UN ESTUDIO DE

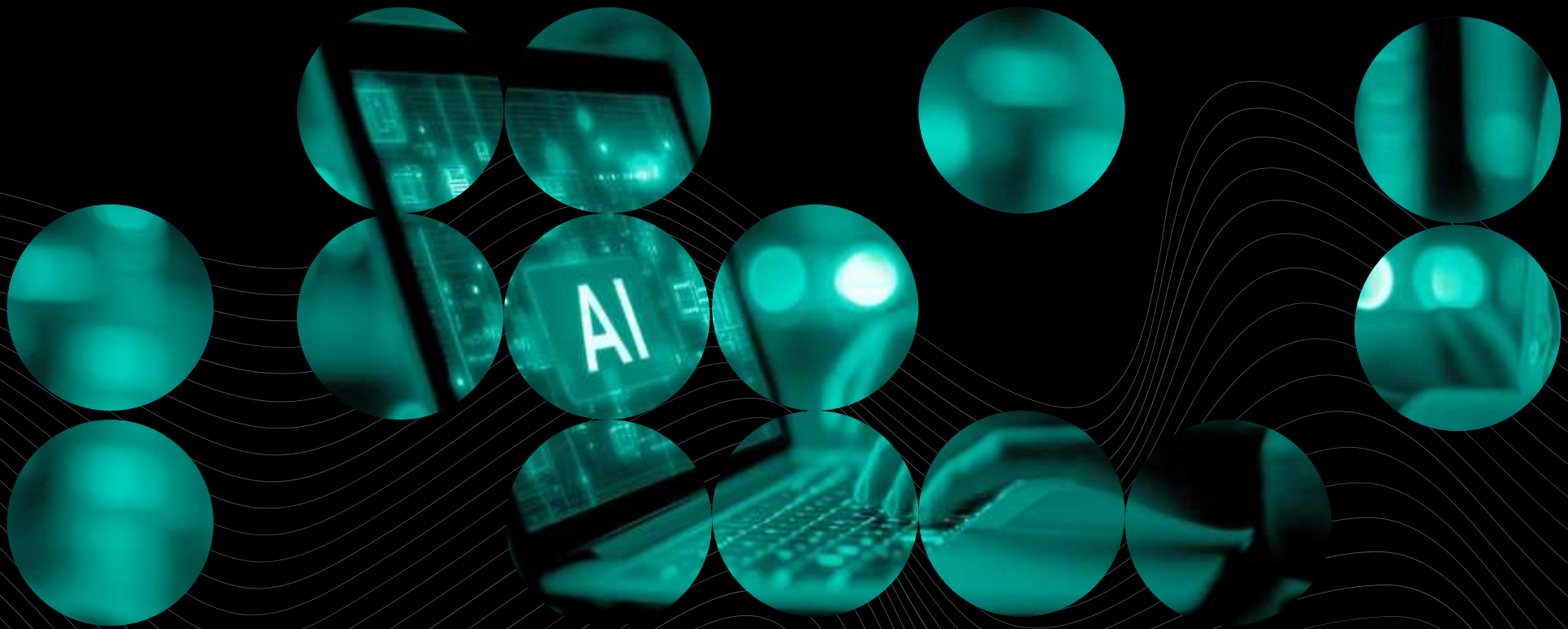
endeavor

CON EL APOYO DE

ORACLE ACID LABS

Agentic AI en el ecosistema emprendedor chileno:

CONDICIONES, PATRONES Y VALOR



Índice

01 **Qué es Agentic AI: Definición operacional y mapa de madurez** 5

02 **Dónde se está capturando valor en el mundo: Panorama global de adopción** 8

Quién está adoptando y con qué profundidad 9

Dónde está la oportunidad dentro del Stack 10

Cómo medir el retorno real 12

Cómo saber si el agente está funcionando bien 13

Quién responde cuando el agente falla 15

03 **Lo que tienen en común los que escalaron** 17

El patrón de origen 19

La decisión que separa el piloto de la escala 20

Cómo justificaron la inversión 21

04 **Qué debe tener la organización antes de empezar** 23

05 **Por dónde partir y en qué orden** 28

06 **Conclusiones** 36

07 **Metodología** 39

08 **Bibliografía** 40

09 **Sobre este estudio** 43

La perspectiva del partner: Oracle

“Los proyectos que son exitosos son los que parten de una estrategia de IA alineada al negocio, donde los agentes son una consecuencia de esa estrategia y no iniciativas aisladas. Las organizaciones que generan mayor valor comienzan identificando procesos de alto impacto, definen métricas de negocio claras y diseñan agentes que transforman esos procesos.

Hay otro factor determinante: no existen agentes inteligentes sin datos inteligentes Los datos son el verdadero activo estratégico. El éxito de iniciativas de agentes depende de la calidad, integración y gobernanza de los datos que proporcionan el contexto para que los agentes tomen decisiones precisas y generen valor para el negocio.”

GLORIA VAILATI
Executive AI Advisor de Oracle South America

La perspectiva del partner: Acid Labs

Dejamos de mirar Agentic AI como una moda. Antes de escalar, ordenamos las bases: gobernanza de datos, seguridad, arquitectura, criterios de uso, ownership y una forma concreta de medir valor. Pasamos de experimentos aislados a una capacidad organizacional trazable, responsable y capaz de escalar. Convertimos nuestra diversidad de industrias, experiencias y talento en frameworks, buenas prácticas y componentes reutilizables que hoy se traducen en nuevos modelos de servicio.

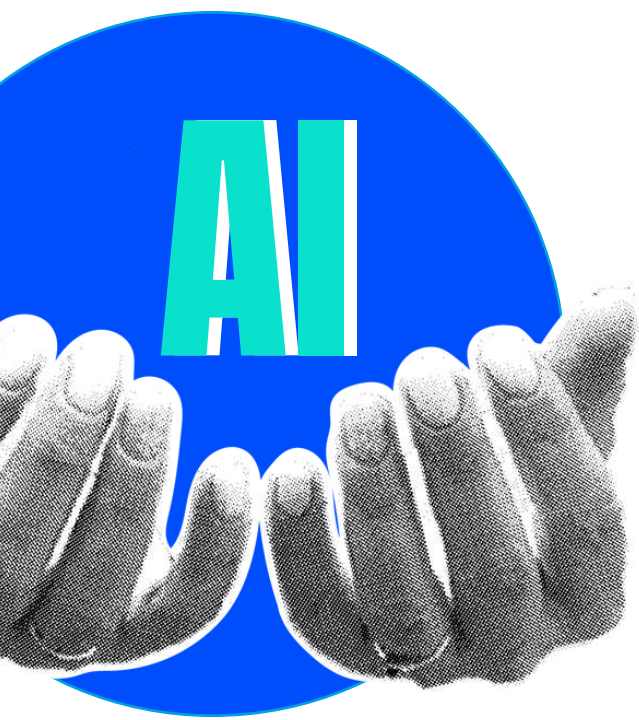
Lo que hemos aprendido en el camino: si los datos no están gobernados y los procesos no son claros, Agentic AI no ordena, amplifica el desorden. El cuello de botella ya no es la tecnología ni el presupuesto: es cuántos procesos tienes realmente listos para delegarle a un agente.Y una apuesta hacia adelante: hoy la conversación es cómo construir agentes internos. La que viene es cómo tu empresa va a interactuar con proveedores agénticos. ¿Vamos a pagar en automático también en el enterprise? La infraestructura de confianza para ese mundo, identidad, límites, auditoría entre empresas, se empieza a construir hoy."

GERT FINDEL
Socio & co-CEO de Acid Labs

01

Qué es Agentic AI

DEFINICIÓN OPERACIONAL
Y MAPA DE MADUREZ



Hoy todo el ecosistema habla de Agentic AI, pero no hay una visión compartida de qué significa. Para algunos un agente es cualquier chatbot que se conecta a una API[1] externa, mientras que para otros es un sistema que logra ejecutar flujos de trabajo complejos casi sin intervención humana. Esta ambigüedad no es menor, ya que puede generar expectativas desalineadas y confusión tanto entre los founders que arman el proyecto como entre los inversionistas que lo financian.

Anthropic define Agentic AI como sistemas donde los LLMs[2] dirigen dinámicamente sus propios procesos y uso de herramientas, manteniendo el control total sobre cómo cumplen las tareas (Anthropic, 2024). OpenAI lo define como sistemas que son capaces de perseguir objetivos complejos con una supervisión directa limitada, y su enfoque se centra en el concepto de "Agenticness", donde se consideran un espectro de 4 dimensiones clave: Complejidad del objetivo, Complejidad del entorno, Adaptabilidad y Ejecución independiente (OpenAI, 2023). McKinsey describe esta tecnología como una evolución clave: los modelos dejan de ser solo para creación de contenidos y pasan a ejecutar tareas complejas de forma autónoma (McKinsey, 2025a). Finalmente A16z, en la voz de una de sus socias lo define como un LLM de múltiples pasos con capacidad de razonamiento y un árbol de decisiones dinámico, es decir, debe ser capaz de decidir sobre la tarea en cuestión y actuar de forma autónoma (Li, citada en Bort, 2025).

Con énfasis distintos, las cuatro definiciones apuntan al mismo lugar: **autonomía en ejecución, capacidad de actuar sobre sistemas reales y orientación hacia objetivos complejos que requieren múltiples pasos**. Esa base común es la que sostiene la definición que usa este estudio.

Para efectos de este estudio, se entenderá por Agentic AI todo sistema de Inteligencia Artificial capaz de perseguir objetivos específicos de forma autónoma, planificando y ejecutando secuencias de acciones sobre sistemas y entornos reales, **sin requerir intervención humana en cada paso**. La diferencia fundamental con un chatbot tradicional radica en la **acción**. El valor no está en el modelo base, sino en su capacidad de ejecutar tareas en lugar de limitarse a responder preguntas. Esta autonomía opera siempre dentro de un marco de supervisión humana, selectiva pero deliberada, que define los límites de acción del agente, los puntos de control y mantiene el accountability sobre las decisiones sensibles.

1 Una Application Programming Interface, o API, son mecanismos que permiten a dos sistemas de software conectarse y comunicarse entre sí mediante un conjunto de protocolos establecidos.

2 LLM o Large Language Models son modelos de aprendizaje profundo que se pre entrenan con gran cantidad de datos.

La definición anterior no es suficiente para tomar decisiones en los distintos procesos de integración. La pregunta que se vuelve relevante para cualquier liderazgo en una empresa u organización que está adoptando estas tecnologías es: **¿en qué punto del proceso se encuentra mi organización y qué hace falta para avanzar?** Este reporte adapta el modelo de fases de adopción de McKinsey (2025d), que segmenta a las empresas en fases de experimentación, pilotaje y escalamiento, y documenta que cerca de dos tercios aún no logra escalar la IA. Capgemini observa algo similar: la mayoría de los agentes no supera la fase de piloto (Capgemini, 2026). Sobre esa base construimos un mapa de tres etapas calibrado al contexto de startups y scaleups, con preguntas diagnósticas para que cada organización identifique dónde está situada.

La utilidad de este mapa es práctica: permite identificar desde qué punto está partiendo una empresa y qué condiciones necesita para que la integración de Agentic AI tenga impacto real en su operación. Desde Endeavor observamos que muchos emprendimientos, PYMEs, startups y scaleups operan en fase de exploración con la convicción de estar piloteando. **Esta brecha de percepción es uno de los principales obstáculos al momento de una integración eficaz.**

FASE	CARACTERÍSTICA CENTRAL	PREGUNTA DIAGNÓSTICA
1. Exploración	Experimentos fragmentados con IA, **sin estrategia formal ni impacto en procesos centrales.** Los agentes operan en **tareas acotadas y repetitivas** dentro de procesos predefinidos.	¿Estamos usando IA de forma aislada, sin que impacte ningún proceso central del negocio? ¿Las iniciativas dependen de personas específicas o hay una visión compartida en el equipo? ¿Tenemos claridad sobre qué problema concreto queremos resolver con IA?
2. Piloto	Agentes desplegados en **funciones específicas con supervisión humana activa y métricas de éxito definidas**. Hay integración parcial con sistemas existentes y gobernanza incipiente.	¿Tenemos al menos un agente corriendo en producción con criterios claros de éxito? ¿Hay una persona responsable de los resultados que genera ese agente? ¿Estamos midiendo su impacto en términos operativos o financieros, aunque sea a nivel de función?
3. Escala	**Agentes integrados en procesos críticos** del negocio, operando en múltiples funciones **con impacto financiero medible**. La organización rediseña flujos de trabajo en torno a capacidades agénticas.	¿Los agentes forman parte de cómo opera el negocio hoy, con KPIs que lo demuestran? ¿La integración de IA implicó rediseñar procesos o solo automatizar lo que ya existía? ¿La dirección de la empresa toma decisiones estratégicas considerando las capacidades agénticas disponibles?

FIGURA 1. Matriz de supervisión humana en Agentic AI. Elaboración propia a partir de OpenAI (2025), World Economic Forum (2025) y McKinsey (2025b).

02

Dónde se está capturando valor en el mundo:

PANORAMA GLOBAL
DE ADOPCIÓN

2.1

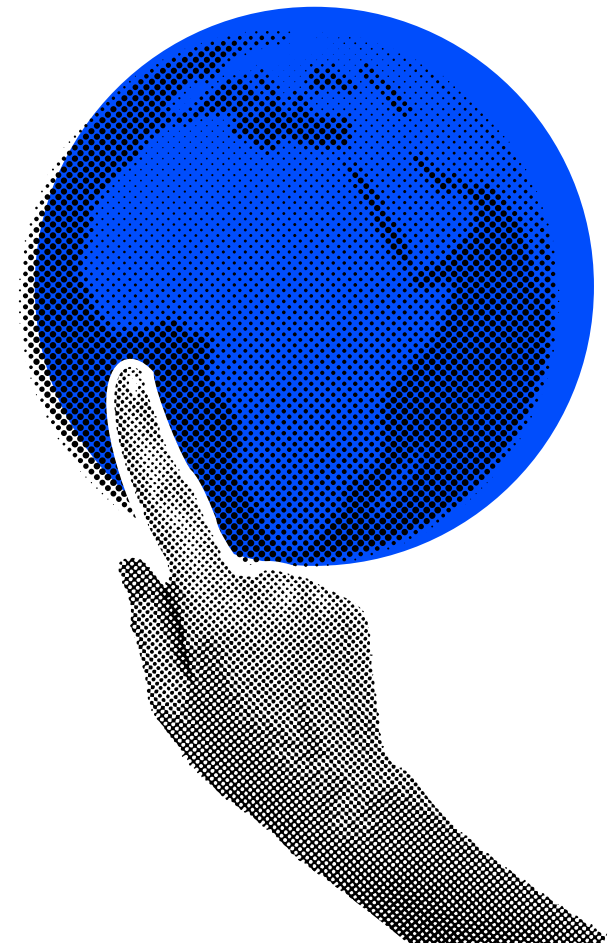
Quién está adoptando y con qué profundidad

En 2025, **el 88% de las empresas afirma haber integrado IA** en alguna función de negocio (McKinsey, 2025d). Sin embargo, la masificación no implica profundidad. La proporción de empresas con implementaciones de Agentic AI en fase de escalamiento sigue siendo acotada. En los sectores que van más avanzados, como la industria tecnológica, las áreas de mayor adopción son Ingeniería de software con 24%, TI con 22% y Desarrollo de producto con 18%. Por debajo de estos niveles se encuentra el resto de las industrias (McKinsey, 2026).

Ransbotham et al. refuerzan ese diagnóstico y agregan una dimensión que los datos de adopción no capturan: **muchas organizaciones incorporan agentes antes de definir una estrategia clara de para qué**. La velocidad de adopción supera la capacidad de definir el problema a resolver, lo que ubica el riesgo principal de Agentic AI en el plano organizacional más que en el tecnológico (Ransbotham et al., 2025). La **ventaja competitiva** no la tendrá quien implemente más rápido, sino quien reorganice sus procesos y estructura en torno a las capacidades agénticas disponibles.

Ese patrón de adopción inicial se repite al analizar la distribución por industria. Se observa una alta concentración de uso de Agentic AI en solo tres sectores: **servicios financieros con el 30%, retail con el 21% y manufactura con el 18%.** Más atrás aparecen viajes y transporte con el 11% y salud y farmacéutica con el 10% (ISG, 2025). El denominador común de los sectores líderes es la disponibilidad de datos estructurados en volumen y la presencia de procesos transaccionales, dos condiciones que aparecen de forma consistente en los sectores con mayor tracción.

En servicios financieros, la riqueza de datos estructurados y no estructurados es la base que hace viable la adopción agéntica en procesos críticos. Los casos de uso más activos son detección de fraudes, interacción con clientes y servicios financieros autónomos. En retail, el foco está en la interacción personalizada con el cliente, incluyendo ventas, upselling y atención, junto con la mejora en las búsquedas dentro de canales digitales. En el sector de la manufactura, el mantenimiento concentra el 42% de los casos de uso del sector, seguido por gestión autónoma de inventarios y monitoreo de seguridad operacional (ISG, 2025).



2.2

Dónde está la oportunidad dentro del Stack

Entender qué industrias lideran la adopción responde al "dónde". La pregunta que sigue es el "cómo": en qué capa del stack tecnológico se está generando valor real y dónde tiene sentido posicionarse. Para efectos de este estudio, se entenderá por **stack tecnológico de Agentic AI** el conjunto de capas que componen su arquitectura: infraestructura (plataformas cloud y fabricantes de hardware), modelos fundacionales que sustentan los productos de IA, y aplicaciones que integran esos modelos en soluciones orientadas a resolver tareas específicas de negocio (Andreessen Horowitz, 2023; Bessemer Venture Partners, 2025).

A pesar del dinamismo del mercado, la distribución del valor entre las distintas capas del stack tecnológico no es uniforme. Los proveedores de infraestructura siguen capturando la mayor parte del flujo de inversión, impulsados por la demanda de capacidad de cómputo. La capa de modelos se consolida en torno a un número reducido de actores respaldados por alto flujo de capital. La capa de aplicaciones, en cambio, muestra dos escenarios. Por un lado las aplicaciones genéricas, que tienen baja retención y sus márgenes son muy bajos, y por otro aplicaciones con alto nivel de integración de flujos de trabajo y datos propietarios, que demuestran solidez en su mercado. La evidencia disponible indica que el 74% del valor económico de la IA lo captura apenas el 20% de las organizaciones, lo que sugiere que aún estamos insertos en una dinámica de concentración más que de difusión (PwC, 2026; Bessemer Venture Partners, 2025).

Este desequilibrio responde a razones de mercado. La capa de modelos se está consolidando con alianzas estratégicas respaldadas por alto acceso a capital (Microsoft/OpenAI, AWS/Anthropic, Google/DeepMind). Al mismo tiempo, la competencia ha generado que el valor del token se desplome. El precio de GPT-4o es cerca de 6 veces menor que el de GPT-4 en su lanzamiento (OpenAI, 2026a, OpenAI, 2026b). Para cualquier startup o scaleup, competir en infraestructura o en el desarrollo de modelos propios resulta inviable frente a actores con ventajas de escala y capital que ninguna empresa emergente puede replicar.

La oportunidad comercial real está en la **capa de aplicación**, pero con matices importantes, dado que no toda aplicación construida sobre modelos es capaz de capturar valor de manera sostenible. Esto se muestra en los datos del mercado: El SaaS horizontal (aquel software diseñado para ser útil en varias industrias) cayó en un 35% en sus múltiplos de valoración en los últimos doce meses, mientras que el SaaS vertical (usado en una vertical o industria) subió un 3%, lo que demuestra un cambio en cómo el mercado valora cada tipo de solución (Redpoint Ventures, 2026). El análisis de **ventajas competitivas** (o moats) de Morningstar sobre 132 empresas refuerza ese diagnóstico, la IA erosiona la ventaja competitiva donde dependía de la fricción del flujo de trabajo o del volumen de usuarios como barrera de salida. Cuando las herramientas de IA reducen esta fricción, la ventaja competitiva desaparece, mientras que aplicaciones con datos propietarios o flujos de trabajo especializados no sufrieron esta pérdida, incluso en algunos casos la amplificó (Morningstar, 2026). Las aplicaciones que sí sostienen su ventaja competitiva son las que logran integrar el modelo con los datos y procesos propios de la organización, y operan como capa de ejecución dentro de un flujo de negocio específico. Bessemer lo describe como el paso de "sistemas de registro a sistemas de acción" (Bessemer Venture Partners, 2025).

Para founders de startups, scaleups, PYMEs y CTOs, la ventaja competitiva no depende del modelo que se elija sino de la **profundidad con que se integre en los procesos y datos propios de la organización.** McKinsey identifica el rediseño de flujos de trabajo como la variable que separa a las organizaciones que extraen valor real de las que sólo añaden una capa de IA sobre lo que ya existía (McKinsey, 2025d). El análisis de gasto en aplicaciones de IA de Andreessen Horowitz confirma ese patrón en la práctica: las empresas con mayor disposición a pagar son las que utilizan aplicaciones verticales con integración profunda, no asistentes genéricos (Andreessen Horowitz, 2025). Mientras Capgemini proyecta que los agentes de IA generarán hasta US$450 mil millones en valor económico hacia 2028 principalmente a través de eficiencia corporativa, la conformación del sector indica que ese valor lo logrará capturar quienes hayan convertido el Agentic AI en una ventaja operativa real, no quienes sólo hayan adoptado tecnología (Capgemini, 2025).

Esta diferencia entre la simple adopción y la integración profunda tiene un impacto financiero directo:

“Las empresas con mayor madurez operativa en IA logran retornos 7,2 veces superiores al promedio, capturando casi tres cuartas partes del valor total generado por esta tecnología.”

(PwC, 2026).

2.3

Cómo medir el retorno real

Capturar ese valor requiere algo más que elegir la capa correcta del stack tecnológico. Requiere saber si la integración está generando retorno real, y esa es precisamente la pregunta que separa a las organizaciones que escalan de las que pilotean sin claridad del impacto real que tienen.

Muchas empresas enfrentan hoy la llamada "Paradoja de la IA Generativa": la tecnología se despliega por toda la organización, pero los reportes financieros aún no reflejan impacto real en los márgenes. Esta trampa ocurre cuando la adopción se mide como un fin en sí mismo. Indicadores como la cantidad de usuarios activos o consultas enviadas suelen considerarse como éxitos aislados, sin conectarse con resultados operativos o financieros. McKinsey plantea que el valor del Agentic AI está en el flujo de trabajo que ejecuta y reinventa, no en el modelo que lo sustenta. La pregunta a responder no es cuántos empleados usan el agente, sino cuánto más rápido, barato o confiable se transforma el proceso donde opera (McKinsey, 2025d).

Antes de definir el marco de medición, la organización debe establecer una línea base del proceso donde se utilizará Agentic AI. Esto implica documentar con precisión el estado actual: tiempo de ciclo, tasa de error, costo por transacción y capacidad del equipo. Sin esa fotografía inicial, cualquier cifra de mejora posterior carece de respaldo. McKinsey lo señala como condición necesaria: no se puede medir el impacto de un agente sobre un proceso que no estaba medido antes de su despliegue (McKinsey, 2025d).

Con la línea base establecida, el retorno real se mide en tres dimensiones que deben operar de forma simultánea. La primera es el ahorro en costos, expresado en horas de trabajo liberadas y reducción del costo por transacción. La segunda es el impacto en ingresos, medido como nuevas oportunidades comerciales habilitadas o mejoras en tasas de conversión atribuibles al agente. La tercera es la reducción de riesgos operativos, calculada a partir de errores prevenidos y costos de cumplimiento evitados. PwC identifica que el 20% de las empresas captura el 74% del valor económico generado por IA, y lo que las diferencia del resto es que miden el retorno en estas tres dimensiones de forma simultánea, no como métricas aisladas (PwC, 2026).

La distinción entre indicadores predictivos e indicadores confirmatorios resulta útil al momento de gestionar la implementación de Agentic AI. Los indicadores predictivos (tasa de éxito de tareas ejecutadas o NPS), tal como lo indica su nombre, anticipan tendencias de uso y adopción de los sistemas antes de que el impacto aparezca en los estados financieros de la empresa. Los confirmatorios (como el costo por transacción o aumento de ingresos atribuibles a la IA) confirman si el impacto es real. Para una implementación exitosa ambos indicadores son necesarios, los primeros para corregir los desvíos con antelación y los segundos para justificar la inversión (McKinsey, 2025d; PwC, 2026).

A partir de estas dimensiones, el ROI total de una implementación de Agentic AI puede calcularse con la siguiente fórmula:

ROI Total = (Ahorros Anuales + Nuevos Ingresos Anuales + Ahorros de Riesgos Anuales – Costo Anual de IA) / Costo Anual de IA

El objetivo de este proceso no es obtener una métrica absoluta, sino contar con un criterio común que permita comparar proyectos, inversiones y decidir cuándo es prudente que el piloto pueda escalar.

2.4

Cómo saber si el agente está funcionando bien

El marco de medición comentado anteriormente responde a la pregunta de si la implementación de Agentic AI en la organización está generando valor. Sin embargo, hay una pregunta previa que es necesario hacerse: ¿Cómo saber si el agente está funcionando correctamente? Implementar un agente sin un esquema de evaluación equivale a contratar un profesional sin definir cómo se medirá su desempeño. Para resolver esto existen los EVALs: pruebas de control de calidad diseñadas para verificar que el sistema responda a las exigencias técnicas y comerciales de la empresa.

La principal diferencia entre los EVALs de Agentic AI y de los de modelos tradicionales es su alcance. En un EVAL simple, un agente procesa un prompt y corrobora si el output está acorde con las expectativas. La evaluación de los sistemas agénticos exige mayor complejidad. Al ejecutar flujos de trabajo de múltiples pasos, las decisiones iniciales de la tecnología alteran el entorno, lo que puede provocar que un desvío menor se acumule y se propague a lo largo de la operación. La evaluación ya no mide una respuesta aislada. Ahora exige revisar el proceso completo del agente, analizando tanto las herramientas que utilizó como su capacidad para resolver errores en el camino (Anthropic, 2026).

La práctica más extendida combina tres tipos de revisiones (graders): *code-based graders, model-based graders y human graders*. Estos tipos de revisiones miden tanto el proceso en sí mismo como el resultado final.

01 **Code-based graders:** Verifican condiciones objetivas, como si se utilizó la herramienta correcta, si los parámetros usados son válidos o si el resultado final cumple con una condición concreta. Son evaluaciones rápidas y fácilmente replicables, pero son más bien rígidos.

02 **Model-based graders:** Se encargan de la verificación de las dimensiones cualitativas, como la claridad del razonamiento, tono de respuesta o cumplimiento de instrucciones. Son capaces de evaluar un gran abanico de outputs, tantos como los posibles dentro de la tarea específica. Son flexibles y escalables, pero requieren una calibración constante para asegurar la precisión.

03 **Human graders**: se reservan para casos de alto riesgo, de baja ocurrencia o cuando no hay un juicio unánime entre expertos, donde el juicio automatizado no es suficiente. Son el gold-standard de calidad, pero los más lentos y costosos de escalar.

Anthropic (2024) recomienda priorizar los code-based cuando sea posible y reservar los model-based para cuando el contexto o subjetividad lo requieran .

Un marco de referencia útil para estructurar los EVALs en contexto de adopción en una empresa, scaleup o startup es el CLEAR: Costo, Latencia, Eficacia, Aseguramiento y Confiabilidad. La relevancia está bien documentada: evaluaciones que solo miden precisión pueden resultar en la adopción de Agentic AI de hasta 10 veces más cara que otras alternativas. El mismo estudio indica que un agente con 60% de tasa de éxito inicial puede caer a un 25%, cuando se mide su consistencia en la réplica de un mismo flujo de trabajo, revelando una fragilidad del sistema que las métricas convencionales pueden ocultar (Mehta, 2025).

El World Economic Forum recomienda integrar EVALs dentro de la arquitectura de gobernanza

de la compañía, con criterios de éxito definidos, rendimientos mínimos y alertas ante desviaciones (WEF, 2025). McKinsey (2025c) refuerza este punto, indicando que las organizaciones que logran escalar estas implementaciones son las que tienen una estructura de evaluación definida y el monitoreo como parte central de la arquitectura.

Anthropic (2024) muestra un mapa práctico de pasos 0 a 8 donde las empresas, scaleups y startups pueden partir en la implementación de EVALs desde las etapas más tempranas, como la recolección de las tareas iniciales, hasta el mantenimiento a largo plazo. La idea central, y lo que fundamenta este mapa, es iniciar lo más precozmente posible con casos reales e iterar de forma continua.

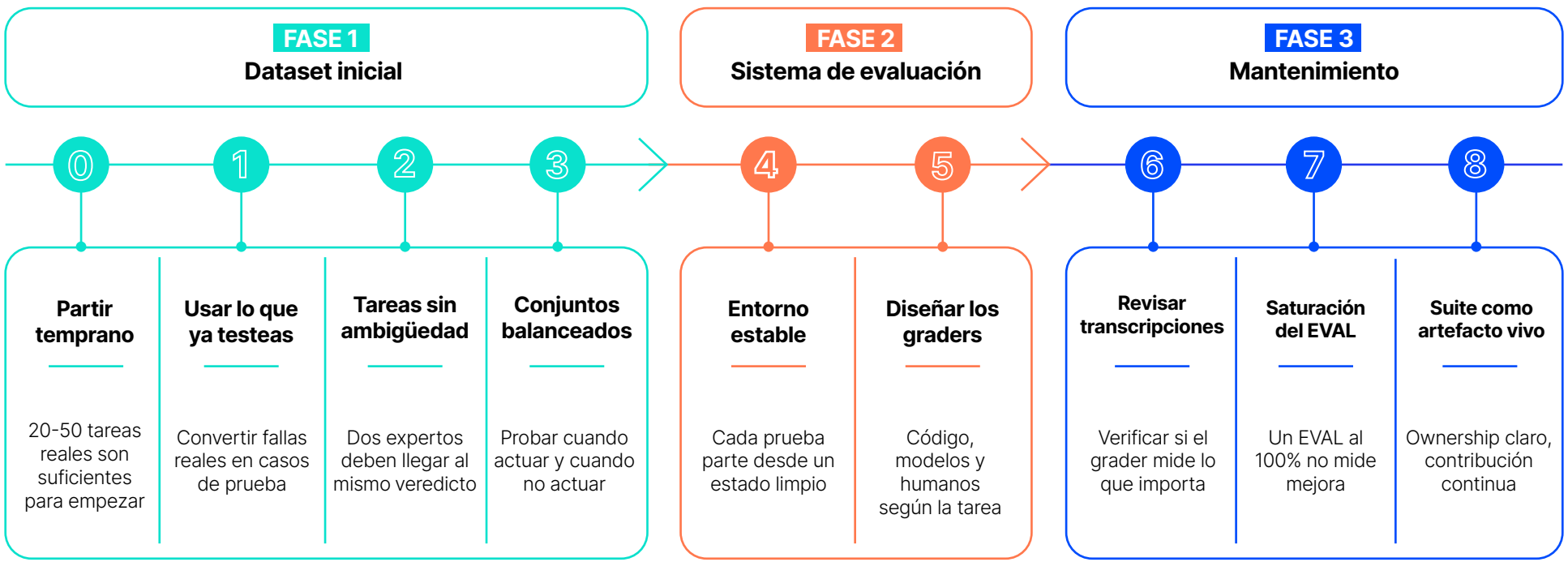


Figura 2: Anthropic (2025). Demystifying evals for AI agents.

2.5
Quién responde cuando el agente falla

Los EVALs responden a la pregunta de si el sistema adoptado funciona correctamente. Pero hay una dimensión que ninguno de esos sistemas puede resolver: quién es responsable cuando el agente yerra, toma una decisión equivocada o genera un daño. En este punto, la asignación de responsabilidad humana deja de ser un asunto puramente técnico para convertirse en una exigencia ética y de diseño organizacional.

Cuando los agentes operan sobre sistemas reales con autonomía creciente, la supervisión humana se vuelve un requisito de diseño. Este enfoque se sustenta en el Reglamento de IA de la Unión Europea (AI Act, 2024), cuyo artículo 14 estipula que los sistemas de alto riesgo deben diseñarse y desarrollarse de tal manera que permitan la supervisión efectiva por humanos, a fin de prevenir riesgos de seguridad y derechos fundamentales. ISG indica que un 25% de las integraciones de Agentic AI operan actualmente con total autonomía, mientras que un 45% son utilizados como un "asesor" donde las decisiones finales son tomadas por humanos (ISG, 2025). Este panorama refleja una postura más bien cauta, donde las organizaciones entienden que entregar autonomía total sin supervisión introduce riesgos difíciles de cuantificar.

La práctica más común describe dos tipos de supervisión. El primer modelo es el Human-in-the-loop (HITL), donde el operador interviene antes de la acción: el agente plantea una solución y el humano la autoriza. Este es el esquema habitual en procesos críticos como operaciones financieras o flujos legales. Por otro lado, está el Human-on-the-loop (HOTL), que sitúa al humano después, donde el agente actúa dentro de límites predefinidos y el humano monitorea métricas y puede intervenir. Las organizaciones con más experiencia en estos ámbitos combinan ambos modelos de supervisión según el tipo de tarea, reservando el HITL para decisiones críticas, irreversibles o de alto impacto (OpenAI, 2025; WEF, 2025).

McKinsey (2025b) describe estos agentes como "Digital Insiders", aludiendo a componentes que actúan de manera interna en los sistemas de la organización con diferentes niveles de acceso y privilegios. Esta naturaleza exige a las organizaciones definir previamente los límites de su autonomía, qué acciones se llevarán a cabo sin la supervisión humana y qué planes de contingencia se activarán en caso de un comportamiento fuera de lo esperado.

Para una Scale Up que esté integrando Agentic AI, el punto de partida es definir cuatro niveles antes de desplegar cualquier adopción de agente:

01 **Qué decisiones puede tomar de forma autónoma**

02 **Qué acciones requieren aprobación o supervisión humana**

03 **Quién es el responsable del agente dentro de la organización**

04 **Qué mecanismo existe para detener o revertir una acción si algo sale mal**

Sin estas bases definidas ex-ante, la supervisión existe solo en la teoría. Como sugiere McKinsey (2025b), un agente sin la debida rendición de cuentas (accountability) asignada deja de ser una herramienta de productividad para transformarse en un riesgo operacional.

FIGURA 3. Matriz de supervisión humana en Agentic AI. Elaboración propia a partir de OpenAI (2025), World Economic Forum (2025) y McKinsey (2025b).

La evidencia global muestra que la ventaja competitiva no depende del presupuesto invertido ni del modelo elegido, sino de los procesos y de la definición de métricas desde el inicio. Para las startups, scaleups y PYMEs chilenas el desafío es la ejecución estratégica de estas adopciones.

03

Lo que tienen en común los que escalaron

Las empresas que acumulan implementaciones sin impacto financiero medible tienen acceso a los mismos modelos, plataformas y en muchos casos presupuestos similares a las que sí han escalado. La diferencia está en las decisiones organizacionales, técnicas y de liderazgo que se tomaron antes y durante la integración. **Esta sección analiza algunos casos con mayor evidencia pública para identificar qué tienen en común quienes lograron efectivamente pasar del piloto al escalamiento, tanto a nivel global como a nivel regional.**

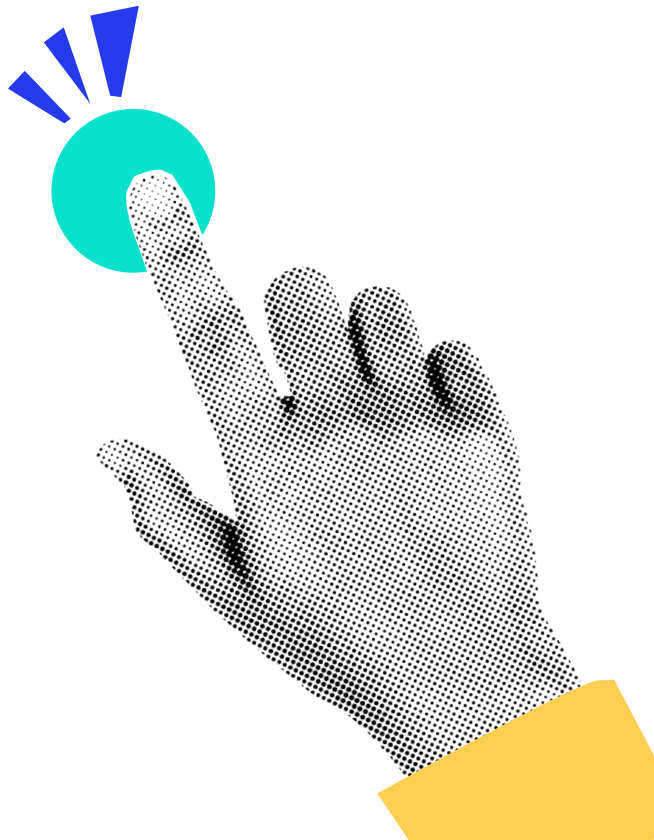
3.1
El patrón de origen

Los casos de éxito comparten una decisión inicial estratégica: se asignaron tareas donde el agente tiene la facultad de decidir y no solo de asistir. Esa diferencia define el impacto final del proyecto.

Un caso destacable es el de Klarna, empresa fintech y banco digital sueco, fundada en 2005. En lugar de diseñar un asistente para el personal, la empresa desplegó un sistema autónomo capaz de resolver disputas y gestionar devoluciones en decenas de mercados. Durante los primeros meses de operación, este agente manejó 2,3 millones de conversaciones, equivalentes al trabajo de 700 trabajadores full time, reduciendo además el tiempo de respuesta de 11 minutos en promedio a menos de 2. Esta implementación generó un impacto financiero de US $40 millones en beneficios solo durante 2024 según los datos reportados por la misma empresa (OpenAI, 2024a). Klarna recalibró parte de esta operación en 2025, reincorporando agentes humanos para casos complejos, lo que muestra que el nivel de autonomía es una decisión que se ajusta con la operación y no se fija una sola vez (CX Dive, 2025). Amazon aplicó una estrategia similar a través de Amazon Q. De acuerdo con información proporcionada por Amazon, en lugar de limitarse a lanzar un asistente de código, la compañía diseñó un agente capaz de migrar aplicaciones de Java de principio a fin. Esto se tradujo en la actualización autónoma de decenas de miles de sistemas, generando eficiencias equivalentes a 4.500 años de trabajo técnico y un ahorro anual estimado de US $260 millones (Amazon Web Services, 2024).

A nivel regional, MercadoLibre implementó el agente Verdi para resolver disputas entre compradores y vendedores. Según MercadoLibre, en sus primeros meses, Verdi gestionó de forma autónoma el 10% de las mediaciones, con potencial, según la empresa, de cubrir tareas equivalentes a 9.000 operadores humanos y decisiones que involucran cerca de US $450 millones anuales(OpenAI, 2024b).

Nubank ilustra el mismo patrón pero en una escala mayor. El neobanco brasileño, que cuenta con 131 millones de clientes en tres países de la región, documenta en su propio blog de ingeniería el paso desde sistemas de búsqueda y respuesta hacia agentes con arquitectura ReAct, capaces de ejecutar acciones reales sobre sistemas financieros tales como renegociación de deuda, gestión de tarjetas y prevención de fraude. El equipo técnico de Nubank indica que si un sistema no puede activar una acción real, es una interfaz de búsqueda, no un agente (Nubank, 2026).



3.2

La decisión que separa el piloto de la escala

Para cada caso analizado hay una decisión que define el curso de la integración: Qué nivel de autonomía está dispuesta a entregarle la organización al agente. Esta traba en el escalamiento ocurre cuando las organizaciones relegan la tecnología a un rol meramente consultivo. Si el agente solo formula sugerencias y espera validación humana constante, la propia estructura bloquea el retorno económico del proyecto.

Klarna optó por un modelo autónomo desde el inicio, asumiendo el riesgo operacional y calibrando la supervisión humana en el proceso. En el caso de Amazon, el sistema realiza la migración completa antes de la revisión técnica. MercadoLibre, por su parte, facultó al agente para resolver las reclamaciones directamente. En estos modelos, el HITL se desplaza desde la aprobación paso a paso hacia el diseño de las reglas de gobernanza que definen el sistema (OpenAI, 2024a).

La segunda decisión consistente es dónde se instala la iniciativa dentro de la organización.

De acuerdo con BCG (2025), las empresas que lideran el despliegue de agentes comparten una característica de su modelo operativo: las gerencias del negocio asumen un control directo de las métricas de éxito. La tecnología se aborda como un vehículo comercial y nunca como un fin corporativo aislado.

El nivel de autonomía del agente y la responsabilidad organizacional de la iniciativa son decisiones de liderazgo que deben tomarse antes de elegir modelos o plataformas. De ellas depende si la adopción tiene condiciones reales para escalar.

3.3

Cómo justificaron la inversión

La sostenibilidad de estas iniciativas depende de su capacidad para demostrar valor. El factor definitivo que separa los proyectos que no escalan y los que sí es la traducción del rendimiento técnico en mejoras financieras explícitas, reflejadas en la rentabilidad de la operación.

En el capítulo anterior se identificaron formas de medir el retorno real de esta inversión, haciendo alusión a su capacidad de reducir costos y riesgos y a la vez aumentar ingresos. Este retorno puede ser valioso desde dos puntos de vista relevantes, la mirada operativa y el impacto financiero de las iniciativas, ambos factores sumamente importantes para la sostenibilidad.

En la medición operativa vemos ejemplos de mejoras en tiempos de proceso, capacidad y tasa de errores. Klarna cifró el impacto de su agente en el equivalente a 700 humanos trabajando a tiempo completo, con una reducción de tiempo de respuesta de casi un 82% y una caída del 25% de consultas repetidas (OpenAI, 2024a). MercadoLibre por su parte, delegó a su agente Verdi la resolución autónoma del 10% de las disputas entre compradores y vendedores, liberando capacidad humana para los casos más complejos (OpenAI, 2024b). En ambos casos, la métrica que importa es el rendimiento del proceso donde opera, más que la adopción del mismo sistema.

El impacto financiero es la segunda dimensión, y que define si la inversión se sostiene. Duolingo reportó en 2025 un crecimiento interanual del 41% en ingresos, con su CEO señalando que el segmento premium con funciones de IA es uno de los pocos casos de empresas rentables atribuibles directamente a características de IA (Reuters, 2025). Nubank midió el impacto en eficiencia antes de escalar la adopción de sus agentes. Su asistente basado en GPT-4o resolvía el 55% de las consultas de primer nivel sin intervención humana, gestionaba más de 2 millones de conversaciones mensuales y redujo en 70% el tiempo de respuesta, cifras que justificaron la inversión posterior en el sistema de agentes que reemplazó ese enfoque (OpenAI, 2025).

Lo que tienen en común estos ejemplos es el paso previo. Antes de desplegar la tecnología, cada organización definió qué indicador iba a medir y con qué línea base. Sin este proceso, cualquier cifra de mejora posterior carece de respaldo. McKinsey lo señala como una práctica distintiva de los líderes, la capacidad de rastrear KPIs específicos y rediseñar flujos, basados en procesos previamente diagnosticados (McKinsey, 2025d).

Para una Scale Up chilena, las preguntas que se desprenden de esta sección son el punto de partida al diseñar una estructura que pueda adoptar un modelo de Agentic AI:

- **¿Hay un proceso concreto donde un agente pueda tomar decisiones más allá de asistir?**
- **¿Cuál es la línea base o de control desde donde se construirá la evaluación de la adopción del agente?**
- **¿Qué área de la estructura es la dueña y responsable de los resultados de la implementación?**

Adicionalmente a esto, startups o scaleups deben tener cuidado de caer en un error fundamental en organizaciones de este tipo y es que esperar al momento o proceso perfecto para implementar cambios o avances tecnológicos. Esto puede llevar a una constante espera que frene la innovación y el progreso. Es por esto que adicionalmente se recomienda ponerse expectativas y métricas modestas al comenzar, para poder implementar procesos que sean reversibles y de bajo costo, cuyos avances puedan ser los primeros paso a proceso de Agentic AI con mayor alcance y rentabilidad.

04

Qué debe tener la organización antes de empezar

Los patrones que se observan en la sección del panorama global de adopción y los casos analizados apuntan hacia un mismo lugar: lo que diferencia un piloto que escala a uno que no avanza pocas veces es el modelo o el presupuesto disponible. Casi siempre es el estado interno de la organización antes que el agente entre en funciones. **A partir de esta evidencia, identificamos cinco dimensiones que, cuando faltan, explican el estancamiento en la implementación de Agentic AI en las empresas.**

Este marco no busca ser exhaustivo. Es una síntesis de lo que tienen en común las implementaciones que funcionaron.

A) Liderazgo y visión estratégica

Implementar Agentic AI es una decisión del negocio más que un proyecto de sistemas. En empresas donde las estructuras organizacionales son más bien pequeñas y el equipo directivo es reducido esto tiene ventajas, ya que no existe una burocracia que haga más difícil la definición de los problemas. El riesgo específico aparece cuando el liderazgo no está involucrado activamente, cuando la iniciativa queda en manos del área técnica sin un mandato comercial, y sin ese mandato es difícil priorizar, financiar y sostener la implementación. El liderazgo de la empresa debe esclarecer tres pilares antes de evaluar plataformas: el proceso específico a transformar, el indicador financiero de éxito y el responsable directo del resultado.

B) Estructura Organizacional

La localización en la estructura de la empresa determina las responsabilidades y agilidad en el cierre de desviaciones y errores. En estructuras organizacionales más pequeñas existe el riesgo de que estas funciones se diluyan. Eso implica que es indispensable una jefatura de proyecto que asuma responsabilidad directa sobre el rendimiento del agente, evitando vacíos de gobernanza que comprometan la continuidad de su implementación.

C) Calidad de datos y sistemas

La viabilidad operativa de un agente depende de la disponibilidad y calidad de los datos. La inconsistencia o fragmentación de los datos anula la eficacia del sistema. Esto no es exclusivo de Agentic AI, sino que es una deficiencia estructural que amenaza cualquier proceso, incluido uno manual.

El requisito fundamental es determinar los requerimientos específicos de información para la tarea asignada, asegurando centralización y depuración previa. El despliegue inicial no exige infraestructura de datos complejas, sino que la conformación de los registros del proceso se encuentren en formatos estandarizados y compatibles.

Adicionalmente, los avances en agentes IA modernos presentan una variedad de herramientas pre-entrenadas con bases de datos complejas y actualizadas. Por lo que la calidad de datos y sistemas puede no requerir de utilización de bases internas o información de empresa, sino que más bien debe depender exclusivamente del tipo de operación o proceso en el cual se requiera emplear Agentic AI. Existirán procesos que requieran de bases de datos internas, cerradas y eficientes, mientras que otros solamente necesitarán acceso a los modelos más recientes para acceder a las capacidades entrenadas más actualizadas.

D) Cultura y Tolerancia al riesgo

Integrar un agente implica ceder autonomía a un sistema para que tome decisiones operativas. Cuando se le asigna autonomía a un sistema que tome decisiones operativas la organización está aceptando de forma implícita un margen de incertidumbre que no existe cuando los procesos los realizan humanos con criterio propio. En una empresa con márgenes estrechos, donde cualquier fallo puede impactar los resultados, esta transición no es sencilla.

El marco de contención mínimo tiene dos componentes: la zona de falla tolerable, donde el agente puede equivocarse sin consecuencias graves y el equipo puede recalibrar, y las líneas rojas, donde un fallo es inaceptable y requiere intervención humana inmediata. El escalamiento exitoso de la tecnología requiere un enfoque orientado al aprendizaje iterativo y no a la aversión al error inicial.

E) Accountability humano, ética y aspectos legales

A medida que el agente gana autonomía, la asignación de responsabilidades ante fallos del sistema pasa a ser una prioridad de gobernanza. Relegar este análisis a etapas posteriores a la implementación constituye un riesgo crítico. Las organizaciones que logran un despliegue escalable incorporan directrices de rendición de cuentas desde la fase de diseño arquitectónico.

El marco de diseño práctico debe delimitar las atribuciones autónomas, los hitos de validación obligatoria por parte del humano, la responsabilidad organizacional del sistema y los protocolos de reversión ante incidentes (OpenAI, 2025). Este alineamiento técnico es crítico ante la aparición de normativas internacionales como la Ley de Inteligencia Artificial de la Unión Europea (EU AI Act, 2024). Chile aún no cuenta con un marco que regule la autonomía del Agentic AI, sin embargo existe el proyecto de ley N° de Boletín 16821-19 presentado el 7 de mayo 2024 ante la cámara de diputados con el objetivo de regular los sistemas de inteligencia artificial y que actualmente se encuentra en el segundo trámite constitucional (BCN, 2024). Sus lineamientos siguen de cerca lo planteado por la Unión Europea y por eso constituyen una buena referencia para empresas chilenas que deseen adelantarse a la legislación.

La matriz siguiente cruza las cinco dimensiones con la etapa de implementación, permitiendo identificar dónde está la empresa y qué condiciones necesita para avanzar.

	01 Exploración	02 Piloto	03 Escala
A. Liderazgo y visión estratégica	La IA la lleva el CTO o el equipo técnico; el founder no define qué problema resolver.	Founder o CTO define el proceso, la métrica y quién responde por el resultado.	La dirección toma decisiones estratégicas considerando las capacidades del agente.
B) Estructura Organizacional	Todos hacen de todo; nadie es dueño de la iniciativa ni de sus resultados.	Un responsable claro; el área afectada participa en el diseño, no solo lo recibe.	Roles definidos con accountability sobre KPIs; el agente tiene dueño en producción.
C) Calidad de datos y sistemas	Datos en planillas, sistemas distintos o sin registro; no hay línea base.	Datos del flujo piloto disponibles, limpios y accesibles para el sistema.	Datos integrados entre procesos; múltiples agentes los consumen.
D) Cultura y Tolerancia al riesgo	El agente solo asiste; cualquier error paraliza la iniciativa o la descarta.	Autonomía acotada con umbrales claros; el error se documenta y se ajusta.	La falla es insumo; el equipo itera sin necesidad de validar cada ajuste.
E) Accountability humano, ética y aspectos legales	Sin definición de límites; nadie sabe qué hace el agente ni quién responde.	HITL en decisiones críticas; un responsable interno y mecanismo de reversión definido.	Gobernanza activa; HITL y HOTL según tarea; posición adelantada al AI Act.

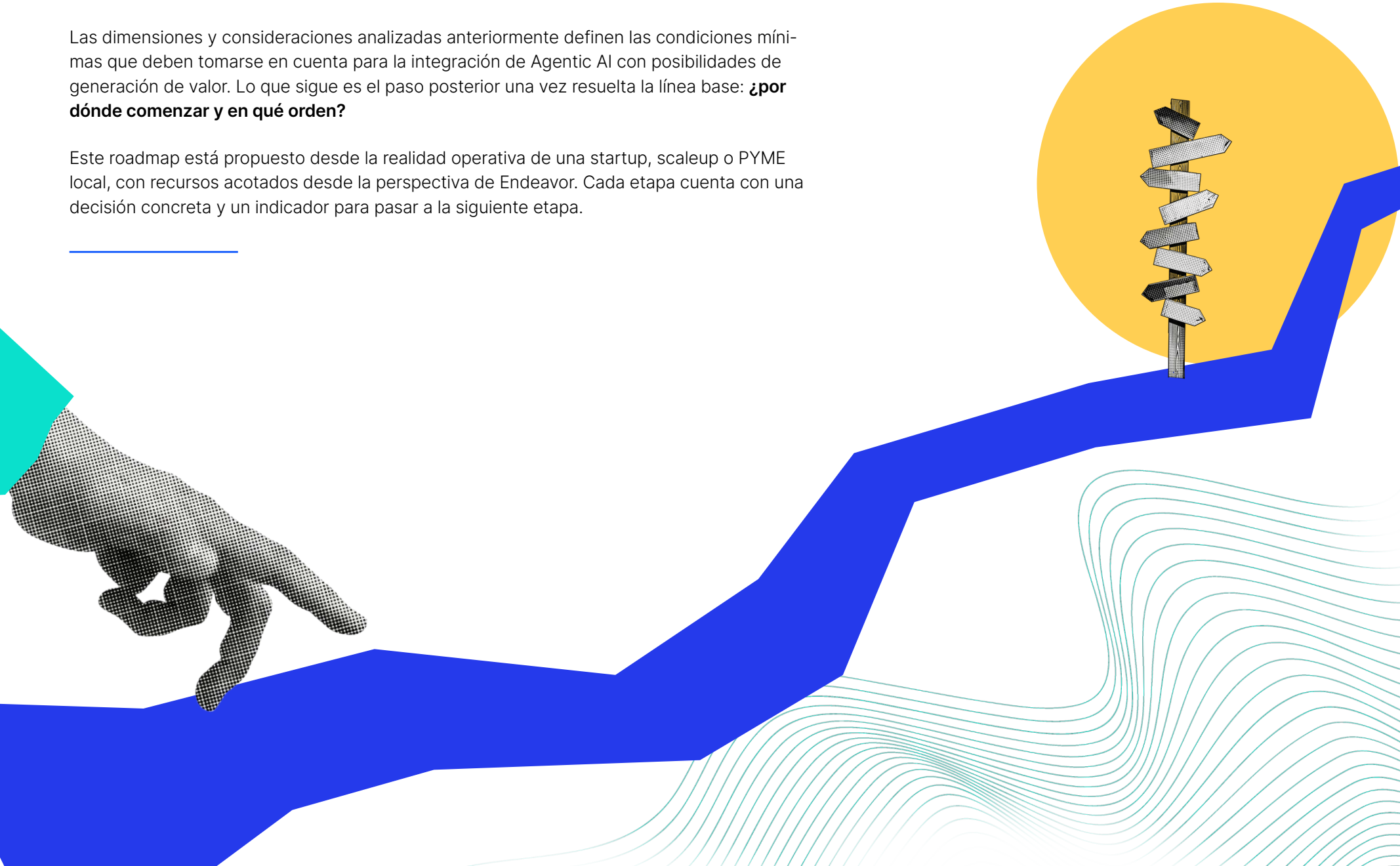
FIGURA 2. Matriz de condiciones por fase de adopción.

05

Por dónde partir y en qué orden

Las dimensiones y consideraciones analizadas anteriormente definen las condiciones mínimas que deben tomarse en cuenta para la integración de Agentic AI con posibilidades de generación de valor. Lo que sigue es el paso posterior una vez resuelta la línea base: **¿por dónde comenzar y en qué orden?**

Este roadmap está propuesto desde la realidad operativa de una startup, scaleup o PYME local, con recursos acotados desde la perspectiva de Endeavor. Cada etapa cuenta con una decisión concreta y un indicador para pasar a la siguiente etapa.



01 02 03 04 05 06

Elegir el proceso >>>

El proceso es el punto de partida porque el valor que puede entregar el agente depende de qué tan bien ejecuta una operación concreta. **Deloitte (2025) indica que los agentes generan retornos cuando se despliegan sobre procesos específicos con impacto medible, no cuando están sobre una plataforma sin un caso de uso definido.**

El error más frecuente en las etapas iniciales es partir desde la herramienta: implementar un agente sin tener claro en qué proceso ni para qué. Esa lógica casi siempre termina en un piloto sin impacto medible.

El punto de partida correcto exige aislar una operación específica dentro del proceso de negocios que cumpla con los siguientes requisitos:

- **Repetibilidad:** el proceso ocurre de forma recurrente y sigue una secuencia lógica de reglas.
- **Disponibilidad de información:** Los datos que usará el agente deben estar disponibles, accesibles y estructurados.
- **Impacto:** La optimización debe reducir costos, liberar tiempo o ayudar a generar aumentos en retornos.
- **Valor:** Identificación del proceso en el cual la optimización puede generar mayor valor.
- **Facilidad de implementación:** Identificar qué tan sencillo es de implementar en tal proceso y la dificultad de revertir los cambios.

Criterio crítico

El equipo a cargo es capaz de describir la operación en menos de cinco oraciones, identificar los hitos de inicio y término y saber con exactitud qué datos serán necesarios para ejecutar cada tarea.

Definir la instrucción >>>

Una instrucción mal definida es una de las principales fuentes de errores en los agentes. Anthropic documenta que la mayoría de los errores observados en implementaciones de agentic AI en empresas no vienen del modelo sino de que las instrucciones no especifican con claridad el objetivo, alcance y criterios de éxito del agente (Anthropic, 2025).

Con el proceso ya seleccionado, el siguiente paso es estructurar la instrucción que recibirá el agente. Una instrucción debe estar bien definida, sin ambigüedades y debe construirse sobre los siguientes componentes:

Objetivos específicos: No debe ser genérico, sino más bien taxativo, con un output medible y auditable.

Flujo secuencial: La ruta y pasos deben estar definidos de manera clara, son los que el agente ejecutará para resolver la tarea.

Criterios de validación: Se deben delinear las reglas en las cuales el agente operará, es una decisión clave que define si el agente está operando de la manera que se espera o está desviándose.

Criterio crítico

Se debe contar con un documento con el detalle preciso de qué hace el agente, cómo lo hace y cómo se diferencia de forma cuantitativa un proceso exitoso de uno fallido.

Validar las condiciones mínimas >>>

Comprometer recursos antes de verificar si se cuenta con las condiciones mínimas puede inducir a fallos al momento de la implementación. McKinsey (2025c) identifica la disponibilidad de los datos, el equipo responsable y la holgura operacional como condiciones previas al despliegue, no para resolver en paralelo.

Con el proceso seleccionado y la instrucción definida, se debe evaluar si la empresa tiene capacidad real de ejecutar el proyecto. Se requiere un diagnóstico honesto antes de comprometer recursos o planificaciones.

Los datos: La información que alimentará el agente debe estar estructurada, disponible y centralizada.

El equipo: Se necesita contar con un responsable interno con competencias para configurar, supervisar y ajustar la herramienta en caso de ser necesario.

Holgura operacional: Se deben contar con recursos y capital humano disponibles para poder realizar la implementación.

Criterio crítico

Las tres condiciones se cumplen a cabalidad, o bien se cuenta con una planificación y recursos asignados para llevarse a cabo.

Establecer métricas >>>

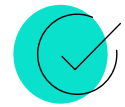
La implementación de un agente sin una medición de la línea base no tiene con qué compararse. Para McKinsey (2025c) es una condición necesaria para poder medir su rendimiento.

Antes de integrar el agente, es necesario tener una radiografía actual del proceso, su tiempo de operación, los costos asociados, la tasa de error promedio y qué capacidad humana tiene. Este es el único parámetro que se tendrá para comparar y demostrar el impacto de la tecnología en su despliegue.

Una línea base apropiada debe obedecer criterios de calidad y operación validados internamente con tal de poder evaluar procesos que sean atribuibles al Agentic AI y que coincidan con las razones de base de porqué se decide implementar este Agente.

En el caso de empresas startups la línea base puede ser una tarea compleja al no tener un registro histórico disponible o un crecimiento acelerado que haga difícil la determinación de metas. Sin embargo, no por eso se debe evitar la formación de métricas de este tipo, estas son fundamentales para el crecimiento de empresas independiente del uso de IA e incluso el proceso de ensayo y error puede resultar muy efectivo para evaluar el proceso productivo en general.

En paralelo, se deben tener EVALs activos desde el inicio, para poder evaluar si el proceso está respondiendo a la necesidad y si el output es el esperado.



Criterio crítico

Contar con la línea base del proceso documentada con métricas que se usarán para evaluar el trabajo del agente, como también pruebas operativas para asegurar la fiabilidad del mismo.

Protocolo de contingencia

Un agente sin accountability asignado es un riesgo operacional antes que una herramienta. Cuando los límites de autonomía no están definidos desde el diseño, la responsabilidad se diluye y los errores no tienen dueño. El EU AI Act lo convierte en requisito formal para sistemas de alto riesgo: la supervisión humana debe estar diseñada en el sistema, no incorporada después (EU AI Act, 2024).

Un agente cometerá errores, y la clave es saber qué se hará cuando ocurran. Se deben definir cuatro controles necesarios que serán las reglas de operación del agente.

- **Límites de autonomía:** Qué decisiones tomará el agente por sí solo.
- **Puntos de control humano:** Qué acciones quedan sin ejecutar hasta recibir una validación humana.
- **Dueño de la operación:** Quién es la persona del equipo responsable final por los resultados y comportamiento del agente.
- **Mecanismo de revisión:** Cuál es el protocolo de emergencia para detener un proceso o deshacerlo en caso de una falla crítica.

Criterio crítico

Los cuatro elementos de control están formalizados por escrito, validados transversalmente y comunicados a todo el equipo que interactúa con la operación del agente.

Decidir cuándo escalar

Escalar un piloto que no ha demostrado consistencia es amplificar un problema, no resolverlo. Las organizaciones que lo hacen bien esperan evidencia de rendimiento estable en condiciones reales antes de expandir, no extrapolando desde las primeras semanas favorables (McKinsey, 2025e).

Para escalar el sistema es necesario verificar las siguientes tres condiciones:

Consistencia: El agente muestra comportamientos estables y predecibles durante un periodo prolongado de tiempo, superando los ajustes iniciales.

Impacto cuantificable: Las mejoras operativas o financieras están medidas de manera correcta y pueden ser un caso de éxito para la empresa.

Gobernanza: El esquema de control, validación y mitigación fueron implementadas de forma correcta y no generaron fricciones en la operación diaria.

Criterio crítico

Cuando estas tres dimensiones están bajo control, el escalamiento se convierte en una decisión estratégica y financiera. Si alguna de ellas no está controlada, se estaría tomando un riesgo innecesario en el escalamiento.

06

Conclusiones

Para el founder o CTO de una PYME, scaleup o startup que aún no ha empezado

Para quienes aún no dan el primer paso, el principal obstáculo no suele ser la falta de presupuesto ni el acceso a la tecnología. El problema real viene del camino lógico que se debe seguir, donde muchas veces se elige la herramienta antes de entender el fondo del problema o proceso que se quiere mejorar. Este error en la secuencia es la razón por la cual la mayoría de las primeras implementaciones fallan o no tienen un resultado medible en el negocio.

La recomendación que se extrae de este reporte, y desde Endeavor, es que antes de evaluar plataformas o modelos, **el equipo debe ser capaz de responder si está definida la operación a transformar, cómo se medirá y saber quién es el responsable del proyecto.**

Realizar una implementación de Agentic AI sin tener estas certezas garantiza el fracaso de la misma, pérdida de recursos y de tiempo. Por el contrario, comenzar por esta secuencia de pasos lógicos, definiendo lo descrito anteriormente, proveerá una base sólida para generar una mejora real y medible en los procesos de negocio de la empresa.

Para la PYME, scaleup o startup que ya está piloteando

En esta etapa existe un problema muy habitual. **El piloto funciona correctamente, está el equipo conforme, pero nadie puede explicar qué impacto real tuvo en el negocio la implementación**. Esto es síntoma de que el proyecto no tuvo un responsable desde el principio y se ejecutó exclusivamente como un experimento.

Destrabar esta situación requiere una decisión organizacional, definiendo quién responde por el sistema y cómo se medirá el éxito o fracaso del proyecto, tal como se le pide resultados a un empleado. Los casos descritos en este reporte muestran que el escalamiento que mostraron empresas como Klarna, MercadoLibre o Amazon no lo lograron usando el mejor modelo, sino **definiendo desde el primer día un mandato explícito de qué hacer y cómo se mediría.**

Para inversionistas, aceleradoras y política pública

Para las entidades que financian, aceleran o diseñan políticas públicas, el criterio de evaluación de las empresas que integran Agentic AI deben dar un giro en su evaluación, dejando atrás la tecnología como tal y enfocarse a medir el impacto operativo real.

La pregunta clave, y que funciona como prueba antes de asignar recursos es: **¿el proceso que se pretende automatizar estaba medido antes de implementar la IA?** Cualquier implementación sin este marco de referencia no tendrá sustento demostrable al momento de evaluarse.

Desde Endeavor, observamos que el cuello de botella no es el acceso a la tecnología, sino que recae en características organizacionales variadas, desde la definición correcta del problema, a asignar responsabilidad y medir desde el comienzo. **Para los diferentes actores del ecosistema implica incorporar el diagnóstico previo como condición para acceder a programas y exigir evidencia que los sistemas muestren el impacto esperado.**

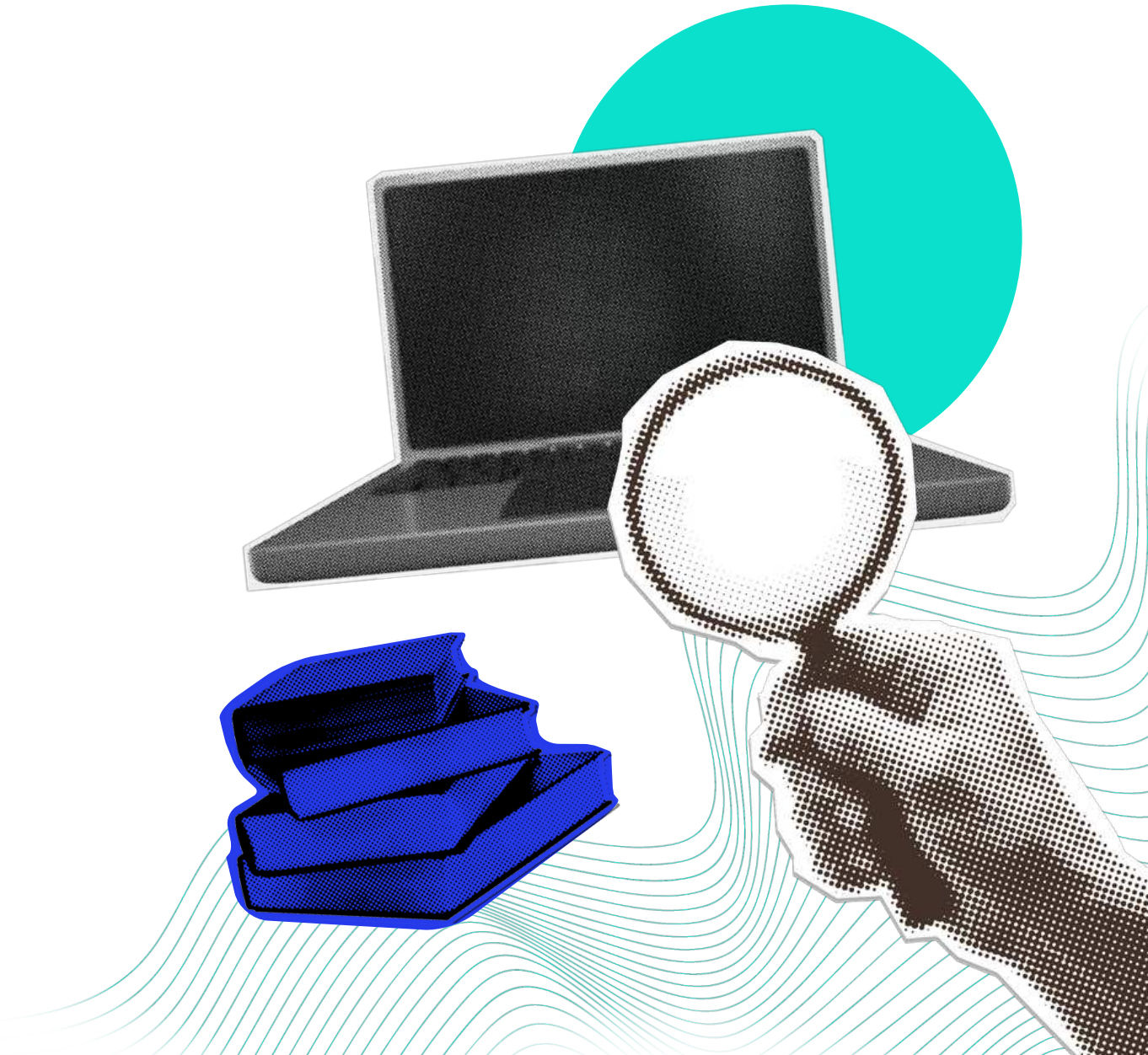
Para empresas grandes

Empresas en un rango de ventas superior pueden verse tentadas a invertir en sistemas de IA o Agentic IA de manera apresurada debido a la disponibilidad de montos de inversión. Sin embargo, la evidencia identificada en los casos de éxito y revisión de literatura identifica un **patrón claro en como empresas similares han logrado asegurar una implementación efectiva.**

El patrón identificado fue la rápida inversión en IA y reemplazo masivo del componente humano, sin embargo, y mediante el uso de indicadores clave para medir la utilidad y justificar la inversión, estas empresas comenzaron a realizar ajustes en el sentido contrario, incentivando la colaboración entre IA y trabajadores humanos, enfatizando la necesidad de que una persona esté **monitoreando y evaluando las tareas en las que Agentic IA puede ser de gran utilidad.**

Por lo tanto, para empresas grandes que estén comenzando, o ya hayan comenzado, su implementación de Agentic AI **es fundamental asegurar un apropiado uso de recursos y efectividad al determinar espacios efectivos en los que estas herramientas pueden mejorar la eficacia y eficiencia del proceso productivo**, seleccionar indicadores clave para medir esta implementación y fomentar la cooperación humana-IA al asignar empleados en posiciones que permitan no sólo supervisar al Agentic AI sino que mejorar exponencialmente la calidad y efectividad de los procesos automatizados.

Metodología



Este estudio utiliza principalmente la revisión de fuentes secundarias como reportes de la industria, papers académicos y publicaciones de organizaciones públicas y privadas líderes en el campo. Los criterios de selección priorizaron fuentes con evidencia publicada entre 2023 y 2026, con preferencia de aquellas que documentan casos reales de la industria.

Los casos analizados fueron seleccionados por la disponibilidad de evidencia pública con fuentes verificables. Se priorizaron casos con métricas documentadas por la misma organización o fuentes independientes.

Para la elaboración de este documento se utilizaron las siguientes herramientas IA como complemento de las metodologías explicadas: Perplexity para la identificación de literatura relevante; NotebookLM para el análisis de documentos previo a su incorporación; y Claude para revisión de redacción general del documento.

Bibliografía

Amazon Web Services. (2024). *Amazon Q Developer just reached a $260 million dollar milestone*. https://aws.amazon.com/blogs/devops/amazon-q-developer-just-reached-a-260-million-dollar-milestone/

Andreessen Horowitz. (2023). *Who owns the generative AI platform? Andreessen Horowitz*. https://a16z.com/who-owns-the-generative-ai-platform/

Andreessen Horowitz. (2025). *The AI application spending report: Where startup dollars really go*. https://a16z.com/the-ai-application-spending-report-where-startup-dollars-really-go/

Anthropic. (2024). *Building effective agents*. https://www.anthropic.com/research/building-effective-agents

Anthropic. (2025). *Effective context engineering for AI agents*. https://www.anthropic.com/engineering/effective-context-engineering-for-ai-agents

Anthropic. (2026). *Demystifying evals for AI agents*. https://www.anthropic.com/engineering/demystifying-evals-for-ai-agents

BCG. (2025). *How companies can prepare for an AI-first future*. https://www.bcg.com/publications/2025/how-companies-can-prepare-for-ai-first-future

Bessemer Venture Partners. (2025). *State of AI 2025*. https://www.bvp.com/atlas/the-state-of-ai-2025

Biblioteca del Congreso Nacional de Chile (BCN). (2024). *Boletín Legislativo (N.º 462, 7 a 20 de mayo de 2024)*. https://www.bcn.cl/boletines/boletinleg.html?id_boletin=6&nro_boletin=462

Bort, J. (2025). *Even a16z VCs say no one really knows what an AI agent is. TechCrunch*. https://techcrunch.com/2025/05/12/even-a16z-vcs-say-no-one-really-knows-what-an-ai-agent-is/

Capgemini. (2025). *Rise of agentic AI: How trust is the key to human-AI collaboration. Capgemini Research Institute*. https://www.capgemini.com/insights/research-library/ai-agents/

Capgemini. (2026). *From pilots to real impact: How enterprises actually scale agentic AI*. https://www.capgemini.com/insights/research-library/from-pilots-to-real-impact-how-enterprises-actually-scale-agentic-ai/

CX Dive. (2025). *Klarna changes its AI tune and again recruits humans for customer service*. https://www.customerexperiencedive.com/news/klarna-reinvests-human-talent-customer-service-AI-chatbot/747586/

Deloitte. (2025). *The business imperative for agentic AI*. https://www.deloitte.com/content/dam/assets-zone1/in/en/docs/services/engineering-ai-data/in-eaid-the-business-imperative-for-agentic-ai.pdf

EU AI Act. (2024). *Regulation on artificial intelligence, Article 14: Human oversight*. https://artificialintelligenceact.eu/article/14/

ISG. (2025). *State of the agentic AI market report 2025. ISG Research*. https://isg-one.com/advisory/artificial-intelligence-advisory/state-of-the-agentic-ai-market-report-2025

Bibliografía

McKinsey & Company. (2025a). *Agentic AI explained: When machines don't just chat but act*. https://www.mckinsey.com/featured-insights/mckinsey-explainers/agentic-ai-explained-when-machines-dont-just-chat-but-act

McKinsey & Company. (2025b). *Deploying agentic AI with safety and security: A playbook for technology leaders*. https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/deploying-agentic-ai-with-safety-and-security-a-playbook-for-technology-leaders

McKinsey & Company. (2025c). *Seizing the agentic AI advantage*. https://www.mckinsey.com/capabilities/quantumblack/our-insights/seizing-the-agentic-ai-advantage

McKinsey & Company. (2025d). *The state of AI in 2025: Agents, innovation, and transformation*. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

McKinsey & Company. (2025e). *Building the foundations for agentic AI at scale*. https://www.mckinsey.com/capabilities/mckinsey-technology/our-insights/building-the-foundations-for-agentic-ai-at-scale

McKinsey & Company. (2026). *Agentic AI advances. The Week in Charts*. https://www.mckinsey.com/featured-insights/week-in-charts/agentic-ai-advances

Mehta, S. (2025). *Beyond accuracy: A multi-dimensional framework for evaluating enterprise agentic AI systems*. https://arxiv.org/abs/2511.14136

Morningstar. (2026). *AI and economic moats: Which stocks are most at risk?* https://global.morningstar.com/en-eu/markets/ai-economic-moats-which-stocks-are-most-risk

Nubank. (2026). Building AI agents for 131 million customers. https://building.nubank.com/building-ai-agents-for-131-million-customers/

OpenAI. (2023). *Practices for governing agentic AI systems*. https://openai.com/index/practices-for-governing-agentic-ai-systems/

OpenAI. (2024a). *Klarna*. https://openai.com/index/klarna/

OpenAI. (2024b). *Mercado Libre presenta Verdi*. https://openai.com/es-419/index/mercado-libre/

OpenAI. (2025). *A practical guide to building agents*. https://cdn.openai.com/business-guides-and-resources/a-practical-guide-to-building-agents.pdf

OpenAI. (2025). Nubank elevates customer experiences with OpenAI. https://openai.com/index/nubank/

OpenAI. (2026a). *GPT-4 model. OpenAI API Documentation*. https://developers.openai.com/api/docs/models/gpt-4

OpenAI. (2026b). *GPT-4o model. OpenAI API Documentation*. https://developers.openai.com/api/docs/models/gpt-4o

PwC. (2026). *Want ROI from AI? Go for growth: PwC's AI performance study*. https://www.pwc.com/gx/en/so-you-can/2026/content/roi-from-ai.pdf

Bibliografía

Ransbotham, S., Kiron, D., Khodabandeh, S., Iyer, S., y Das, A. (2025). *The emerging agentic enterprise: How leaders must navigate a new age of AI. MIT Sloan Management Review y Boston Consulting Group.* https://sloanreview.mit.edu/ai2025

Redpoint Ventures. (2026). 2026 market update. https://www.redpoint.com/reports/2026-market-update

Reuters. (2025). *Duolingo says AI features profitable as it beats revenue estimates.* https://www.reuters.com/business/duolingo-says-ai-features-profitable-it-beats-revenue-estimates-raises-forecast-2025-11-05/

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

Sobre este estudio

Este estudio fue realizado por el equipo de Endeavor Research Chile

EQUIPO ENDEAVOR RESEARCH CHILE

Andrés Alvarado Valencia
Director Endeavor Research

Rocío López Garafulic
Investigadora

Nicolás Araya Domínguez
Investigador

El desarrollo de este estudio fue posible gracias al apoyo de:

ORACLE ACID LABS

Para conocer otros estudios de Endeavor Research visita:
www.endeavor.cl/research

Cómo citar:
López, R., Alvarado, A., & Araya, N. (2026). Agentic AI en el ecosistema emprendedor chileno. Endeavor.

UN ESTUDIO DE

endeavor

CON EL APOYO DE

ORACLE ACID LABS

Agentic AI en el ecosistema emprendedor chileno

CONDICIONES, PATRONES Y VALOR