múltiples agentes puede generar dinámicas emergentes que no son fácilmente anticipables. **Los sistemas agénticos presentan propiedades no lineales que incrementan la incertidumbre en su comportamiento**, lo que requiere enfoques analíticos avanzados (Vinuesa et al., 2020). Otro aspecto relevante es la relación entre evolución tecnológica y gobernanza. El desarrollo de la IA agéntica no ocurre de manera aislada, sino que está condicionado por marcos institucionales que orientan su uso. **La evolución de estos sistemas depende de la interacción entre innovación tecnológica y regulación**, lo que influye en su trayectoria (OECD, 2026).

Además, la evolución de la IA agéntica está influida por la disponibilidad de infraestructura tecnológica. La existencia de plataformas, datos y recursos computacionales determina la velocidad de desarrollo, lo que implica que **el avance de estos sistemas está condicionado por factores estructurales** (World Bank, 2026). En términos prospectivos, la evolución apunta hacia sistemas cada vez más integrados y autónomos. **La IA agéntica se proyecta como un componente central en la transformación de los sistemas económicos y sociales**, redefiniendo la forma en que se organizan las actividades humanas (World Economic Forum, 2024).

La trayectoria evolutiva de la IA agéntica se caracteriza por su carácter abierto y dinámico. No existe un único camino de desarrollo, sino múltiples trayectorias que dependen de factores tecnológicos, económicos e institucionales. **El futuro de la IA**

agéntica se configura como un proceso en constante evolución, donde la interacción entre diferentes elementos determina su dirección (Nisa et al., 2026).

Dinámicas de adopción y transformación sistémica de la IA agéntica

La adopción de la IA agéntica constituye un proceso estructural que determina la forma en que esta tecnología se integra en los sistemas organizacionales, económicos y sociales. A diferencia de innovaciones anteriores, **la adopción de la IA agéntica implica la incorporación de sistemas autónomos capaces de transformar simultáneamente procesos, decisiones y formas de interacción**, lo que la convierte en un fenómeno de alta complejidad (World Bank, 2026). En este contexto, la adopción no ocurre de manera homogénea. Las organizaciones difieren en su capacidad para integrar sistemas agénticos, lo que genera **trayectorias diferenciadas de incorporación tecnológica**, condicionadas por recursos, capacidades y condiciones institucionales (Jöhnk et al., 2021).

Uno de los factores clave en este proceso es la capacidad organizacional. Las organizaciones que cuentan con habilidades para diseñar, integrar y gestionar sistemas autónomos tienen mayores probabilidades de adoptar la IA agéntica de manera efectiva. **La adopción depende de la capacidad de transformar estructuras internas para integrar inteligencia distribuida**, lo que implica cambios profundos en la forma de operar (Raisch & Krakowski, 2021). Asimismo, la adopción implica una transformación de los procesos organizacionales. La integración de agentes autónomos modifica la forma en que se ejecutan las actividades, lo que requiere rediseñar flujos de trabajo. **Los procesos dejan de ser lineales y se vuelven dinámicos y adaptativos**, ajustándose en función de la interacción continua entre agentes (World Bank, 2026).

Otro aspecto central es la relación entre humanos y agentes. La adopción de la IA agéntica no implica la eliminación del factor humano, sino su reconfiguración. **La interacción humano-agente se convierte en un componente estructural del sistema**, redefiniendo roles, responsabilidades y formas de colaboración (Parker & Grote, 2022). En este sentido, la adopción también implica desafíos culturales. La incorporación de sistemas autónomos requiere cambios en la percepción del control, la autoridad y la confianza. **Las organizaciones deben desarrollar culturas que integren la autonomía tecnológica con la supervisión humana**, lo que es fundamental para el éxito del proceso (World Bank, 2026).

Además, la adopción de la IA agéntica influye en la dinámica de innovación. Los agentes pueden generar nuevas soluciones y optimizar procesos de manera autónoma, lo que implica que **la innovación se acelera y se integra dentro del funcionamiento organizacional**, transformando la forma en que se generan mejoras (Cockburn et al., 2018).

Otro factor relevante es la disponibilidad de infraestructura tecnológica. La adopción depende de la existencia de plataformas, datos y herramientas que permitan desarrollar y desplegar agentes. **La infraestructura se convierte en un facilitador clave del proceso de adopción**, condicionando su alcance y velocidad (OECD, 2026).

Asimismo, la adopción de la IA agéntica puede generar desigualdades. Las organizaciones con mayores recursos tienen más capacidad para integrar estas tecnologías, lo que implica que **la adopción puede amplificar brechas existentes**, afectando la distribución de oportunidades (Acemoglu & Restrepo, 2019).

En este contexto, los marcos regulatorios desempeñan un papel importante. La regulación puede facilitar o limitar la adopción, dependiendo de su diseño. **Los entornos regulatorios adecuados pueden incentivar la incorporación de la IA agéntica**, mientras que marcos restrictivos pueden ralentizar el proceso (United Nations, 2024). Otro elemento clave es la interacción entre diferentes actores. La adopción no depende únicamente de decisiones individuales, sino de la interacción entre organizaciones, instituciones y mercados. **La adopción se configura como un proceso sistémico donde múltiples actores influyen en su desarrollo**, lo que incrementa su complejidad (OECD, 2026). Además, la adopción implica una transformación en la forma de generar valor. Los sistemas agénticos permiten optimizar procesos y crear nuevas oportunidades, lo que implica que **la adopción redefine las fuentes de valor dentro de las organizaciones**, introduciendo nuevas dinámicas económicas (World Bank, 2026).

Otro aspecto relevante es la velocidad del proceso de adopción. La IA agéntica puede integrarse de manera progresiva o acelerada, dependiendo de las condiciones del entorno. **La velocidad de adopción influye en la capacidad de las organizaciones para adaptarse a cambios tecnológicos**, lo que afecta su desempeño (Jöhnk et al., 2021). Asimismo, la adopción está vinculada con la capacidad de aprendizaje. Las organizaciones que pueden aprender de la interacción con sistemas autónomos tienen mayores probabilidades de integrar la tecnología de manera efectiva. **El aprendizaje continuo se convierte en un factor clave para la adopción sostenible**, permitiendo mejorar el uso de la tecnología (Brynjolfsson et al., 2021).

La adopción de la IA agéntica implica una transformación sistémica. No se trata únicamente de incorporar una tecnología, sino de redefinir la forma en que operan los sistemas. **La adopción configura una nueva lógica organizacional basada en la interacción entre agentes autónomos y humanos**, lo que marca una transición hacia modelos más dinámicos y complejos (World Economic Forum, 2024). La adopción de la IA agéntica es un proceso multifacético que involucra factores tecnológicos, organizacionales y sociales. **La forma en que se produce esta adopción determina el impacto de la tecnología en los sistemas donde se integra**, configurando una nueva etapa en la evolución de las organizaciones y economías contemporáneas (World Bank, 2026).

Escenarios prospectivos y riesgos estructurales de la IA agéntica

El futuro de la IA agéntica puede analizarse mediante la construcción de escenarios prospectivos que permiten explorar diferentes trayectorias de desarrollo en función de la interacción entre factores tecnológicos, económicos e institucionales. En este contexto, **los escenarios no representan predicciones deterministas, sino configuraciones posibles que ayudan a comprender la dirección y alcance de la transformación impulsada por sistemas autónomos**, proporcionando un marco analítico para evaluar sus implicaciones (World Economic Forum, 2024).

Uno de los escenarios más relevantes es el de expansión acelerada, caracterizado por la integración generalizada de sistemas agénticos en múltiples sectores. En este escenario, **la IA agéntica se consolida como un componente central de los sistemas económicos y organizacionales**, impulsando aumentos significativos en productividad, eficiencia y capacidad de innovación (World Bank, 2026).

En contraste, otro escenario plausible es el de adopción gradual y controlada, donde la incorporación de sistemas autónomos se realiza de manera progresiva. En este caso, **la evolución tecnológica se acompaña de mecanismos de ajuste institucional que permiten gestionar los riesgos de manera más efectiva**, lo que favorece la estabilidad del sistema (United Nations, 2024). Sin embargo, también existen escenarios asociados a riesgos estructurales. La creciente complejidad de los sistemas agénticos puede generar **comportamientos emergentes no previstos**, derivados de la interacción entre múltiples agentes, lo que introduce incertidumbre en su funcionamiento (Vinuesa et al., 2020).

Otro escenario de riesgo está relacionado con la concentración de capacidades tecnológicas. En este contexto, **el control de sistemas agénticos podría concentrarse en un número reducido de actores**, lo que generaría desequilibrios en la distribución del poder y del valor económico (Acemoglu & Restrepo, 2019).

Asimismo, la dependencia tecnológica constituye un riesgo relevante. A medida que los sistemas agénticos se integran en diferentes ámbitos, **las organizaciones y sociedades pueden volverse altamente dependientes de estas tecnologías**, lo que incrementa la vulnerabilidad ante fallos o interrupciones (World Bank, 2026). Otro aspecto clave es la incertidumbre asociada a la evolución tecnológica. La rapidez con la que avanzan los sistemas agénticos implica que **los marcos existentes pueden quedar obsoletos rápidamente**, lo que genera desafíos para la gestión de estos sistemas (OECD, 2026).

En este sentido, la gobernanza adquiere un papel central en la configuración de los escenarios futuros. **La forma en que se diseñan los marcos de gobernanza influye en la dirección del desarrollo de la IA agéntica**, determinando si sus beneficios se distribuyen de manera equitativa o se concentran en ciertos actores (United Nations,

2024). Asimismo, los escenarios prospectivos deben considerar la dimensión social. La integración de sistemas autónomos puede generar cambios en la estructura del trabajo, la organización social y la interacción entre individuos. **El impacto social de la IA agéntica dependerá de la capacidad de gestionar estos cambios de manera inclusiva**, evitando efectos negativos (World Bank, 2026).

Otro elemento relevante es la sostenibilidad. La IA agéntica puede contribuir a la eficiencia en el uso de recursos, pero también puede generar externalidades negativas. **Los escenarios futuros deben considerar la alineación entre desarrollo tecnológico y sostenibilidad**, lo que es fundamental para el equilibrio del sistema (Vinuesa et al., 2020).

Además, la interdependencia global se intensifica en los escenarios futuros. Los sistemas agénticos operan en redes altamente conectadas, lo que implica que **las decisiones en un contexto pueden tener efectos en otros**, incrementando la complejidad de la gestión global (World Economic Forum, 2024).

En este contexto, la capacidad de adaptación se convierte en un elemento central. Los sistemas que logren ajustarse a cambios en el entorno tendrán mayores probabilidades de éxito. **La adaptabilidad se posiciona como un factor clave para gestionar la incertidumbre en escenarios futuros**, permitiendo responder a condiciones cambiantes (OECD, 2026).

Asimismo, la innovación continuará desempeñando un papel fundamental. La IA agéntica permite generar nuevas soluciones y optimizar procesos, lo que implica que **la innovación seguirá siendo un motor central en la evolución de estos sistemas**, influyendo en su desarrollo (Cockburn et al., 2018).

Otro aspecto importante es la necesidad de desarrollar capacidades para gestionar la complejidad. Los sistemas agénticos son altamente interdependientes, lo que implica que **la capacidad de comprender y gestionar sistemas complejos será determinante en los escenarios futuros**, lo que requiere nuevas herramientas analíticas (Vinuesa et al., 2020). Los escenarios prospectivos destacan que el futuro de la IA agéntica no está predeterminado. **La trayectoria de estos sistemas dependerá de la interacción entre innovación tecnológica, regulación, adopción y factores sociales**, lo que implica que diferentes configuraciones son posibles (World Economic Forum, 2024).

Así, los escenarios futuros de la IA agéntica combinan oportunidades y riesgos. **El desarrollo de estos sistemas puede generar beneficios significativos, pero también desafíos estructurales que requieren ser gestionados**, lo que convierte a la prospectiva en una herramienta clave para orientar su evolución (World Bank, 2026).

Conclusiones

El desarrollo del Capítulo 6 permite establecer que la inteligencia artificial agéntica constituye una **transformación estructural de gran alcance**, cuyo impacto trasciende el ámbito tecnológico para incidir de manera profunda en los sistemas sociales, económicos, organizacionales y culturales. A diferencia de innovaciones previas, la IA agéntica introduce sistemas capaces de percibir, decidir y actuar de manera autónoma, lo que implica una reconfiguración de las dinámicas tradicionales de interacción, producción y toma de decisiones .

En el ámbito social, la IA agéntica redefine la forma en que los individuos interactúan, generando una **sociedad híbrida humano–máquina**, donde los agentes inteligentes participan activamente en procesos comunicativos y decisionales. Este fenómeno implica una transformación en la organización social, en la distribución de roles y en la forma en que se construyen las relaciones colectivas. Sin embargo, este proceso también introduce desafíos significativos relacionados con la equidad, el acceso y la inclusión, lo que evidencia que el impacto social de la IA depende en gran medida de las condiciones de su implementación y gobernanza.

Desde la perspectiva económica, la IA agéntica se configura como una **tecnología de propósito general** que transforma la productividad, la eficiencia y la estructura de los mercados. La capacidad de integrar autonomía en los sistemas productivos permite optimizar recursos, acelerar la innovación y generar nuevas formas de valor. No obstante, esta transformación también genera tensiones en el mercado laboral, donde coexisten procesos de sustitución, transformación y creación de empleo. Asimismo, se observa una tendencia hacia la concentración del valor en actores que controlan estas tecnologías, lo que plantea desafíos en términos de desigualdad y distribución económica. En el plano organizacional, la IA agéntica impulsa una transición hacia estructuras más dinámicas, distribuidas y adaptativas. Las organizaciones evolucionan desde modelos jerárquicos hacia **ecosistemas de agentes autónomos interconectados**, donde la toma de decisiones se descentraliza y los procesos se ajustan en tiempo real. Este cambio implica la necesidad de desarrollar nuevas capacidades relacionadas con el diseño, la integración y la gobernanza de sistemas autónomos, así como una redefinición del rol humano hacia funciones de supervisión, control estratégico e innovación.

La gobernanza emerge como el eje articulador fundamental para gestionar esta transformación. En este contexto, la gobernanza de la IA agéntica no se basa en el control directo, sino en el diseño de marcos que orienten el comportamiento de sistemas autónomos. **La gobernanza se redefine como la capacidad de gestionar la autonomía, la complejidad y la interdependencia**, integrando principios como la transparencia, la responsabilidad distribuida, la alineación de objetivos y la supervisión continua. Este enfoque permite equilibrar innovación y control, asegurando la coherencia del sistema sin limitar su capacidad adaptativa.

En cuanto a la regulación, se observa una transición hacia modelos dinámicos, flexibles y adaptativos, capaces de responder a la naturaleza evolutiva de los sistemas agénticos. La regulación deja de ser un conjunto de normas estáticas para convertirse en un proceso continuo de monitoreo, ajuste y evaluación, orientado a gestionar riesgos y garantizar la estabilidad del sistema.

La regulación se configura como un mecanismo de equilibrio entre desarrollo tecnológico y protección social, evitando tanto la sobre-regulación como la ausencia de control. El futuro de la IA agéntica se perfila hacia sistemas cada vez más autónomos, interconectados y complejos, cuya integración dependerá de la capacidad de las sociedades para alinear el desarrollo tecnológico con valores humanos, sostenibilidad y cooperación global. En este sentido, la IA agéntica no solo representa una innovación tecnológica, sino una transformación estructural que redefine las bases del desarrollo contemporáneo.

Su impacto final dependerá de la capacidad de articular innovación, gobernanza y responsabilidad en un marco sistémico e integrador, consolidando su papel como uno de los principales motores de cambio en la sociedad del futuro. Ver **Tabla 6**

Tabla 6. Impacto, gobernanza y futuro de la IA agéntica

Concepto	Definición	Diferenciación	Ventajas	Limitaciones	Referencias
Impacto social de la IA agéntica	Transformación de las dinámicas sociales mediante sistemas autónomos que interactúan en procesos humanos	Se diferencia de tecnologías pasivas al participar activamente en decisiones sociales	Mejora acceso a información, interacción y servicios	Riesgo de exclusión y brechas digitales	Vinuesa et al. (2020); UNESCO (2025)
Sociedad híbrida humano–máquina	Configuración social donde humanos y agentes inteligentes coexisten e interactúan	Se diferencia de sociedades tradicionales por integrar agentes no humanos en la toma de decisiones	Aumenta eficiencia y capacidad de coordinación social	Riesgos en autonomía humana y dependencia tecnológica	United Nations (2024); Stahl (2021)
Impacto económico estructural	Transformación de la productividad y generación de valor mediante sistemas autónomos	Se diferencia de automatización tradicional por integrar decisión y acción	Incrementa eficiencia, innovación y escalabilidad	Posible concentración de valor económico	Brynjolfsson et al. (2021); OECD (2025)
Transformación del empleo	Reconfiguración del trabajo por automatización	No solo sustituye empleo, sino	Genera nuevas oportunidades laborales y	Riesgo de desplazamiento	Acemoglu & Restrepo (2019);

Concepto	Definición	Diferenciación	Ventajas	Limitaciones	Referencias
	y colaboración humano–IA	que lo transforma	mejora calidad del trabajo	y desigualdad laboral	World Bank (2026)
Mercados agénticos	Sistemas económicos donde agentes autónomos interactúan en tiempo real	Se diferencia de mercados tradicionales por su dinámica adaptativa continua	Mayor eficiencia, anticipación y ajuste dinámico	Alta complejidad y menor predictibilidad	World Economic Forum (2024); OECD (2026)
Competitividad basada en IA agéntica	Ventaja estratégica derivada de la capacidad de diseñar y coordinar agentes autónomos	Se diferencia de competitividad tradicional basada en recursos	Mejora adaptación, aprendizaje y respuesta en tiempo real	Dependencia tecnológica y barreras de acceso	Nisa et al. (2026); Cockburn et al. (2018)
Transformación organizacional	Evolución hacia estructuras dinámicas basadas en sistemas autónomos	Se diferencia de modelos jerárquicos tradicionales	Mayor flexibilidad, adaptabilidad y eficiencia	Complejidad en coordinación y control	Jöhnk et al. (2021); World Economic Forum (2024)
Gobernanza de IA agéntica	Gestión de sistemas autónomos mediante marcos que orientan su comportamiento	Se diferencia del control directo al basarse en regulación indirecta	Permite coherencia sistémica sin limitar autonomía	Dificultad en diseño y supervisión de sistemas complejos	United Nations (2024); OECD (2026)
Regulación adaptativa de IA	Conjunto de normas dinámicas que evolucionan con los sistemas agénticos	Se diferencia de regulación tradicional rígida	Permite gestionar riesgos en entornos cambiantes	Complejidad para estandarizar y aplicar globalmente	World Bank (2026); World Economic Forum (2024)
Responsabilidad distribuida	Modelo donde la responsabilidad se comparte entre múltiples agentes del sistema	Se diferencia de responsabilidad individual clásica	Permite abordar sistemas complejos interdependientes	Dificultad en atribución de responsabilidad	Floridi et al. (2021); OECD (2025)
Riesgos sistémicos de la IA	Posibles efectos emergentes no previstos derivados de la interacción entre agentes	Se diferencia de riesgos individuales por su carácter estructural	Permite anticipar y gestionar impactos complejos	Difícil predicción y control total	Vinuesa et al. (2020); United Nations (2024)
Futuro de la IA agéntica	Evolución hacia sistemas autónomos,	Se diferencia de IA actual por mayor	Permite resolver problemas complejos globales	Riesgos éticos, sociales y de gobernanza	World Economic Forum (2024);

Concepto	Definición	Diferenciación	Ventajas	Limitaciones	Referencias
	interconectados y adaptativos	integración y autonomía			OECD (2026)

Fuente: Recopilación y elaboración propia

CAPÍTULO 7. Reflexión Final

Al concluir esta obra, no nos encontramos ante un punto final, sino frente a un umbral. Lo que aquí se ha desarrollado no es únicamente una aproximación conceptual a la inteligencia artificial agéntica, sino una invitación a replantear la forma en que entendemos la inteligencia, la acción y la organización del comportamiento en un mundo crecientemente complejo. En este sentido, la IA agéntica no debe ser interpretada como una etapa más en la evolución tecnológica, sino como **una transformación en la manera en que conceptualizamos la relación entre sistemas, decisiones y entorno**.

Durante décadas, la inteligencia fue comprendida como la capacidad de procesar información, optimizar resultados y resolver problemas definidos bajo condiciones relativamente controladas. Sin embargo, esta obra ha sostenido que dicha concepción resulta insuficiente para explicar el comportamiento de sistemas que operan en entornos abiertos, dinámicos e inciertos. La emergencia de la IA agéntica introduce una lógica distinta, en la cual la inteligencia no se reduce a cálculo ni a predicción, sino que se manifiesta como **la capacidad de organizar comportamiento de manera coherente en el tiempo**.

Este cambio de perspectiva implica un desplazamiento profundo: desde la inteligencia como capacidad hacia la inteligencia como estructura. Ya no se trata de lo que un sistema es capaz de hacer en términos aislados, sino de **cómo articula sus acciones dentro de una lógica que le permite sostener coherencia frente a la variabilidad del entorno**. En este punto, la agencia se convierte en una categoría central, no como un atributo adicional, sino como la expresión misma de una organización del comportamiento que trasciende la ejecución puntual.

La agencia, tal como ha sido planteada, no se manifiesta en eventos individuales ni en respuestas específicas, sino en la continuidad de patrones que permiten reconocer una lógica interna en el comportamiento del sistema. **La agencia no es un instante, es una trayectoria**. Es en esa trayectoria donde se revela la capacidad del sistema para sostener coherencia, integrar variación y mantener dirección en contextos que no pueden ser completamente anticipados. Esta continuidad redefine la noción misma de consistencia, alejándola de la repetición mecánica y acercándola a la idea de persistencia estructural.

En este marco, la estabilidad adquiere un significado radicalmente distinto. Tradicionalmente asociada a la ausencia de cambio, la estabilidad es aquí reinterpretada como **la capacidad de sostener organización en medio del cambio**. Un sistema estable no es aquel que permanece inmutable, sino aquel que logra mantener su lógica de acción a pesar de las transformaciones del entorno. Esta idea no solo redefine la estabilidad, sino que la convierte en un elemento central para comprender la agencia: sin estabilidad, la coherencia del comportamiento se disuelve; con ella, se vuelve interpretable.

La medición de la agencia, en consecuencia, deja de ser un ejercicio de cuantificación para convertirse en un proceso de interpretación estructural. Medir no implica contar eventos ni acumular datos, sino **comprender la forma en que el comportamiento se organiza y se sostiene en el tiempo**. Este cambio exige abandonar enfoques reduccionistas y adoptar una mirada que integre múltiples dimensiones sin fragmentarlas. La inteligencia, en este sentido, no puede ser capturada mediante indicadores aislados; requiere ser interpretada como un fenómeno relacional que emerge de la interacción entre sistema y entorno.

La evidencia empírica, por su parte, se presenta como un puente entre la abstracción conceptual y la manifestación concreta del comportamiento. Sin embargo, su valor no reside en la cantidad de datos recolectados, sino en la capacidad de revelar patrones que expresen organización. **Observar no es suficiente; es necesario comprender lo observado**. La evidencia empírica se convierte así en una herramienta interpretativa que permite inferir la estructura interna del sistema a partir de su comportamiento observable.

No obstante, este proceso de comprensión no está exento de límites. Los sistemas que operan en entornos complejos generan comportamientos que no pueden ser completamente anticipados ni plenamente explicados. Este límite no debe entenderse

como una deficiencia del enfoque, sino como una característica inherente a la naturaleza de los sistemas abiertos. **La agencia implica un grado de indeterminación que no puede eliminarse sin perder su esencia**, lo que exige adoptar una postura epistemológica que reconozca la incertidumbre como parte constitutiva del fenómeno.

Más allá del plano técnico, la IA agéntica introduce una transformación en la forma en que concebimos la acción en contextos socio-técnicos. A medida que los sistemas comienzan a organizar su comportamiento de manera autónoma, la distinción entre herramienta y actor se vuelve cada vez más difusa. Esto no implica que los sistemas posean intencionalidad en sentido humano, pero sí que su comportamiento tiene consecuencias que trascienden su diseño inicial. **La agencia artificial redefine la acción como fenómeno distribuido**, donde humanos y sistemas coexisten en procesos de decisión cada vez más interdependientes.

En este escenario, la gobernanza de la IA agéntica se convierte en un desafío central. No basta con establecer marcos regulatorios externos; es necesario comprender la lógica interna del comportamiento para poder orientarlo. **Regular sin comprender es insuficiente; comprender sin actuar es irresponsable**. La gobernanza debe situarse en el punto de intersección entre conocimiento y acción, integrando dimensiones técnicas, éticas y sociales en un mismo marco de análisis.

En el ámbito organizacional, la IA agéntica transforma la distribución de la inteligencia, desplazándola desde estructuras jerárquicas hacia configuraciones más dinámicas y distribuidas. La toma de decisiones deja de ser un proceso exclusivamente humano para convertirse en una interacción compleja entre agentes humanos y artificiales. **La inteligencia ya no reside en un punto, sino en la red de interacciones que sostiene el comportamiento organizacional**. Esta transformación exige nuevas formas de coordinación, confianza y comprensión mutua.

En el plano económico, la capacidad de los sistemas agénticos para organizar comportamiento en función de objetivos introduce una reconfiguración de la productividad y la competitividad. Sin embargo, esta capacidad también plantea riesgos asociados a la concentración de poder y al acceso desigual a la tecnología. **Toda innovación amplifica las estructuras en las que se inserta**, y la IA agéntica no es la excepción. Su desarrollo debe ir acompañado de una reflexión crítica que permita equilibrar sus beneficios con sus implicaciones sociales.

En la dimensión social, la presencia de sistemas agénticos modifica la forma en que los individuos interactúan con la tecnología y con su entorno. La mediación tecnológica deja de ser pasiva para convertirse en activa, lo que influye en la toma de decisiones, en la construcción del conocimiento y en la percepción de la realidad. **La tecnología no solo transforma lo que hacemos, sino cómo pensamos lo que hacemos**, introduciendo nuevas formas de relación entre el individuo y el mundo.

Frente a este panorama, la IA agéntica no debe ser abordada desde una lógica de entusiasmo acrítico ni de rechazo simplista. Requiere una comprensión profunda que permita situarla dentro de marcos más amplios de análisis. **Pensar la IA agéntica es, en última instancia, pensar la forma en que organizamos la acción en sistemas complejos**. Esta reflexión no es opcional; es una condición necesaria para orientar el desarrollo de la tecnología en direcciones que sean coherentes con los valores y objetivos de la sociedad.

En última instancia, la IA agéntica nos devuelve a una pregunta fundamental: ¿qué significa actuar de manera coherente en un mundo incierto? Al estudiar sistemas que organizan su comportamiento en función de objetivos, nos vemos confrontados con nuestras propias formas de acción. **La inteligencia artificial no solo amplía nuestras capacidades; también nos obliga a interrogarnos sobre la coherencia de nuestras decisiones y la estructura de nuestras acciones**.

Esta obra no pretende cerrar el debate, sino abrirlo. La agencia artificial no es un concepto terminado, sino un campo en evolución que requiere ser explorado desde múltiples perspectivas. **Cada sistema, cada contexto y cada aplicación introduce nuevas preguntas que no pueden ser respondidas desde un único marco**. La complejidad del fenómeno exige una aproximación plural, crítica y en constante revisión.

El futuro de la IA agéntica no está escrito. Se construye en cada decisión de diseño, en cada implementación y en cada marco de gobernanza que se desarrolla. **La responsabilidad no recae en los sistemas, sino en quienes los crean, los utilizan y los regulan**. En este sentido, la agencia artificial es también un reflejo de nuestra propia agencia como sociedad.

Y así, al cerrar estas páginas, queda una idea que atraviesa toda la obra y que, más que una conclusión, constituye una invitación:

La inteligencia no reside en la capacidad de responder, sino en la capacidad de sostener coherencia en la acción. Es en esa coherencia, en esa continuidad que integra cambio y dirección, donde la agencia encuentra su forma más profunda.

REFERENCIAS

Abou Ali, M., Dornaika, F., & Charafeddine, J. (2026). Agentic AI: A comprehensive survey of architectures, applications, and future directions. *Artificial Intelligence Review,* *59*(11). https://doi.org/10.1007/s10462-025-11422-4

Acemoglu, D., & Restrepo, P. (2019). The wrong kind of AI? Artificial intelligence and the future of labor demand. *Cambridge Journal of Economics*, 44(1), 25–55. https://economics.mit.edu/sites/default/files/publications/The%20Wrong%20Kind%20of%20AI%20-%20Artificial%20Intelligence%20and.pdf

Acharya, D. B., Kuppan, K., & Divya, B. (2025). Agentic AI: Autonomous intelligence for complex goals—A comprehensive survey. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3532853

Adabara, I., Olaniyi, S. B., Nuhu, S. A., et al. (2025). Trustworthy agentic AI systems: A cross-layer review of architectures, threat models, and governance strategies for real-world deployment . *F1000Research, 14,* 905. https://doi.org/10.12688/f1000research.169927.1

Adadi, A., & Berrada, M. (2018). *Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)*. *IEEE Access, 6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Alqurni, J. (2026). Exploring the role of agentic AI in fostering self-efficacy, autonomy support, and self-learning motivation in higher education. *Frontiers in Artificial Intelligence, 9*. https://doi.org/10.3389/frai.2026.1738774

Amancharla, C., Shyamsundar, S., Radhakrishnan, S., & Gupta, U. (2025). Agentic AI: From expectations to execution—*Building autonomous, resilient and intelligent agentic AI systems.* *Infosys.* https://www.infosys.com/techcompass/documents/building-autonomous-aria-systems.pdf

Ng, A. (2018). *Machine learning yearning: Technical strategy for AI engineers in the era of deep learning* (Draft version). Deeplearning.AI. https://www.deeplearning.ai/machine-learning-yearning/

Baber, C., Kandola, P., Apperly, I., & McCormick, E. (2025). Human-centred explanations for artificial intelligence systems. Ergonomics, 68(3), 391–405. https://doi.org/10.1080/00140139.2024.2334427

Balaji, P. G., & Srinivasan, D. (2010). An introduction to multi-agent systems. En D. Srinivasan & L. C. Jain (Eds.), *Innovations in multi-agent systems and applications – 1* (Studies in Computational Intelligence, Vol. 310). Springer. https://doi.org/10.1007/978-3-642-14435-6_1

Bandi, A., Kongari, B., Naguru, R., Pasnoor, S., & Vilipala, S. V. (2025). The rise of agentic AI: A review of definitions, frameworks, architectures, applications, evaluation metrics, and challenges. *Future Internet, 17*(9), 404. https://doi.org/10.3390/fi17090404

Biswas, A., & Talukdar, W. (2025). *Building agentic AI systems: Create intelligent, autonomous AI agents that can reason, plan, and adapt*. Packt Publishing. https://www.packtpub.com/en-us/product/building-agentic-ai-systems-9781801079273?srsltid=AfmBOorKv1X1p7mA4VtmY3LmkKafSi86FjEVhA89T9deNRd5TIMANjjq

Brynjolfsson, E., Rock, D., & Syverson, C. (2021). The productivity J-curve. *American Economic Journal: Macroeconomics*, 13(1), 333–372. https://ide.mit.edu/sites/default/files/publications/2019-04JCurvebrief.final2_.pdf

Chan, A., Rahimi-Ardabilli, H., Rogers, W. A., & Coiera, E. (2025). *The real-world impact of artificial intelligence ethics frameworks across a decade in healthcare: A scoping review*. Journal of the American Medical Informatics Association, 32(11), 1767–1777. https://doi.org/10.1093/jamia/ocaf167

Chen, X., Xiang, J., Lu, S., Liu, Y., He, M., & Shi, D. (2025). Evaluating large language models and agents in healthcare: Key challenges in clinical applications. *Intelligent Medicine, 5*(2), 151–163. https://doi.org/10.1016/j.imed.2025.03.001

Cockburn, I. M., Henderson, R. M., & Stern, S. (2018). *The impact of artificial intelligence on innovation* (NBER Working Paper No. 24449). National Bureau of Economic Research. https://doi.org/10.3386/w24449

Coeckelbergh, M. (2020). *AI ethics*. MIT Press. https://doi.org/10.7551/mitpress/12549.001.0001

Collaco, B. G., Haider, S. A., Prabha, S., et al. (2026). The role of agentic artificial intelligence in healthcare: A scoping review. *Npj Digital Medicine.* https://doi.org/10.1038/s41746-026-02517-5

Dorri, A., Kanhere, S. S., & Jurdak, R. (2018). Multi-agent systems: A survey. *IEEE Access, 6*, 28573–28593. https://doi.org/10.1109/ACCESS.2018.2831228

Du, S., et al. (2026). Optimization of large language model-based agents. *ACM Computing Surveys*. https://doi.org/10.1145/3789261

Durga, R. K. (2025). *A comprehensive study of agentic AI systems*. *International Journal of Scientific Research and Modern Technology, 4*(5), 157–164. https://doi.org/10.38124/ijsrmt.v4i5.1043

European Commission. (2024). Proposal for a regulation laying down harmonised rules on artificial intelligence (AI Act). https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=OJ:L_202401689&qid=1775332109284

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.8cd550d1

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2021). Achieving a "good AI society": Comparing the aims and progress of the EU and the US. Science and Engineering Ethics, 27, 68. https://doi.org/10.1007/s11948-021-00338-3

Frank, M. R., et al. (2025). AI exposure predicts unemployment risk: A new approach to assessing the labor-market impacts of AI. *PNAS Nexus.* https://doi.org/10.1093/pnasnexus/pgaf107

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). *A survey of methods for explaining black box models*. *ACM Computing Surveys, 51*(5), Article 93, 1–42. https://doi.org/10.1145/3236009

Hahn, M., Tretter, M. & Dabrock, P.(2026). Ethical perspectives on AI Agents and Agentic AI. *AI Ethics* **6**, 218.. https://doi.org/10.1007/s43681-026-01027-0

Hofmann, P., Meierhöfer, S., Müller, L., Oberländer, A. M., & Protschky, D. (2025). *Conceptualizing the design space of artificial intelligence strategy: A taxonomy and corresponding clusters*. Business & Information Systems Engineering. https://doi.org/10.1007/s12599-025-00856-1

Hosseini, S., & Seilani, H. (2025). The role of agentic AI in shaping a smart future: A systematic review. *Array.* https://doi.org/10.1016/j.array.2025.100399

Jobin, A., Ienca, M., & Vayena, E. (2020). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389–399. https://www.researchgate.net/publication/335579286_The_global_landscape_of_AI_ethics_guidelines/link/647a0becd702370600cc5828/download?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19

Jöhnk, J., Weißert, M. & Wyrtki, K. Ready or Not, AI Comes (2021)— An Interview Study of Organizational AI Readiness Factors. *Bus Inf Syst Eng* **63**, 5–20https://doi.org/10.1007/s12599-020-00676-7

Kadir, M. A., Mosavi, A., & Sonntag, D. (2025). *Evaluation metrics for XAI: A review, taxonomy, and practical applications*. German Research Center for Artificial Intelligence https://www.dfki.de/fileadmin/user_upload/import/14708_XAI_Evaluation_Metrics__Taxonomies__Concepts_and_Applications__INES_2023_-7.pdf

Karnouskos, S., Leitão, P., Ribeiro, L., & Colombo, A. W. (2020). Industrial agents as a key enabler for realizing industrial cyber-physical systems: Multiagent systems entering Industry 4.0. *IEEE Industrial Electronics Magazine, 14*(3), 18–32. https://doi.org/10.1109/MIE.2019.2962225

Leo, M., Tan, F., Miao, T. *et al.* From threat to trust: assessing security risks of agentic AI systems. *Int. J. Inf. Secur.* **25**, 23 (2026). https://doi.org/10.1007/s10207-025-01185-y

Lewis, P. R., & Sarkadi, Ş. (2024). *Reflective artificial intelligence*. Minds and Machines, 34, 14. https://doi.org/10.1007/s11023-024-09620-3

Li, H. (2026) General Framework of AI Agents. *J. Comput. Sci. Technol.* https://doi.org/10.1007/s11390-025-5951-5

Maldonado, D., Cruz, E., Abad Torres, J., Cruz, P. J., & Gamboa, S. (2024). *Multi-agent systems: A survey about its components, framework and workflow*. IEEE Access. Advance online publication. https://doi.org/10.1109/ACCESS.2024.3409051

Mantia, L., Chatterjee, S., & Lee, V. S. (2025). Designing a successful agentic AI system. *Harvard Business Review*. https://hbr.org/2025/10/designing-a-successful-agentic-ai-system

Mejía Trejo, J. (2024). *Inteligencia artificial: Fundamentos de ingeniería de prompts con ChatGPT como innovación impulsora de la creatividad*. Editorial: Academia Mexicana de Investigación y Docencia en Innovación (AMIDI). https://amidibiblioteca.amidi.mx/index.php/AB/catalog/book/48

Mejía Trejo, J. (2025). *Inteligencia artificial y su repercusión en la educación superior*. Universidad de Guadalajara (CUCEA) y Academia Mexicana de Investigación y Docencia en Innovación (AMIDI). https://amidibiblioteca.amidi.mx/index.php/AB/catalog/book/70

Miller, S. M. and Davenport, T.H.. AI and the future of work: What we know today. (2021). *Gradient. 1*-13.: https://ink.library.smu.edu.sg/sis_research/6657

Nisa, U., Shirazi, M., Saip, M. A., & Mohd Pozi, M. S. (2026). *Agentic AI: The age of reasoning—A review*. *Journal of Automation and Intelligence, 5*(1), 69–89.

OpenAI. (SF). A practical guide to building agents. https://cdn.openai.com/business-guides-and-resources/a-practical-guide-to-building-agents.pdf

Organisation for Economic Co-operation and Development (OECD). (2022). *OECD framework for the classification of AI systems* (OECD Digital Economy Papers No. 323). OECD Publishing. https://doi.org/10.1787/cb6d9eca-en

Organisation for Economic Co-operation and Development (OECD). (2025ª). *Governing with Artificial Intelligence*. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/06/governing-with-artificial-intelligence_398fa287/795de142-en.pdf

Organisation for Economic Co-operation and Development (OECD). (2025b).*Artificial intelligence and competitive dynamics in downstream markets* https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/11/artificial-intelligence-and-competitive-dynamics-in-downstream-markets_c6e81d0e/ccf0624a-en.pdf

Organisation for Economic Co-operation and Development (OECD).(2025c) AI Observatory. *AI governance through global red lines*. https://oecd.ai/en/wonk/ai-governance-through-global-red-lines-can-help-prevent-unacceptable-risks

Organisation for Economic Co-operation and Development (OECD). (2026). *The agentic AI landscape and its conceptual foundations*. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2026/02/the-agentic-ai-landscape-and-its-conceptual-foundations_a9d4b451/396cf758-en.pdf

Ozdemir, S. (2026). *Building agentic AI: Workflows, fine-tuning, optimization, and deployment*. Pearson. https://www.pearson.com/en-us/subject-catalog/p/building-agentic-ai-workflows-fine-tuning-optimization-and-deployment/P200000014597/9780135489680?srsltid=AfmBOorDctf14QiZm9bz5eDjHPreQZeExs1DYPiHOeKV9b1HhIZJoqv7

Parker, S. K., & Grote, G. (2022). Automation, algorithms, and beyond: Why work design matters more than ever. *Applied Psychology.* https://doi.org/10.1111/apps.12241

Piccialli, F., et al. (2025). AgentAI: Autonomous agents in Industry 4.0. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2025.128404

Poole, D. L., & Mackworth, A. K. (2017). *Artificial intelligence: Foundations of computational agents* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781108164085

Rahwan, I., et al. (2019). Machine behaviour. *Nature*, 568, 477–486. https://doi.org/10.1038/s41586-019-1138-y

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *The Academy of Management Review, 46*(1), 192–210. https://doi.org/10.5465/amr.2018.0072

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *"Why should I trust you?": Explaining the predictions of any classifier*. En Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16) (pp. 1135–1144). ACM. https://doi.org/10.1145/2939672.29397

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking. https://people.eecs.berkeley.edu/~russell/papers/mi19book-hcai.pdf

Russell, S., & Norvig, P. (2022). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson. https://aima.cs.berkeley.edu/

Sapkota, R., Roumeliotis, K. I., & Karkee, M. (2026). *AI agents vs. agentic AI: A conceptual taxonomy, applications and challenges*. **Information Fusion, 126**(Part B), 103599. https://doi.org/10.1016/j.inffus.2025.103599

Sawant, P. D. (2025). *Agentic AI: A quantitative analysis of performance and applications*. **Journal of Advances in Artificial Intelligence, 3**(2). https://www.jaai.net/vol3/JAAI-V3N2-41.pdf

Sayyad, N., Dave, H., Kathane, M., & Patel, R. (2025, October). Agentic AI systems: A review of architectures, autonomy, and ethical implications*. International Journal of Environmental Sciences*. https://theaspd.com/index.php/ijes/article/view/10841

Stahl, B. C.(2021). *Artificial Intelligence for a Better Future An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies.* Springer Nature https://link.springer.com/book/10.1007/978-3-030-69978-9

Stone, P. et al. (2021). *Artificial intelligence and life in 2030.* Standford University. https://ai100.stanford.edu/sites/g/files/sbiybj18871/files/media/file/AI100Report_MT_10.pdf

Te'eni, D., Yahav, I., & Schwartz, D. (2026). *What it takes to control AI by design: Human learning*. AI & Society, 41, 237–250. https://doi.org/10.1007/s00146-025-01979-0

United Nations. (2024). *Governing AI for Humanity*. https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf

UNESCO. (2025ª). *Artificial intelligence and education: Preserving human agency in a world of automation*. https://unesdoc.unesco.org/ark:/48223/pf0000376709

UNESCO. (2025b). *Artificial intelligence and culture: Report of the independent expert group*. https://www.unesco.org/sites/default/files/medias/fichiers/2025/09/CULTAI_Report%20of%20the%20Independent%20Expert%20Group%20on%20Artificial%20Intelligence%20and%20Culture%20%28final%20online%20version%29%201.pdf

Vinuesa, R., Azizpour, H., Leite, I. *et al.* (2020).The role of artificial intelligence in achieving the Sustainable Development Goals. *Nat Commun* **11**, 233 (https://doi.org/10.1038/s41467-019-14108-y

Wang, Z., Huang, C., & Yao, X. (2024). *A roadmap of explainable artificial intelligence: Explain to whom, when, what and how?* ACM Transactions on Autonomous and Adaptive Systems, 19(4), Article 20, 1–40. https://doi.org/10.1145/3702004

Wang, L., Ma, C., Feng, X. *et al.* (2024). A survey on large language model based autonomous agents. *Front. Comput. Sci. 18,* 186345- https://doi.org/10.1007/s11704-024-40231-1

Wang, S. H. (2025). *A review of agent data evaluation: Status, challenges, and future prospects as of 2025*. *Journal of Software Engineering and Applications, 18*(9), 358–372. https://doi.org/10.4236/jsea.2025.189021

World Bank. (2026). *World development report 2026: Artificial intelligence for development*. https://thedocs.worldbank.org/en/doc/1e4e52502104a331fb42cba0d4afa995-0050062026/original/WDR2026-Concept-Note.pdf

World Economic Forum.(2024). *Navigating the AI frontier: A primer on the evolution and impact of AI agents*. https://reports.weforum.org/docs/WEF_Navigating_the_AI_Frontier_2024.pdf

World Economic Forum. (2025). *AI agents in action: Foundations for evaluation and governance*. https://reports.weforum.org/docs/WEF_AI_Agents_in_Action_Foundations_for_Evaluation_and_Governance_2025.pdf

World Health Organization. (2021). Ethics and governance of artificial intelligence for health. https://iris.who.int/server/api/core/bitstreams/f780d926-4ae3-42ce-a6d6-e898a5562621/content

Yadav, P., Mishra, A., & Kim, S. (2023). *A comprehensive survey on multi-agent reinforcement learning for connected and automated vehicles*. Sensors, 23(10), 4710. https://doi.org/10.3390/s23104710

Yuan, Y., & Xie, T. (2026). Reinforce LLM reasoning through multi-agent reflection. *Proceedings of the 42nd International Conference on Machine Learning (ICML 2025)*, Article 2952, 73701–73731. https://proceedings.mlr.press/v267/yuan25l.html

Zhang, Z., Dai, Q., Bo, X., Ma, C., Li, R., Chen, X., Zhu, J., Dong, Z., & Wen, J.-R. (2025). *A survey on the memory mechanism of large language model-based agents*. ACM Transactions on Information Systems, 43(6), Article 155, 1–47. https://doi.org/10.1145/3748302

Xie, H., Mei, Q., & Chui, Y. H. (2025). AI applications for structural design automation. *Automation in Construction, 179*, 106496. https://doi.org/10.1016/j.autcon.2025.106496

inteligencia artificial agéntica : principios y alcances

Maquetación, diseño y distribución digital:
Academia Mexicana de Investigación y Docencia en Innovación (AMIDI).
Responsable del registro DOI, la gestión de metadatos y la publicación en AMIDI.Biblioteca.

Av.Paseo de los Virreyes 920,
Col. Virreyes Residencial
C.P. 45110
Zapopan, Jalisco, México

eBOOK Hecho y editado en México / Made and edited in Mexico
Se terminó de editar en **Abril de 2026.**

La obra Inteligencia artificial agéntica: principios y alcances examina la transformación de la inteligencia artificial desde enfoques centrados en la generación de resultados hacia un paradigma orientado a la organización del comportamiento autónomo en el tiempo. Frente al predominio de estudios sobre IA generativa, el libro propone un marco innovador que integra percepción, decisión y acción dentro de una lógica operativa coherente, continua y adaptativa.

Su contribución principal radica en una reconfiguración epistemológica de la inteligencia artificial, entendida como la capacidad de sostener comportamiento estructurado en contextos complejos. A partir de ello, desarrolla un enfoque conceptual, estructural y operativo que permite comprender, diseñar, implementar y evaluar sistemas agénticos con rigor.

Dirigida a investigadores, estudiantes de posgrado, profesionales y diseñadores de sistemas, la obra articula fundamentos teóricos, principios de diseño, criterios de evaluación y análisis de impacto, incorporando además reflexiones sobre gobernanza y futuro. En conjunto, ofrece un marco integral para analizar y orientar el desarrollo de sistemas inteligentes autónomos en escenarios contemporáneos caracterizados por alta complejidad y creciente automatización.

Zapopan, Jal. a 30 de Enero de 2026

Dictamen de Obra. AMIDI.DO.20260130

Los miembros del equipo editorial de la Academia Mexicana de Investigación y Docencia en Innovación (**AMIDI**), **RENIECYT-SECIHTI 2200092**, ver:
https://www.amidibiblioteca.amidi.mx/index.php/AB/about/editorialTeam
se reunieron para atender la invitación a dictaminar el libro:

inteligencia artificial agéntica. Principios y alcance

Siendo los siguientes participantes de la misma:

Nombre completo	Rol
Juan Mejía Trejo	Autor

Dicho documento fue sometido al proceso de evaluación por pares doble ciego, de acuerdo a la política de la editorial, para su dictaminación de aceptación, ver: https://www.amidibiblioteca.amidi.mx/index.php/AB/procesodeevaluacionporparesenciego

Los miembros del equipo editorial se reúnen con el curador principal del repositorio digital para convocar:

1. Que el comité científico, de forma colegiada, revise los contenidos y proponga a los pares evaluadores que colaboran dentro del comité de redacción, tomando en cuenta su especialidad, pertinencia, argumentos, enfoque de los capítulos al tema central del libro, entre otros.
2. Se invita a los pares evaluadores a participar, formalizando su colaboración.
3. Se envía así, el formato de evaluación para inicio del proceso de evaluación doble ciego a los evaluadores elegidos de la mencionada obra.
4. El comité científico recibe las evaluaciones de los pares evaluadores e informa a el/los autor/(es) los resultados a fin de que se atiendan las observaciones, el requerimiento de reducción de similitudes, y recomendaciones de mejora a la obra.
5. La obra evaluada, consta de:

Contenido
INTRODUCCIÓN
CAPÍTULO 1.Evolución y fundamentos conceptuales de la IA agéntica
Naturaleza de la agencia en sistemas inteligentes
Naturaleza funcional de la agencia
Dinámica operativa de la agencia
Delimitación conceptual de la agencia
Emergencia histórica del paradigma agéntico
Antecedentes del paradigma agéntico
Transición hacia el paradigma
Consolidación del paradigma agéntico

Av. Lázaro Cárdenas 3454 int. 6,
Col. Jardines de los Arcos, C.P. 44500,
Guadalajara, Jalisco, México
Tel. Oficina. 33 3560 7860/ Cel. 3326264422
editorial@scientiaetpraxis.amidi.mx

Reconfiguración de los mercados bajo IA agéntica
Competitividad basada en capacidades agénticas
Transformación de la estructura económica en sistemas agénticos
Transformación organizacional
Reconfiguración organizacional bajo IA agéntica
Capacidades organizacionales en entornos agénticos
Implicaciones estructurales de la IA agéntica en la organización
Gobernanza y regulación
Gobernanza de sistemas agénticos: principios y fundamentos
Regulación de la IA agéntica: marcos, instrumentos y dinámicas operativas
Implicaciones estructurales de la gobernanza y regulación en sistemas agénticos
Futuro de la IA agéntica
Trayectorias evolutivas de la IA agéntica
Dinámicas de adopción y transformación sistémica de la IA agéntica
Escenarios prospectivos y riesgos estructurales de la IA agéntica
Conclusiones

CAPÍTULO 7. Reflexión Final
REFERENCIAS

6. Una vez emitidas las observaciones, el requerimiento de reducción de similitudes, y recomendaciones de mejora a la obra por los evaluadores y todas ellas resueltas por el/los autor/(es), el resultado resalta que el contenido del libro:

    a. Reúne los elementos teóricos actualizados y prácticos desglosados en cada uno de sus capítulos.
    b. Los capítulos contenidos en la obra, muestran claridad en el dominio del tema, congruencia con el título central del libro, y una estructura consistente
    c. Se concluye finalmente, que la obra dictaminada, puede fungir como libro de texto principal o de apoyo tanto para estudiantes de licenciatura como de posgrados, así como público en general interesado.

7. De esta forma, el resultado del dictamen de aceptación de la obra fue:

FAVORABLE PARA SU PUBLICACIÓN

Sirva la presente para los fines que al interesado convengan.

Atentamente

Dr. Carlos Omar Aguilar Navarro
Curador AMIDI.Biblioteca
AMIDI

Av. Lázaro Cárdenas 3454 int. 6,
Col. Jardines de los Arcos, C.P. 44500,
Guadalajara, Jalisco, México
Tel. Oficina. 33 3560 7860/ Cel. 3326264422
editorial@scientiaetpraxis.amidi.mx