LA CADENA DE VALOR DE LA INTELIGENCIA ARTIFICIAL

Tensiones geopolíticas y la carrera global por la competitividad de América Latina y el Caribe en el mundo

Raúl Katz
Fernando Callorda
Ramiro Valencia

Naciones Unidas
CEPAL
Global Gateway
Alianza Digital UE-ALC

NACIONES UNIDAS

CEPAL

Gracias por su interés en esta publicación de la CEPAL

¿Desea compartir esta obra?

Utilice el enlace ubicado en la portada final de este PDF, que dirige a la versión oficial de la publicación en el sitio web de la CEPAL. Por favor, no cargue ni distribuya copias de este archivo en sitios web o plataformas externas.

Deseo registrarme

Si desea recibir información oportuna sobre nuestros productos editoriales y actividades, le invitamos a registrarse. Podrá definir sus áreas de interés y acceder a nuestros productos en otros formatos.
https://mailchi.mp/cepal/suscripciones-cepal

Recursos digitales

Conozca nuestras redes sociales y otras fuentes de difusión en el siguiente enlace: **https://bit.ly/m/CEPAL**

Documentos de Proyectos

La cadena de valor de la inteligencia artificial

Tensiones geopolíticas y la carrera por la competitividad de América Latina y el Caribe en el mundo

Raúl Katz
Fernando Callorda
Ramiro Valencia

NACIONES UNIDAS
CEPAL
Global Gateway

Este documento fue preparado por Raúl Katz, Fernando Callorda y Ramiro Valencia, Consultores de la División de Desarrollo Productivo y Empresarial de la Comisión Económica para América Latina y el Caribe (CEPAL), bajo la coordinación de Sebastián Rovira, Oficial Superior de Asuntos Económicos, y Valeria Jordán, Oficial de Asuntos Económicos, ambos de esa misma División.

Los autores agradecen especialmente los valiosos comentarios y aportes de Valeria Jordán, que contribuyeron al análisis, la coherencia conceptual y la orientación de política pública del documento.

El documento se elaboró en el marco de la Alianza Digital Unión Europea-América Latina y el Caribe y contó con el financiamiento de la Unión Europea, a través de la estrategia Global Gateway.

Las Naciones Unidas y los países que representan no son responsables por el contenido de vínculos a sitios web externos incluidos en esta publicación.

No deberá entenderse que existe adhesión de las Naciones Unidas o los países que representan a empresas, productos o servicios comerciales mencionados en esta publicación.

Ni la Unión Europea ni ninguna persona que actúe en su nombre es responsable del uso que pueda hacerse de la información contenida en esta publicación. Los puntos de vista expresados en este estudio son de los autores y no reflejan necesariamente los puntos de vista de la Unión Europea.

Las opiniones expresadas en este documento, que no ha sido sometido a revisión editorial, son de exclusiva responsabilidad de los autores y pueden no coincidir con las de las Naciones Unidas o las de los países que representan.

Publicación de las Naciones Unidas
LC/TS.2026/117
Distribución: L
Copyright © Naciones Unidas, 2026
Todos los derechos reservados
Impreso en Naciones Unidas, Santiago
S.2600465[S]

Esta publicación debe citarse como: Katz, R., Callorda, F. y Valencia, R. (2026). La cadena de valor de la inteligencia artificial: tensiones geopolíticas y la carrera por la competitividad de América Latina y el Caribe en el mundo. *Documentos de Proyectos* (LC/TS.2026/117). Comisión Económica para América Latina y el Caribe.

La autorización para reproducir total o parcialmente esta obra debe solicitarse a la Comisión Económica para América Latina y el Caribe (CEPAL), División de Documentos y Publicaciones, publicaciones.cepal@un.org. Los Estados Miembros de las Naciones Unidas y sus instituciones gubernamentales pueden reproducir esta obra sin autorización previa. Solo se les solicita que mencionen la fuente e informen a la CEPAL de tal reproducción.

Índice

Resumen ejecutivo 9

Introducción 17

I. La taxonomía y organización industrial de la inteligencia artificial 19
A. Definición y marco teórico para el estudio de la cadena de valor de la inteligencia artificial 19
B. Del ecosistema digital a la cadena de valor de la inteligencia artificial 20
1. Primer período: segmentación de cadenas de valor alrededor de telecomunicaciones y contenidos (hasta 1995-2005) 21
2. Segundo período: expansión de plataformas y OTT (2005-2022) 22
3. Tercer período: irrupción de la inteligencia artificial y el poder de cómputo (2023 - ¿?) 23
C. Estructura de la cadena de valor de la inteligencia artificial 27
1. Eslabones de la cadena de valor de inteligencia artificial 28
2. Insumos intermedios de los eslabones de la cadena de valor: aspectos críticos 31
3. El papel de las telecomunicaciones 34

II. La cadena de valor de la inteligencia artificial en Estados Unidos: estructura, dinámicas y dimensionamiento 37
A. La cadena de valor de la IA en Estados Unidos 37
1. Posiciones originales en la cadena de valor 38
2. Movimientos de integración vertical en la cadena de valor de la inteligencia artificial de Estados Unidos 47
3. Los puntos de presión en la cadena de valor de inteligencia artificial 56
B. Dimensionamiento económico de la cadena de valor de la inteligencia artificial en Estados Unidos 64
1. Ingresos 66
2. Costos operativos 68
3. Rentabilidad 69

4. Inversión de capital ... 72
5. Capitalización de mercado ... 74
6. Conclusión ... 76

III. La cadena de valor de la inteligencia artificial en China y Europa: estructura y dinámicas geopolíticas ... 81
A. La cadena de valor de la inteligencia artificial en la República Popular China ... 81
1. Estructura de la cadena de valor de la IA en China ... 82
2. Movimientos de integración vertical en la cadena de valor de la inteligencia artificial en China ... 94
3. Los puntos de presión en la cadena de valor de inteligencia artificial china ... 95
B. La cadena de valor de la inteligencia artificial en Europa ... 96
1. Movimientos de integración vertical en la cadena de valor de la inteligencia artificial en Europa ... 106
2. Los puntos de presión en la cadena de valor de inteligencia artificial en Europa ... 107
C. Dimensiones geopolíticas de la cadena de valor de la inteligencia artificial ... 108

IV. La cadena de valor de la inteligencia artificial en América Latina ... 111
A. Estructura de la cadena de valor ... 111
1. Equipamiento ... 112
2. Infraestructura de cómputo de la IA ... 113
3. Adquisición y preparación de bases de datos ... 116
4. Modelos fundacionales ... 117
5. Software y herramientas ... 120
6. Desarrollo e implementación de aplicaciones ... 121
7. Los puntos de presión en la cadena de valor de inteligencia artificial en América Latina ... 124
B. Dimensionamiento económico de la cadena de valor de la inteligencia artificial en América Latina ... 127
1. Equipamiento ... 127
2. Infraestructura de cómputo de la IA ... 128
3. Preparación y gestión de datos ... 131
4. Modelos fundacionales ... 131
5. Plataformas de software y herramientas ... 132
6. Desarrollo e implementación de aplicaciones ... 134
7. Conclusión ... 136

V. Elementos de política de desarrollo productivo para la inteligencia artificial en América Latina y el Caribe ... 139
A. Posicionamiento estratégico y especialización de empresas digitales en la cadena de valor ... 141
1. Posicionamiento estratégico en eslabones con empresas regionales ... 141
2. Especialización en verticales productivos ... 143
3. Consolidación de empresas regionales operando en mercados fragmentados ... 144
4. Madurez de mercados específicos para ensamblaje de hardware ... 146
5. Articulación regional para la negociación estratégica de materiales críticos ... 147
B. Escalamiento empresarial y consolidación productiva ... 148
1. Promoción de la inversión, innovación y exportación ... 148
2. Consolidación de clústeres y alianzas para potenciar el tejido empresarial ... 149
3. Desarrollo de campeones regionales en la. cadena de valor de la IA ... 150
4. Especialización funcional dentro de los eslabones priorizados ... 151

C. Financiamiento, modelo de negocios y acceso a mercados 152
1. Brecha de financiamiento en desarrollo de producto 152
2. Financiamiento para escalamiento 153
3. Estrategia exportadora de servicios tecnológicos 154
D. Infraestructura, cómputo y derrames 156
1. Aprovechamiento estratégico de infraestructura internacional 156
2. Acceso eficiente y asequible a capacidades de cómputo para IA 157
3. Desarrollo de modelos soberanos 158
4. Alineamiento geopolítico selectivo 160
E. Talento, conocimiento aplicado y desarrollo tecnológico 163
1. Formación de talento aplicado a perfiles de diseño y desarrollo 163
2. Retención de talento y masa crítica regional 164
3. Vinculación entre empresa y universidad 165
4. Desarrollo tecnológico con orientación comercial 167
F. Capacidades institucionales transversales habilitantes 168
1. Regulación habilitante para la innovación y adopción confiable de IA 168
2. Gobernanza de datos y espacios de datos sectoriales 170
3. Capacidades institucionales TOPP para implementar una política productiva de IA 171
G. Síntesis de recomendaciones de política 172

VI. Conclusiones 177

Bibliografía 181

Anexos 195
Anexo A1 196
Anexo A2 198
Anexo A3 202
Anexo A4 204
Anexo A5 207

Cuadros

Cuadro 1 Cadena de valor de la IA: valor de mercado por eslabón de empresas líderes globales, 2025 12
Cuadro 2 Ejes de política y orientación estratégica 15
Cuadro 3 Comparación entre segundo y tercer períodos de evolución de la cadena de valor 24
Cuadro 4 Valor de mercado sectorial de empresas en el ecosistema digital, a marzo de cada año 25
Cuadro 5 Participación en tráfico total de Chatbots, diciembre de 2025 46
Cuadro 6 Alianzas estratégicas de los principales participantes en la cadena valor para entrar en negocios adyacentes o complementarios 51
Cuadro 7 Ventas y CAPEX de jugadores provenientes de otras cadenas adyacentes (2025) 59
Cuadro 8 Cadena de Valor de IA (Estados Unidos): ingresos totales, 2025 77
Cuadro 9 Cadena de Valor de IA (Estados Unidos): promedio de costos operativos/ingresos, 2025 77
Cuadro 10 Rentabilidad (Estados Unidos): margen EBITDA por eslabón, 2025 78
Cuadro 11 Inversión, Estados Unidos, 2025 78
Cuadro 12 Cadena de valor de la IA en Estados Unidos: valor de mercado por eslabón, 2025 78

Cuadro 13 Poder de procesamiento de semiconductores usados en IA 83
Cuadro 14 China: empresas privadas para el desarrollo de GPUs 85
Cuadro 15 China: cuota de mercado de proveedores de servicios de cloud 89
Cuadro 16 China: principales modelos fundacionales 92
Cuadro 17 China: estrategia de IA de operadores de telecomunicaciones 94
Cuadro 18 Europa: lista de factorías de IA 99
Cuadro 19 Europa: modelos fundacionales de IA 102
Cuadro 20 América Latina y el Caribe: centros de datos, 2015-2025 113
Cuadro 21 Países de América Latina: visitantes a principales sitios de herramientas de IA, noviembre 2025 119
Cuadro 22 Países de America Latina: distribución de herramientas y empresas de IA 122
Cuadro 23 Posicionamiento en el ranking de instituciones de formación de IA, 2025 126
Cuadro 24 Ingresos de proveedores de infraestructura en América Latina, 2025 129
Cuadro 25 Inversiones en América Latina en el despliegue de centros de datos por proveedores de servicios de la nube, 2023-2025 129
Cuadro 26 Dimensionamiento económico de Scala Data Centers, 2024 130
Cuadro 27 América Latina: Métricas de la cadena de valor de inteligencia artificial 131
Cuadro 28 Dimensionamiento económico de Globant, 2025 135
Cuadro 29 América Latina: métricas de eslabones de Plataformas de software y herramientas 136
Cuadro 30 Métricas por eslabón de la cadena de la IA en América Latina, 2025 137
Cuadro 31 América Latina: número de empresas en el ecosistema de IA por país, 2025 144
Cuadro 32 Matriz de alineamiento geopolítico selectivo por eslabón de la cadena de valor de la IA en ALC 162
Cuadro 33 Resumen de objetivos, estrategias y líneas de acción de políticas enfocadas al desarrollo de las industrias asociadas a la inteligencia artificial 173
Cuadro A1.1 Equipamiento: ingresos, 2025 196
Cuadro A1.2 Infraestructura: ingresos, 2025 196
Cuadro A1.3 Preparación y gestión de datos: ingresos, 2025 196
Cuadro A1.4 Modelos fundacionales: ingresos, 2025 197
Cuadro A1.5 Plataformas de software y herramientas: ingresos, 2025 197
Cuadro A1.6 Desarrollo e implementación de aplicaciones: ingresos, 2025 197
Cuadro A2.1 Equipamiento: costos operativos, 2025 198
Cuadro A2.2 Infraestructura: costos operativos, 2025 199
Cuadro A2.3 Preparación y gestión de datos: costos operativos, 2025 199
Cuadro A2.4 Modelos fundacionales: costos operativos, 2025 200
Cuadro A2.5 Plataformas de software y herramientas: costos operativos, 2025 200
Cuadro A2.6 Desarrollo e implementación de aplicaciones: costos operativos, 2025 201
Cuadro A3.1 Infraestructura: rentabilidad, 2025 202
Cuadro A3.2 Preparación y gestión de datos: rentabilidad, 2025 202
Cuadro A3.3 Modelos fundacionales: rentabilidad, 2025 203
Cuadro A3.4 Plataformas de software y herramientas: rentabilidad, 2025 203
Cuadro A3.5 Desarrollo e implementación de aplicaciones: rentabilidad, 2025 203
Cuadro A4.1 Equipamiento: intensidad de capital, 2025 204
Cuadro A4.2 Infraestructura: intensidad de capital, 2025 204
Cuadro A4.3 Preparación y gestión de datos: Intensidad de capital, 2025 205
Cuadro A4.4 Modelos Fundacionales: intensidad de capital, 2025 205
Cuadro A4.5 Plataformas de software y herramientas: intensidad de capital, 2025 206
Cuadro A4.6 Desarrollo e implementación de aplicaciones: intensidad de capital, 2025 206
Cuadro A5.1 Equipamiento: capitalización de mercado, marzo 2026 207
Cuadro A5.2 Infraestructura: capitalización de mercado, marzo 2026 207

Cuadro A5.3 Preparación y gestión de datos: capitalización de mercado, marzo 2026 207
Cuadro A5.4 Modelos fundacionales: capitalización de mercado, marzo 2026 208
Cuadro A5.5 Plataformas de software y herramientas: capitalización de mercado, marzo 2026 208
Cuadro A5.6 Desarrollo e implementación de aplicaciones: capitalización de mercado, marzo 2026 208

Gráficos

Gráfico 1 Desempeño de modelos cerrados vs. abiertos 45
Gráfico 2 Costo promedio de construcción y operación de un centro de cómputo por un ciclo de operaciones de 20 años 57
Gráfico 3 Estados Unidos: consumo energético de centros de datos (2025-2030) 58
Gráfico 4 OpenAI: proyección de ingresos (1T2025 vs. 3T2025) 60
Gráfico 5 Estados Unidos: adopción de IA por empresas, 2022 63
Gráfico 6 Ingresos por eslabón de la cadena de valor de la IA en Estados Unidos, 2025 67
Gráfico 7 Margen EBITDA por eslabón la cadena de valor de la IA en Estados Unidos, mínimos, máximos y promedio 71
Gráfico 8 CAPEX por eslabón de la cadena de valor, 2025 74
Gráfico 9 Capitalización de mercado consolidada por etapa de la cadena de valor de la IA en Estados Unidos, marzo de 2026 76
Gráfico 10 China: oferta y demanda de chips de inferencia de IA, 2025-2030 84
Gráfico 11 Capacidad comparada de manufactura de chips de AI, 2025-2023 86
Gráfico 12 China: zonas de proveedores de servicios de computación en la nube, 2011-2025 90
Gráfico 13 Europa: zonas de proveedores de servicios de computación en la nube, 2011-2025 98
Gráfico 14 América Latina y el Caribe: despliegue de Zonas de proveedores de servicios en la nube, 2011-2025 114
Gráfico 15 Superficie actual y proyectada de centros de datos 116

Diagramas

Diagrama 1 Estructura de la cadena de valor de IA (incluyendo insumos intermedios) 9
Diagrama 2 Movimientos de integración vertical simultáneos en la cadena de valor de IA 11
Diagrama 3 América Latina, cadena de valor de IA y la captura de valor regional 13
Diagrama 4 Primer período: cadenas de valor de telecomunicaciones y contenidos audiovisuales 21
Diagrama 5 Segundo período: cadena de valor del ecosistema digital 22
Diagrama 6 Transición de la cadena de valor del ecosistema digital a la de inteligencia artificial 23
Diagrama 7 Estructura del "AI stack" (ejemplo) 26
Diagrama 8 Evolución del *Technology Stack* 27
Diagrama 9 Estructura de la cadena de valor de IA 27
Diagrama 10 Funciones del eslabón de infraestructura 30
Diagrama 11 Estructura de la cadena de valor de IA (incluyendo insumos intermedios) 32
Diagrama 12 Posicionamiento original de firmas en la cadena de valor de la IA 38
Diagrama 13 Inversiones cruzadas de NVIDIA y OpenAI 52
Diagrama 14 Movimientos de integración vertical simultáneos en la cadena de valor de IA 55
Diagrama 15 Precios de ChatGPT (2026) 59
Diagrama 16 Equipamiento: propiedad intelectual versus ensamblaje 70
Diagrama 17 China: estructura de la cadena de valor de la IA 82
Diagrama 18 Europa: cadena de valor de IA 97
Diagrama 19 Objetivos geoestratégicos de tres polos mundiales de desarrollo de la inteligencia artificial 108

Diagrama 20 Estrategias de posicionamiento geopolíti 110
Diagrama 21 América Latina y el Caribe: cadena de valor de IA 112
Diagrama 22 Cadena de valor de la IA en América Latina y el Caribe: posicionamiento regional y espacios de intervención de política 141

Recuadro

Recuadro 1 Aspectos particulares que condicionan el desarrollo de LATAMGPT 125

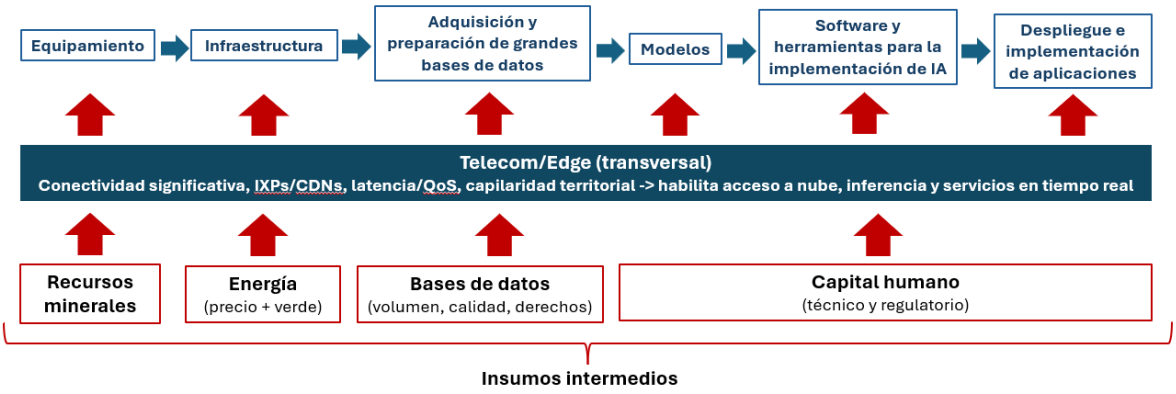
Mapa

Mapa 1 Estados Unidos: mapa de centros de datos desplegados o por desplegar para procesamiento de IA 56

Resumen ejecutivo

Este estudio analiza, modela y cuantifica la transformación de la cadena de valor de la industria digital provocada por la irrupción de la inteligencia artificial (IA), identificando los mecanismos de captura de valor y sus efectos sobre la organización industrial. Aunque parte de las tendencias internacionales, su propósito es generar un marco analítico que permita determinar la posición de América Latina y el Caribe en dicha cadena y definir políticas públicas para que la región construya una posición competitiva a nivel global.

Diagrama 1
Estructura de la cadena de valor de IA (incluyendo insumos intermedios)



Fuente: Elaboración propia.
Nota: Cada eslabón depende de insumos intermedios cuyo acceso condiciona costos y capacidades. Las telecomunicaciones, antes un eslabón central del ecosistema digital, operan ahora como insumo intermedio.

La cadena de valor de la IA se organiza hoy en seis eslabones encadenados, desde los insumos básicos hasta el usuario final: (1) equipamiento (semiconductores y ensamblaje de servidores), (2) infraestructura de cómputo (centros de datos y nube), (3) preparación y gestión de datos, (4) modelos fundacionales, (5) plataformas de software y herramientas y (6) desarrollo e implementación de

aplicaciones. A lo largo del documento, los "eslabones iniciales" aluden a los primeros (equipamiento e infraestructura) y los "eslabones finales" a los últimos (software, herramientas y aplicaciones). La incorporación de los insumos intermedios —energía, minerales críticos, datos, talento y conectividad— permite identificar los cuellos de botella que condicionan los costos y las capacidades de cada eslabón.

Mensajes clave

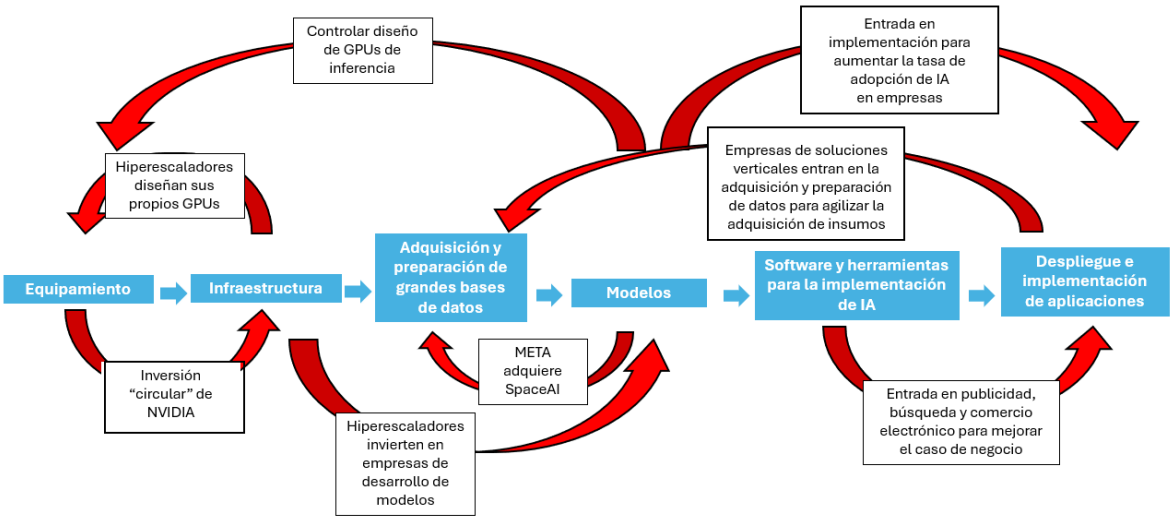
- **La captura de valor se concentra en los eslabones iniciales.** El equipamiento (semiconductores y servidores) y la infraestructura de cómputo reúnen cerca del 86% de la valoración de mercado de la cadena y el 95% de sus ingresos, reflejando rentas asociadas a la exclusividad tecnológica y el control de activos escasos.
- **La cadena se reorganiza por integración vertical.** Los grandes actores se expanden hacia adelante y hacia atrás (cómputo, modelos, semiconductores, datos). El riesgo financiero se eleva porque buena parte de la inversión en centros de datos y GPUs se financia con deuda, los ratios de inversión sobre ingresos son muy altos y existen compromisos cruzados o "circulares" entre proveedores y clientes, lo que abre interrogantes sobre la viabilidad de algunos planes de inversión.
- **La IA es ya un terreno de competencia geopolítica.** Estados Unidos, China y Europa despliegan estrategias divergentes —liderazgo privado, cadena soberana paralela y regulación pionera— que condicionan el espacio de maniobra de América Latina y el Caribe.
- **América Latina y el Caribe captura valor en los eslabones finales.** La región muestra capacidades reales en software, herramientas, integración y aplicaciones sectoriales, con presencia extranjera dominante en infraestructura y un desarrollo embrionario en modelos soberanos.
- **Existe una brecha crítica en modelos fundacionales.** El financiamiento de iniciativas soberanas regionales es órdenes de magnitud inferior al de referentes comparables, lo que orienta la estrategia hacia la especialización vertical antes que hacia modelos de propósito general.
- **La soberanía digital se juega en la capacidad de decisión, no en la autarquía.** Avanzar en soberanía digital no implica desarrollar modelos de frontera propios en cada país, sino construir las capacidades institucionales para decidir con autonomía dónde y cómo insertarse en la cadena: negociar en mejores términos con los proveedores tecnológicos, gobernar el acceso a los datos estratégicos, evaluar riesgos en los sectores críticos y cooperar a nivel regional. Sobre esa base, la región puede especializarse en los eslabones donde tiene capacidades y escalar sus empresas, en lugar de dispersar esfuerzos intentando replicar toda la cadena.

A. Una cadena de valor en transformación

La cadena de valor del ecosistema digital no es estática. Ha atravesado tres períodos: la segmentación inicial entre telecomunicaciones y contenidos audiovisuales; la convergencia impulsada por la masificación de Internet y la emergencia de las plataformas de agregación y *Over-the-Top* (OTT); y, desde el lanzamiento público de ChatGPT en octubre de 2022, una nueva configuración centrada en la IA y la computación en la nube. En esta etapa, quienes operan modelos fundacionales y proveen capacidad de cómputo se convierten en actores clave, mientras que las telecomunicaciones pasan a funcionar como insumo intermedio que habilita la interacción entre eslabones.

En su origen, la cadena estaba estructurada en torno a firmas que ocupaban posiciones específicas. Casi de inmediato, esas empresas iniciaron movimientos de integración vertical —hacia adelante y hacia atrás— motivadas por el control estratégico de etapas clave, la garantía de adopción de sus productos, la reducción de costos de transacción y la búsqueda o preservación de posiciones de rentabilidad superior.

Diagrama 2
Movimientos de integración vertical simultáneos en la cadena de valor de IA



Fuente: Elaboración propia.

Estos movimientos y los voluminosos planes de inversión plantean interrogantes sobre su viabilidad económica y su posible inestabilidad. El despliegue de infraestructura eleva el perfil de riesgo por su escala, plazos y sofisticación técnica; ciertos actores *pure play* enfocados exclusivamente en IA enfrentan presiones de financiamiento; y la construcción de centros de datos, apoyada parcialmente en instrumentos de deuda, puede afectar los costos de capital y las estrategias de salida de los inversionistas.

Dónde se crea y se captura el valor

El dimensionamiento económico de la cadena en Estados Unidos revela una asimetría estructural: el valor se concentra donde existe exclusividad tecnológica y poder de fijación de precios.

- **Ingresos.** Los eslabones iniciales concentran la mayor parte de la facturación. El equipamiento (semiconductores y ensamblaje de servidores) reúne cerca del 61% de los ingresos.
- **Costos operativos.** La IA impone una estructura de costos recurrente que rompe el paradigma de costo marginal cercano a cero del software tradicional, donde cada usuario adicional cuesta casi nada. En la IA cada interacción consume cómputo y energía en tiempo real (la "inferencia"). Por eso el costo marginal no tiende a cero y la presión es especialmente alta en infraestructura y en el entrenamiento de modelos fundacionales.
- **Rentabilidad.** Coexisten márgenes muy altos en los nodos críticos —el diseño y la manufactura de GPUs (NVIDIA y TSMC)— con una rápida pérdida de margen en los eslabones que se estandarizan; los laboratorios de modelos fundacionales, por su parte, operan a pérdida para financiar su desarrollo y su adopción.
- **Inversión de capital.** La intensidad de capital es máxima en la infraestructura de nube y en la adquisición de cómputo para el entrenamiento de modelos, con ratios CAPEX/ingresos del 77% y 88%, respectivamente.
- **Capitalización de mercado.** El equipamiento concentra el 67,4% de la valoración total de la cadena, seguido por la infraestructura (18,6%) y los laboratorios de modelos fundacionales (9,9%); los eslabones restantes registran valoraciones marcadamente menores (cuadro 1).

Cuadro 1
Cadena de valor de la IA: valor de mercado por eslabón de empresas líderes globales, 2025

Eslabón	Empresas	Valoración (*En millones de dólares*)	Porcentaje
1. Equipamiento	NVIDIA[a] AMD[a] TSMC[a] Foxconn[b] Supermicro[b] Dell Technologies[b]	9 097 000	67,4
2. Infraestructura de cómputo para IA	AWS[c] Microsoft Azure[c] Google Cloud[c] Oracle[c] CoreWeave[a]	2 515 529	18,6
3. Preparación y gestión de datos	ScaleAI[d] Gretel.aie[e] Pinecone[d]	31 000	0,2
4. Modelos fundacionales	OpenAI[f] Anthropic[f] xAI[d]	1 330 000	9,9
5. Plataformas de software y herramientas	Databricks[d] Hugging Face[d] LangChain[d] Suno[d]	142 200	1,1
6. Desarrollo e implementación de aplicaciones	Palantir[a] Harvey[d] Abridge[d]	375 232	2,8
Total		13 490 961	100

Fuente: Elaboración propia.
[a] Valoración bursátil compilada de Yahoo Finance.
[b] Incluye la valoración completa de la empresa y no solamente aquella asignada al ensamblaje de servidores IA.
[c] Valoración bursátil total de Yahoo Finance ponderada por ingresos en segmento de la nube, lo que podría subestimar el valor asignado a los servicios de la nube.
[d] Valoración estipulada a partir de la última ronda de financiamiento.
[e] Precio de adquisición.
[f] Valoración pre-oferta pública (Fuente: Reportes públicos).

La dimensión geopolítica

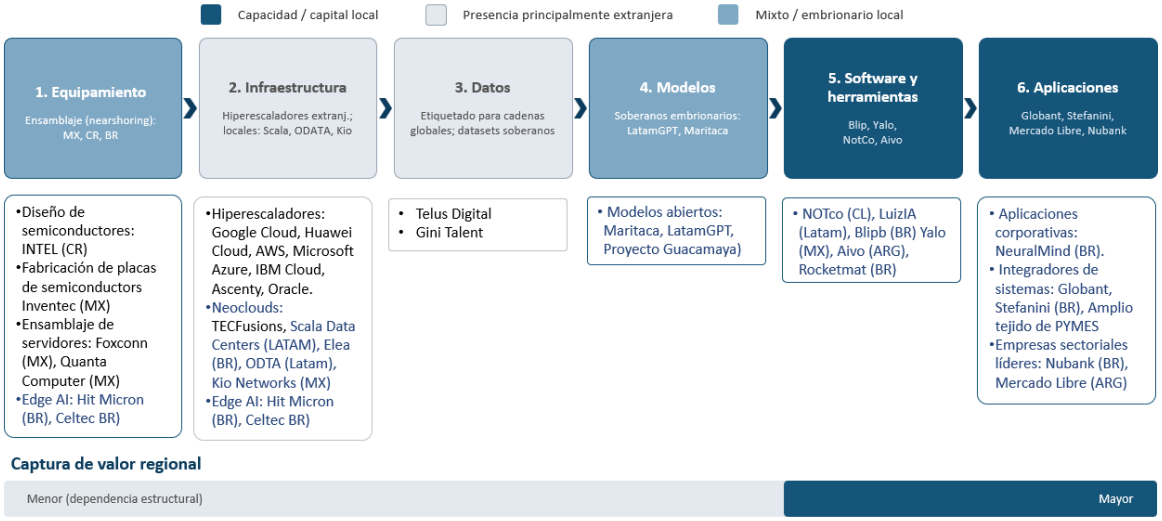
El desarrollo y despliegue de la IA es un terreno de competencia entre naciones, con implicancias económicas, políticas y militares. Tres polos articulan estrategias divergentes:

- **Estados Unidos** apuesta por el liderazgo del sector privado, busca remover barreras regulatorias a la innovación y asegura el acceso a insumos intermedios, al tiempo que despliega controles sobre el aprovisionamiento tecnológico —en especial de GPUs— para contener el avance chino.
- **China** no busca posicionarse en unos pocos eslabones, sino desplegar una cadena de valor paralela y completa con factores de producción soberanos, replicando su trayectoria en industrias de capital intensivo, con el "sur global" como mercado objetivo.
- **Europa** —como alianza de "middle powers"— vincula su cadena a un marco regulatorio pionero (EU AI Act y Plan de Acción de IA), orientado a tecnologías fiables que preserven los valores democráticos y la privacidad mientras mejora la competitividad de sus firmas.

B. América Latina y el Caribe en la cadena de la IA

La región incluye países de desarrollo medio y naciones en desarrollo. Economías como Brasil, Chile y Uruguay pueden apalancar sistemas educativos de calidad y ecosistemas de innovación para jugar un papel relevante, aunque ello exige aumentar significativamente la inversión. La cadena de valor de la IA en la región muestra avances en el desarrollo local de software, herramientas e implementación de sistemas; una presencia principalmente extranjera en infraestructura de cómputo; y un movimiento embrionario en modelos fundacionales soberanos.

Diagrama 3
América Latina, cadena de valor de IA y la captura de valor regional



Fuente: Elaboración propia.
Nota: Texto en azul indica desarrollo/capital local.

El dimensionamiento económico regional revela un ecosistema en transición, con una marcada asimetría entre la inversión en activos físicos y la generación de ingresos por propiedad intelectual, y márgenes inferiores a los del mercado estadounidense:

- **Infraestructura como cimiento.** La región consolidó una base material robusta, pasando de 95 a 475 centros de datos en una década, con 53 zonas cloud y 20 accesos Edge en 2025. Empresas locales como Scala Data Centers (US$ 4.000 millones) y ODATA (US$ 1.850 millones) demuestran que existe valor en esta etapa, aunque con márgenes EBITDA (33%–37%) inferiores a los de mercados de frontera.
- **Modelos fundacionales y especialización.** Aun con presupuestos muy inferiores a los de sus pares (LatamGPT ronda los US$ 0,80 millones; AMALIA, US$ 5,5; ALIA, US$ 45; SEA-LION, US$ 65), la región ya desarrolla modelos soberanos. La oportunidad mayor está en la especialización vertical: afinar y adaptar modelos con datos locales en sectores donde la región tiene ventajas, generando capacidades propias sin depender de igualar la inversión de la frontera.
- **"Unicornios" en plataformas y herramientas.** Es el eslabón de mayor eficiencia en la captura de valor. Empresas como Blip, Yalo y NotCo superan los US$ 100–150 millones de ingresos anuales y exportan propiedad intelectual, capturando el "margen intermedio" entre los modelos y el usuario final, adaptados a los canales preferidos en la región (WhatsApp/Messenger).

- **La IA como multiplicador de la economía real.** El mayor impacto no proviene de las empresas que venden IA, sino de las que la utilizan para transformar sectores tradicionales: Mercado Libre (comercio electrónico y logística) y Nubank (finanzas) logran eficiencias inalcanzables sin IA.
- **El integrador como exportador de talento.** Firmas como Globant y Stefanini (Woopi) muestran que el capital humano avanzado es el activo más resiliente de la cadena local, con ingresos atribuidos a la IA que superan los US$ 350 millones en casos individuales.

Entre los puntos de presión propios de la región destacan la subinversión que limita el desarrollo de modelos soberanos y la fragmentación institucional que diluye los esfuerzos nacionales. Más allá de Brasil, México y Chile —con capacidades más completas a lo largo de la cadena—, el resto de los países encuentra mayor viabilidad en la especialización vertical, afinando modelos y construyendo aplicaciones en sectores donde la región dispone de datos y ventajas comparativas, como agricultura, biodiversidad, minería y fintech.

Estos tres elementos —el desarrollo embrionario de modelos soberanos, la dependencia de infraestructura de cómputo controlada por actores externos y la dotación de minerales críticos— configuran la cuestión de la soberanía digital regional. En este estudio, soberanía no equivale a reproducir de forma integral los eslabones de frontera (la vía que sigue China y que resulta inviable para la región), sino a preservar la capacidad de decisión sobre los esquemas de aprovechamiento de los activos controlados por terceros: qué se adopta, bajo qué reglas, con qué datos y con qué grado de control en los usos críticos.

C. Lineamientos de política

Para construir una posición competitiva, la política pública debe priorizar antes que abarcar: concentrar el esfuerzo en los eslabones y verticales viables en lugar de extenderlo a toda la cadena. La estrategia se organiza como una secuencia coherente: seleccionar eslabones y verticales viables; escalar empresas y consolidar campeones regionales; resolver las restricciones de financiamiento, internacionalización y acceso a mercados; y sostener el esfuerzo con infraestructura, cómputo, talento y transferencia tecnológica. Cinco ejes de política, articulados por tres condiciones institucionales transversales, ordenan esa estrategia (cuadro 2).

A ello se suman dos orientaciones específicas. En materia de minerales críticos, la región debería avanzar desde la exportación de recursos en bruto hacia una articulación regional que permita negociar mejores condiciones de intercambio, procesamiento, transferencia tecnológica y acceso preferente a segmentos más avanzados de la cadena. En el plano geopolítico, conviene adoptar una estrategia de diversificación de alianzas y alineamientos selectivos según el tipo de capacidad que se busque desarrollar.

Cuadro 2
Ejes de política y orientación estratégica

Eje de política	Orientación estratégica
Posicionamiento y especialización	Especializarse en los eslabones con capacidades observables (software, integración, aplicaciones sectoriales); construir escala mediante clústeres, consorcios y redes de colaboración; priorizar verticales con activos diferenciales, datos y demanda empresarial identificable.
Escalamiento empresarial y consolidación productiva	Fortalecer campeones regionales y la especialización funcional; consolidar el tejido de firmas pequeñas mediante plataformas de articulación, servicios compartidos y proyectos con empresas ancla.
Financiamiento, modelo de negocio y acceso a mercados	Estructurar instrumentos públicos y de coinversión que cubran desde la prueba de concepto hasta el escalamiento; estimular la I+D privada con incentivos fiscales y crédito; construir plataformas exportadoras de servicios tecnológicos.
Infraestructura, cómputo y derrames	Aprovechar la infraestructura ya instalada (475 centros de datos, 551 MW hacia 1,6 GW en 2031) en lugar de duplicarla; habilitar acceso compartido y subsidiado a cómputo para startups y pymes; impulsar modelos soberanos de especialización vertical.
Talento, conocimiento aplicado y desarrollo tecnológico	Anticipar y monitorear la demanda de perfiles de diseño y desarrollo; concentrar y retener talento avanzado en centros de excelencia regionales; vincular universidad y empresa con orientación comercial.
Condiciones institucionales transversales habilitantes	
Capacidades TOPP · Gobernanza de datos · Regulación habilitante	Capa común que atraviesa todos los eslabones: marcos regulatorios proporcionales al riesgo y sandboxes; espacios de datos sectoriales con reglas de acceso e interoperabilidad; fortalecimiento de capacidades institucionales (diagnósticos TOPP, coordinación interministerial, prospectiva).

Fuente: Elaboración propia, sobre la base del capítulo V del documento.

En síntesis, el desarrollo de una industria regional de IA dependerá de la articulación simultánea —no aislada— de las políticas de desarrollo productivo, innovación, financiamiento, infraestructura, talento, gobernanza de datos y regulación habilitante. La región no debe aspirar a reproducir la frontera global, sino a construir un ecosistema con densidad productiva e institucional suficiente para escalar las capacidades que ya posee y convertirlas en una base estable de competitividad tecnológica. Para América Latina y el Caribe, especializarse es la manera realista de ejercer soberanía digital: concentrarse en los eslabones donde puede competir le permite conservar capacidad de decisión sobre las tecnologías, los datos y los activos estratégicos que definen la productividad de sus industrias clave.

Introducción

El siguiente estudio tiene por objeto analizar, modelar y cuantificar la transformación de la cadena de valor de la industria digital ocasionada por la irrupción de la inteligencia artificial (IA), identificando los mecanismos de captura de valor y sus efectos sobre la organización industrial. Si bien el estudio comienza analizando estas tendencias a nivel internacional, el objetivo es generar un marco de análisis que permita determinar la posición de América Latina y el Caribe en dicha cadena y definir políticas públicas para que la región construya una posición competitiva global.

La irrupción de la inteligencia artificial está modificando de manera profunda la estructura del ecosistema digital, al introducir nuevas capas tecnológicas, reordenar las relaciones entre actores y desplazar los puntos de creación y apropiación de valor. A diferencia de olas previas de digitalización, la IA no solo agrega nuevas funcionalidades a plataformas, redes o servicios existentes, sino que reorganiza la articulación entre infraestructura, datos, modelos, herramientas y aplicaciones. Por ello, analizar el cambio en la cadena de valor del ecosistema digital resulta necesario para comprender cómo se reconfigura la organización industrial, cuáles son los nuevos factores estratégicos de competitividad y de qué manera se redistribuyen las oportunidades y dependencias entre países, empresas y regiones.

El análisis de la cadena de valor de la inteligencia artificial permite entender cómo esta tecnología reconfigura la organización industrial del ecosistema digital y cómo se distribuyen sus beneficios económicos entre distintos actores y eslabones. La IA no constituye un sector único, sino una secuencia de actividades interdependientes —desde el equipamiento y la infraestructura hasta los datos, los modelos, las herramientas y las aplicaciones— que presentan estructuras de mercado, requerimientos de inversión y oportunidades de aprendizaje muy diferentes. Por ello, el enfoque de cadena resulta indispensable para identificar los mecanismos de captura de valor, los puntos de concentración económica y los cuellos de botella que condicionan la competitividad.

Para América Latina y el Caribe, este marco analítico es particularmente importante porque permite distinguir entre los eslabones donde predominan dependencias estructurales y aquellos donde existen márgenes más factibles para construir capacidades, escalar actividades y orientar políticas públicas. Así, el análisis de la cadena de valor de la IA ofrece una base más precisa para vincular la

transformación digital con la agenda de desarrollo productivo, en la medida en que permite examinar no solo la adopción de la tecnología, sino también la posición que la región ocupa en la nueva estructura de creación y apropiación de valor.

El estudio está organizado en cinco grandes partes, cada una de ellas incluyendo varios capítulos. La primera parte está enfocada en el estudio de la estructura de la cadena de valor de la inteligencia artificial. En primer lugar, se analiza el proceso de desarrollo de esta cadena, vinculándola a las transformaciones ocurridas en los últimos veinte años en el ecosistema digital (capítulo III). El estudio de la cadena de valor de la inteligencia artificial analiza los diferentes eslabones que la componen, así como el uso de insumos intermedios para cada uno de ellos, definiendo la taxonomía actual compuesta por la manufactura de semiconductores especializados, el despliegue de infraestructura de centros de datos, el desarrollo de modelos fundacionales alimentados por las plataformas de datos, el desarrollo de aplicaciones y agentes, y la implementación en empresas (capítulo IV).

La segunda parte utiliza el marco teórico de la cadena de valor de la inteligencia artificial para estudiar su composición y las dinámicas de integración vertical en Estados Unidos (capítulo V). En ese contexto, el análisis identifica puntos de presión, es decir situaciones que podrían generar riesgos en términos de retornos o sostenibilidad financiera. El capítulo VI se adentra en el dimensionamiento económico por eslabón, identificando aquellas etapas de la cadena donde se está produciendo la mayor creación de valor, medida ésta por ingresos, rentabilidad, o capitalización de mercado.

La tercera parte retoma la estructura de la cadena, pero la aplica a la comprensión de dinámicas en la República Popular China (capítulo VII) y los países de Europa (capítulo VIII). Este análisis es similar al realizado en el caso estadounidense, identificando movimientos de integración vertical entre eslabones y puntos de presión posibles, resultando en riesgos económicos. Habiendo analizado China y Europa, es posible estudiar las dinámicas de competencia geopolítica entre ellas y Estados Unidos (capítulo IX).

La cuarta parte se enfoca la estructura de la cadena de valor y organización industrial de la cadena de valor de la IA en el contexto de América Latina y el Caribe. El capítulo X estudia la estructura de la cadena y sus actores principales mientras que el capítulo XI analiza su dimensionamiento económico.

Los análisis anteriores proveen el contexto que permite esbozar elementos de política de desarrollo productivo que apunten al desarrollo de la inteligencia artificial en la región, lo que es desarrollado en la quinta parte.

I. La taxonomía y organización industrial de la inteligencia artificial

La primera parte de este estudio está organizada en tres capítulos. El capítulo II introduce brevemente el marco teórico y definición del concepto de cadena de valor en el marco del análisis de organización industrial. El capítulo III estudia el proceso que llevó a la transformación del ecosistema digital resultando en la emergencia de la cadena de valor de la inteligencia artificial (IA). El capítulo IV define la estructura de la cadena de valor de la IA y cada uno de sus eslabones, yendo de la manufactura de GPUs [1], el despliegue de infraestructura de la nube, el desarrollo de modelos fundacionales, el despliegue de software y herramientas y el desarrollo de aplicaciones en mercados verticales.

A. Definición y marco teórico para el estudio de la cadena de valor de la inteligencia artificial

El análisis de la estructura de mercados está basado en el concepto de cadena de valor, tal como fuera definido por Stigler en su artículo publicado en 1951, donde el autor alude implícitamente a la integración y fragmentación de cadenas productivas [2]. Extendiendo el concepto de funciones de una firma tal y como lo define Adam Smith, Stigler lo aplica al funcionamiento de industrias para entender las dinámicas de especialización, integración y fragmentación vertical en relación con su ciclo de vida:

> "Si se considera el ciclo de vida de industrias, la desintegración vertical es de esperar. Las empresas en nuevas industrias no adhieren a esta organización. Dado que estas requieren nuevos tipos o cualidades de materiales o insumos, los fabrican internamente; para ello,

[1] GPUs o Graphics Processing Units son el componente esencial de la arquitectura de procesamiento de IA, debido a que su estructura de procesamiento paralelo les permite operar miles de tareas simultáneamente necesarias para el entrenamiento de modelos de IA y su inferencia.

[2] Stigler, G. (1951). "The division of labor is limited by the extent of the market", *The Journal of Political Economy*, Vol. 59, No. 3 (June), pp. 185-193 (Traducción del autor).

> deben superar problemas técnicos en el uso de sus productos y no pueden esperar a que los posibles usuarios los superen; deben persuadir a sus clientes para que abandonen otros productos. Estas industrias jóvenes deben diseñar sus equipos especializados y a menudo fabricarlo, y deben comprometerse a reclutar (históricamente, a menudo importando) mano de obra cualificada. Cuando la industria ha alcanzado cierto tamaño y perspectivas, muchas de estas tareas son lo suficientemente importantes como para ser entregadas a especialistas. Resulta rentable para otras empresas suministrar equipos y materias primas, encargarse de la comercialización del producto y la utilización de subproductos, e incluso formar mano de obra cualificada."

Es así como Stigler extiende conceptos como la división del trabajo usado para entender el funcionamiento de firmas (por ejemplo, cadena de aprovisionamiento, manufactura y distribución) a la dinámica de cadenas productivas. Este marco permite entender no solo el posicionamiento de diferentes jugadores en una cadena productiva sino también los respectivos modelos de negocio, dinámicas competitivas, alianzas estratégicas y modelos de integración vertical y horizontal [3].

Apuntando a entender la dinámica de fronteras entre firmas y mercados, Coase planteó el concepto de costos de transacción como variable explicativa de cambios en cadenas de valor [4]. A partir de ello, Williamson generó el análisis empírico, vinculando costos de transacción entre firmas especializadas en eslabones de la cadena de valor y la integración vertical. De acuerdo con este concepto, empresas operando en eslabones específicos de cadenas de valor tienden a integrarse verticalmente cuando el mecanismo de mercado se vuelve ineficiente debido a la alta especificidad de los activos y la incertidumbre contractual. Williamson consideró que la reintegración de la cadena de valor es ocasionada por dos tipos de factores. Por un lado, empresas tienden a integrarse verticalmente buscando el control estratégico de ciertos estadios de la cadena (por ejemplo, la relación con el usuario final), o buscando reducir los costos de transacción entre estadios de la cadena [5]. Los costos de transacción que pueden estimular tendencias a la integración vertical incluyen contratos indefinidos para la adquisición de insumos, imperfecciones en información, incentivos para el comportamiento ineficiente de firmas, y los costos y beneficios internos de organizaciones en relación con las "fronteras de eficiencia" de una empresa (por ejemplo, economías de escala o experiencia) [6].

Estos conceptos son fundamentales para entender la dinámica y evolución de la cadena de valor del ecosistema digital.

B. Del ecosistema digital a la cadena de valor de la inteligencia artificial

El estudio de la cadena de valor del ecosistema digital demuestra que, siguiendo a los autores citados arriba, esta no es estática, sino que se ha ido transformando en el tiempo, sugiriendo un desplazamiento del centro de gravedad de las industrias digitales y la influencia de tendencias a la integración vertical. Esta dinámica permite identificar tres periodos de evolución diferentes, cada uno de los cuales indica un cambio estructural en los actores dominantes, la lógica de organización industrial, y el foco tecnológico. Además de la determinación de cambios en los diferentes eslabones dominantes, cada período indica modificaciones en los activos determinantes de poder económico, las prioridades de inversión y el énfasis regulatorio.

---

[3] Este concepto no debe ser confundido con el estudiado en las denominadas "Cadenas de Valor Globales" donde si bien se estudia la composición de cadenas productivas, el enfoque de análisis es la distribución de eslabones entre países y regiones. Ver Gereffi, G. & Sturgeon, T. (2013), "Global value chain-oriented industrial policy: the role of emerging economies", *Global value chains in a changing world.*

[4] Coase, R. (1937). *The nature of the firm*, 4 Economica 386.

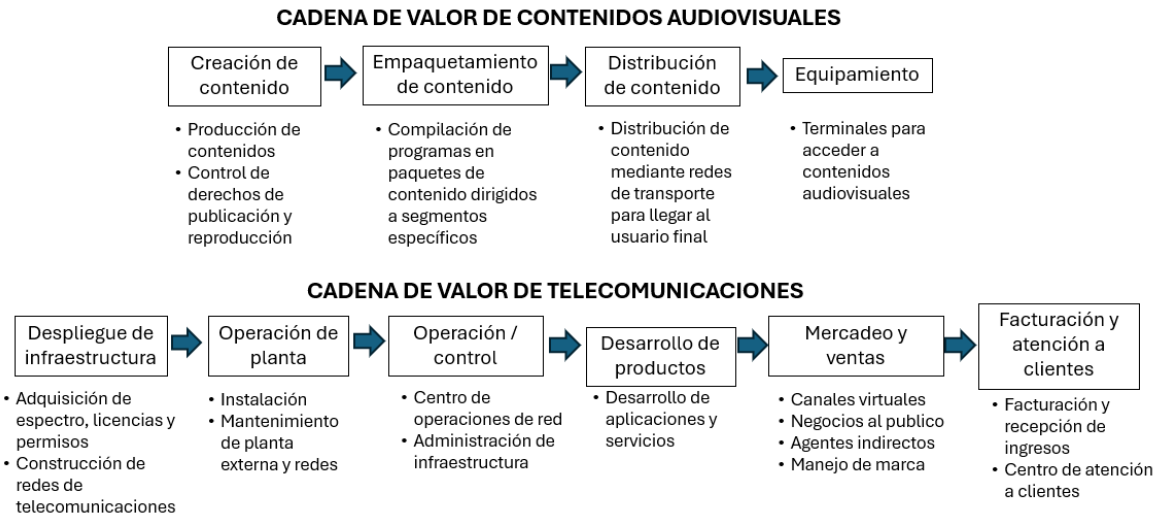
[5] Williamson, O. (1985). *The Economic Institutions of Capitalism*. New York: The Free Press, pp. 85-130.

[6] Para un análisis detallado de Williamson en relación a teorías neoclásicas, ver Joskow, P. "Vertical integration", *Issues in Antitrust Law and Policy* 273.

1. Primer período: segmentación de cadenas de valor alrededor de telecomunicaciones y contenidos (hasta 1995-2005)

Hasta mediados de la década del noventa, la oferta de servicios de telecomunicaciones y la distribución de contenidos audiovisuales se realizaba a partir de dos cadenas productivas independientes (ver diagrama 4).

Diagrama 4
Primer período: cadenas de valor de telecomunicaciones y contenidos audiovisuales



Fuente: Katz (2015).

Cada cadena incluía operadores que cumplían funciones específicas como transporte, o provisión de equipamiento de red o dispositivos de acceso. Las cadenas de valor que definían la estructura de cada uno de los dos sectores eran compartimientos estancos con pocas posibilidades de interrelación. Al mismo tiempo, muchas de las empresas que operaban en cada cadena de valor estaban integradas verticalmente, y ocupaban la mayor parte de los eslabones. Por ejemplo, los operadores de telecomunicaciones se encargaban de construir redes, operarlas, desarrollar productos y ofrecerlos al usuario. Por el lado de la distribución de contenidos, y tomando los contenidos audiovisuales como ejemplo, si bien existían empresas que se encargaban de producir programas, empaquetarlos y distribuirlos, también existían empresas con un foco en eslabones específicos: los estudios de cine y televisión producían programas, los canales empaquetaban contenidos con vistas a atraer una audiencia determinada, y los operadores de TV por cable se encargaban de distribuirlos. El cruce entre cadenas de valor (por ejemplo, operadores de cable trataban de entrar en el negocio de las telecomunicaciones o firmas de telecomunicaciones intentaban diversificarse en la distribución de contenido) ocurría ocasionalmente, aunque esto raramente resultaba en un éxito empresario. En este sentido, tanto los sectores de contenidos audiovisuales como los de las telecomunicaciones reconocían que existía una oportunidad de negocio asociada a la diversificación en cadenas de valor adyacentes, pero no encontraban las plataformas tecnológicas que les permitiesen hacerlo[7].

[7] Recordar, por ejemplo, que los primeros intentos de operadores de telecomunicaciones para entrar en el negocio de la distribución de contenidos estaban basados en "redes fijas hibridas" que constituían dos redes paralelas unidas físicamente en el mismo conducto.

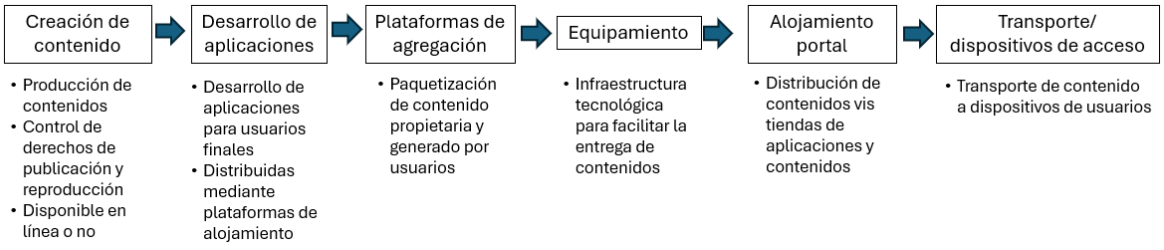
Bajo esta configuración de la cadena de valor, exceptuando la creación de contenidos audiovisuales, el centro de la cadena residía en la infraestructura física y la operación de redes, y el valor era capturado en acceso, interconexión y expansión de cobertura. El marco regulatorio relevante en este periodo estaba condicionado por conceptos tales como la universalización de servicios, la competencia en telecomunicaciones, la asignación de espectro, la interconexión entre redes, y el monitoreo de calidad y precios. Es importante mencionar que estos principios no desaparecen en las nuevas configuraciones de la cadena del ecosistema digital, sino que se agregan a nuevas consideraciones y objetivos de política pública.

2. Segundo período: expansión de plataformas y OTT (2005-2022)

La digitalización masiva de contenidos y plataformas de transporte, y la difusión de Internet ocasionaron un cambio fundamental en esta organización industrial, proveyendo un facilitador de la convergencia de las cadenas de valor. Consideramos la irrupción de Facebook y Google como instancias temporales del inicio de esta nueva etapa.

A partir de este momento, la segmentación original entre distribución de información y conectividad desaparece y las cadenas productivas de las industrias de medios, telecomunicaciones, y productos culturales se integran, generando una estructura de interdependencias única, que denominamos el ecosistema digital (ver diagrama 5).

Diagrama 5
Segundo período: cadena de valor del ecosistema digital



Fuente: Katz (2015).

En esta nueva serie de interrelaciones, el transporte (o sea las telecomunicaciones) ya no es una cadena productiva independiente, sino que se ha transformado en uno de los estadios de la nueva configuración integrada. Si bien los componentes del ecosistema digital mencionados al inicio todavía están presentes, la contribución de cada uno ha cambiado. Lo que es aún más importante, nuevos estadios emergen en esta organización industrial. Por ejemplo, los desarrolladores de aplicaciones digitales (una función inexistente en la cadena productiva anterior) crean nuevas propuestas de valor como ser la vinculación entre oferta y demanda para productos específicos (la compra de pasajes de avión, el trueque de habitaciones en casas de particulares, y el encuentro de oportunidades de trabajo). Estas funciones ya existían con anterioridad, pero su funcionamiento no tenía la eficiencia resultante de las plataformas digitales. Otras nuevas aplicaciones con un papel específico se enfocan en la provisión de servicios de comunicación (como ser *Whatsapp*). Al mismo tiempo, emergen plataformas clave cuya función es agregar ofertas desde la búsqueda (*Google*) a las redes sociales (*Facebook*, LinkedIn). En esta nueva cadena de valor digitalizada, la conectividad y el transporte representan los insumos que permiten a los diferentes componentes o estadios interrelacionarse entre sí y llegar al usuario final.

En el marco de esta nueva cadena de valor se desencadenaron simultáneamente dos dinámicas opuestas de comportamiento de firmas participantes: la fragmentación de la cadena y su integración. Uno de los factores que determinaron la fragmentación de la cadena productiva eran las economías de

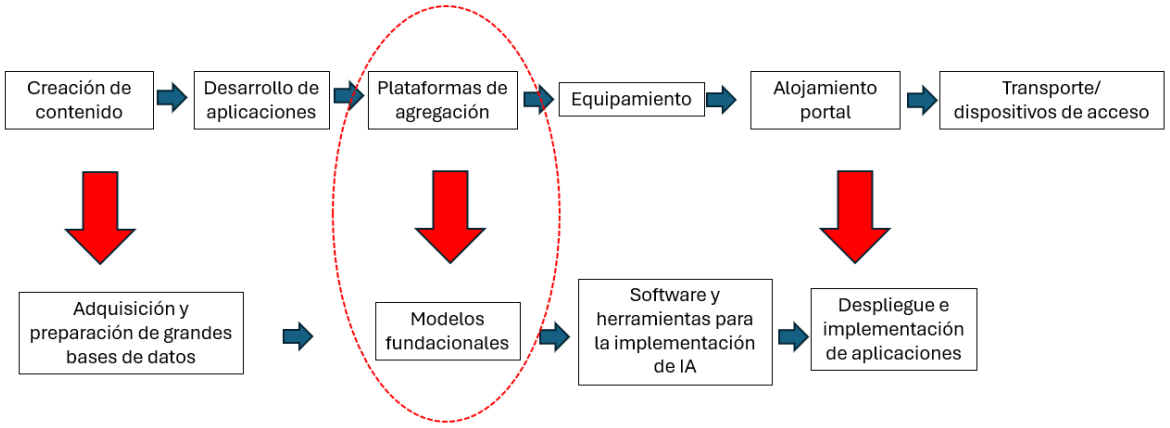
escala o de especialización. La tendencia a la fragmentación facilitada por la modularización y los "ensambladores" comenzó a ser afectada por la decisión de ciertas firmas que ocupaban posiciones específicas a adquirir empresas en estadios adyacentes de la cadena. Casi inmediatamente con la emergencia de la cadena de valor fragmentada, las firmas que ocupaban posiciones únicas comienzan a buscar oportunidades de diversificación en otros estadios, buscando integrarse verticalmente. Una motivación central para la entrada en otros estadios de la cadena de valor es la búsqueda de nuevas posiciones de rentabilidad superior, o para preservar posiciones de dominancia.

Bajo esta segunda configuración, las plataformas de agregación (Facebook, Google), OTT (Netflix, YouTube, WhatsApp) reconfiguran la captura de valor vía datos, atención a clientes y efectos de red. En este estadio, el marco regulatorio está influenciando por la competencia entre plataformas, la protección de datos personales, la ciberseguridad, y la regulación de mercados digitales.

3. Tercer período: irrupción de la inteligencia artificial y el poder de cómputo (2023 - ¿?)

El lanzamiento público de ChatGPT en octubre de 2022 indica el comienzo de la transformación de la cadena de valor de una estructura dominada por plataformas de agregación y OTT a una centrada alrededor de la inteligencia artificial y la computación en la nube. En esta transformación, si bien ciertos eslabones continúan existiendo los modelos de negocio cambian radicalmente (ver diagrama 6).

Diagrama 6
Transición de la cadena de valor del ecosistema digital a la de inteligencia artificial



Fuente: Elaboración propia.

En este nuevo contexto, la centralidad de las plataformas de agregación (Google, Meta) del ecosistema digital será apalancada por el desarrollo de los modelos fundacionales de IA. Estos se perpetuarán operando en el centro de la nueva cadena, mientras que la generación de contenido original devendrá la producción de datos para alimentarlos, y los portales de aplicaciones se transformarán en plataformas para la entrega de aplicaciones verticales.

Así, como en el período anterior cuando las plataformas de agregación y OTT emergieron como un nuevo eslabón de la cadena, transformando su centralidad y dinámicas de competencia, en la cadena de la IA los proveedores de servicios de esta tecnología, como ser aquellos que operan grandes modelos de lenguaje (*Large Language Models*), se convierten en actores clave, posicionándose en un nuevo eslabón. De la misma manera que la agregación de contenidos cambió el paradigma básico de la cadena en la segunda etapa, los proveedores de servicios de IA transforman la cadena en la medida de que actúan como plataformas intermediarias entre usuarios, empresas, desarrolladores y sistemas, ofrecen

herramientas y APIs[8] que se convierten en nuevos estándares para construir soluciones y modelos de negocio, y, sobre todo, procesan y generan datos a gran escala, lo que requiere elevados recursos de cómputo y conectividad.

Como se demostrará en el próximo capítulo, la capacidad de cómputo de los chips de alta potencia (GPU/TPU), los modelos fundacionales, la inferencia de modelos de IA a gran escala y las nuevas interfaces (copilotos/voz/multimodal) se transforman en activos críticos. En este marco, la integración vertical entre modelos se generaliza (chips-nube-modelos-distribución) y emergen nuevos "cuellos de botella" para la creación de valor (cómputo, energía, datos y talento). En esta tercera fase, las telecomunicaciones dejan de ser un eslabón independiente (como en las dos anteriores) y pasan a ser un habilitador sistémico que condiciona el desempeño simultáneo de varios eslabones (cómputo, nube, inferencia). En este contexto, emergen nuevas políticas públicas y preocupaciones regulatorias como la gobernanza algorítmica, los marcos de seguridad, la propiedad intelectual, el acceso competitivo a infraestructura y la sostenibilidad energética.

En consecuencia, los cambios estratégicos clave en la cadena de valor del ecosistema digital ocurridos en el tercer periodo son tres:

- **Convergencia entre telecomunicaciones y cómputo en la nube**: Los proveedores de servicios de telecomunicaciones ya no solo proveen conectividad, están entrando al negocio del *edge computing* y formalizando alianzas con proveedores de servicios de computación en la nube para ofrecer servicios de IA.
- **Concentración de poder**: Los actores que dominan el desarrollo de grandes modelos de lenguaje ejercen una influencia creciente en múltiples eslabones de la cadena (infraestructura, modelo, aplicación, API, estándar).
- **Dependencia tecnológica creciente**: Gobiernos, empresas y usuarios dependen cada vez más de los grandes modelos de lenguaje para acceder a servicios esenciales.

Estas transformaciones implican cambios en la organización industrial del ecosistema, los focos de inversión, el marco de gobernanza y los requerimientos regulatorios (ver cuadro 3).

Cuadro 3
Comparación entre segundo y tercer períodos de evolución de la cadena de valor

Dimensión	Segundo período: Emergencia de plataformas y OTT (2005-2022)	Tercer período: Irrupción de la inteligencia artificial y el poder de cómputo (2023 - ¿?)
Actores clave	Operadores de telecomunicaciones, fabricantes de hardware, desarrolladores de software, proveedores de servicios OTT	Hiperescaladores, desarrolladores de modelos fundacionales, integradores de plataformas verticales de IA
Cadena de valor	Creación de contenido >desarrollo de aplicaciones >plataformas de agregación >equipamiento >alojamiento>transporte, dispositivos	Equipamiento>infraestructura> adquisición/preparación de datos>modelos>software y herramientas>despliegue e implementación de aplicaciones
Dimensión	Segundo período: Emergencia de plataformas y OTT (2005-2022)	Tercer período: Irrupción de la inteligencia artificial y el poder de cómputo (2023 - ¿?)
Organización industrial	Segmentada: actores especializados por eslabón. Operadores de telecomunicaciones dominan la conectividad; OTT dominan la agregación de contenidos	Concentración creciente: pocos actores dominan múltiples eslabones (infraestructura, modelo, servicio). Mayor integración vertical
Focos de inversión	Expansión de redes (fibra, 4G/5G), licencias de espectro, hardware	Capacidad de cómputo (GPUs, centros de datos), entrenamiento de modelos fundacionales, APIs de IA

[8] Los denominados *Application Programming Interfaces* (APIs) son un conjunto de reglas que permiten a diferentes aplicaciones de software comunicarse entre sí, actuando como intermediarios para intercambiar datos y funcionalidad.

Dimensión	Segundo período: Emergencia de plataformas y OTT (2005-2022)	Tercer período: Irrupción de la inteligencia artificial y el poder de cómputo (2023 - ¿?)
Marco regulatorio y de gobernanza	Regulación centrada en servicio universal, competencia, protección de datos personales, neutralidad de red	Gobernanza algorítmica en construcción, protección de datos, desafíos de transparencia y soberanía tecnológica
Vacíos de gobernanza y regulación	Regulación de datos centrada en telcos y OTT. Escasa coordinación multilateral	EE. UU.: Vulnerabilidad a concentración de mercado, falta de normas para auditoría de modelos IA, opacidad de modelos fundacionales, limitada participación de países en desarrollo en la gobernanza global

Fuente: Elaboración propia con contribución de CEPAL.

La evolución de la cadena de valor en las etapas arriba mencionadas puede ser verificada con base en el estudio del valor de mercado de las 88 empresas globales de mayor capitalización bursatil en el ecosistema digital. Al categorizar estas empresas en telecomunicaciones, software y servicios informáticos, plataformas de servicios de internet, hardware y plataformas de comercio electrónico, puede observarse que entre el 2010 y el 2022 (o sea la segunda etapa de desarrollo del ecosistema digital) se produce una reconfiguración en la cadena de valor, indicando el crecimiento del valor relativo de software (como SAP, Salesforce, Adobe) y servicios informáticos (IBM, Accenture, Tata) y las plataformas de comercio electrónico (Amazon, Alibaba, Doordash). A partir del 2022, los sectores de semiconductores (NVIDIA, TSMC, y AMD) e infraestructura (Dell, HPE, y Quanta Computer) pasan de ser un sector secundario a un componente principal. A lo largo de estos dos periodos, las telecomunicaciones mantienen un valor de mercado estable, con lo cual pasan de ser un componente central del ecosistema al proveedor de un insumo intermedio (ver cuadro 4).

Cuadro 4
Valor de mercado sectorial de empresas en el ecosistema digital, a marzo de cada año
(En miles de millones de dólares y en porcentajes)

	2010	2020	2025
Valor de mercado (mil millones de dólares)			
Telecomunicaciones	2 244 286	2 445 146	2 638 121
Plataformas de comercio electrónico	207 286	2 156 088	3 036 655
Plataformas de servicios de internet y software	1 080 208	6 340 196	9 049 672
Hardware	1 420 340	2 943 697	11 543 837
Software y servicios informáticos	1 220 172	3 662 345	6 756 029
Total	**6 172 292**	**17 547 473**	**33 024 315**
Porcentaje			
Telecomunicaciones	**36,4**	13,9	8,0
Plataformas de comercio-e	3,4	12,3	9,2
Plataformas de servicios de internet y software	17,5	**36,1**	27,4
Hardware	23,0	16,8	**35,0**
Software y servicios informáticos	19,8	20,9	20,5

Fuente: Observatorio de Desarrollo Digital de la CEPAL (ODD) sobre la base de Bloomberg; análisis de los autores.

Los valores del cuadro 4 son ilustrativos de los cambios estructurales que se producen en el ecosistema digital en los últimos quince años:

- El valor de mercado en términos absolutos entre el 2010 y el 2025 ha crecido más de cinco veces, potenciado principalmente por el comercio electrónico y las plataformas de servicios de internet (segunda etapa de transformación del ecosistema digital).

- Desde el 2020, impulsado por el valor de empresas en el diseño y manufactura de semiconductores, la participación de hardware en el valor de mercado del ecosistema ha crecido de 16.8% a 35%, el cambio estructural más significativo directamente asociado a la irrupción de la IA.
- Determinada por el estancamiento en el valor de las empresas de telecomunicaciones, su participación ha disminuido del 36.4% en el 2010 a 8% en el 2025.

La literatura que estudia la estructura de industrias digitales no es consistente y/o rigurosa conceptualmente en el análisis de la cadena de valor. Por ejemplo, el concepto de "*AI stack*" es frecuentemente usado para asimilarlo al de cadena productiva. En términos generales, el "*AI stack*" es un ensamblaje de herramientas, tecnologías, modelos e infraestructura necesarios para construir, entrenar, desplegar y gestionar aplicaciones. En este sentido, el "*AI stack*" presenta una visión estructurada y modular del ecosistema requerido para desarrollar aplicaciones de IA e implementarlas en el mercado (ver ejemplo en el diagrama 7).

Diagrama 7
Estructura del "AI stack" (ejemplo)

Aplicaciones	Aplicación para el desarrollo de IA generativa	• Aplicaciones para generar respuestas en lenguaje natural a una amplia variedad de preguntas y "prompts", como ser **chatbots** y **traductores**. Es fundamental minimizar alucinaciones y basar el modelo en la realidad • **IA** (creatividad: datos históricos + **Búsqueda** (exactitud: datos en tiempo real) + **barandas** (relevancia)
API	PaLM API	APIs para **desarrolladores profesionales** (punto de acceso de LLMs de Google para construir una variedad de aplicaciones) y **aficionados** (herramientas de prototipado)
Grandes modelos de lenguaje	Familia de modelos PaLM	• Modelos de lenguaje pre-entrenados con **propósito general,** puede ser refinados para tareas especificas • Diseñados para desempeñar tareas basadas en lenguajes genéricos, como traducción, resúmenes y generación de dialogo
Modelos fundacionales de ML	ULM	• Modelos pre-entrenados de **propósito general** sobre lenguajes versátiles • **Pueden ser refinados** para una variedad de tareas de lenguaje natural, como clasificación de texto, o análisis de sentimientos
Arquitectura	Transformador	Arquitectura de red neuronal para tareas de procesamiento de lenguaje natural lo que planteas dos capacidades: • Toma de decisiones • Control de atención que pueden ser resumidas con un concepto único de **atención personal**
Infraestructura	Clústeres de TPU	Tensor Processing Units (TPU) diseñadas específicamente para cargas de trabajo de aprendizaje profundo y optimizadas para desempeñar operaciones matriciales, usadas comúnmente en redes neuronales

Fuente: Google.

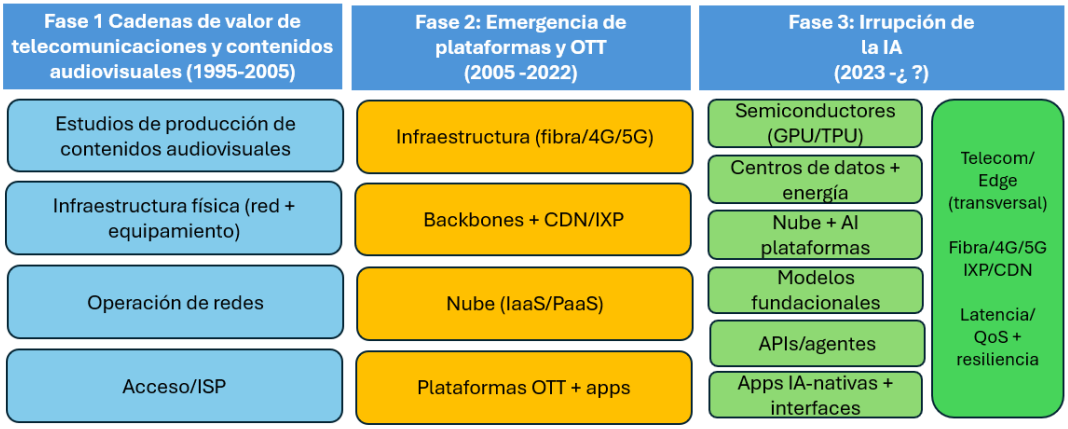
Si bien algunos componentes del "*AI stack*" pueden ser asimilados a eslabones de la cadena de valor (por ejemplo, infraestructura o modelos fundacionales), la estructura de este no permite entender los mecanismos económicos y de organización industrial que guían el funcionamiento de la industria.

De todas maneras, el *stack* en el caso de la IA permite representar la arquitectura funcional del ecosistema (conectividad/edge, cómputo, nube, modelos, herramientas, APIs/agentes, aplicaciones e interfaces) y, con ello, identificar puntos de control y mecanismos de captura de valor asociados a economías de escala y alcance, efectos de red, costos de cambio e interoperabilidad. Por lo tanto, el concepto es adecuado para discutir temas como organización industrial, competencia, riesgos de *lock-in* y desafíos de gobernanza en mercados digitales. Por otro lado, la cadena de valor, en cambio,

reconstruye el proceso económico-productivo que transforma insumos estratégicos —energía, minerales, datos y capital humano— en capacidades (infraestructura, cómputo, modelos, herramientas) y en servicios (implementación y uso), habilitando un análisis directo de inversión, encadenamientos productivos, dependencia tecnológica y oportunidades de política industrial y desarrollo productivo.

En este sentido, ambos enfoques se complementan y fortalecen la recomendación de políticas: el *stack* sustenta diagnósticos sobre poder de mercado y reglas del juego por capa, mientras la cadena permite diseñar intervenciones sobre capacidades, insumos críticos y trayectorias de inversión necesarias para cerrar brechas y capturar valor por eslabón. Es importante mencionar que, de la misma manera que en el caso de las tres etapas de desarrollo de la cadena productiva del ecosistema digital descritas arriba, el *technology stack* también ha incurrido transformaciones concurrentes (ver diagrama 8).

Diagrama 8
Evolución del *Technology Stack*

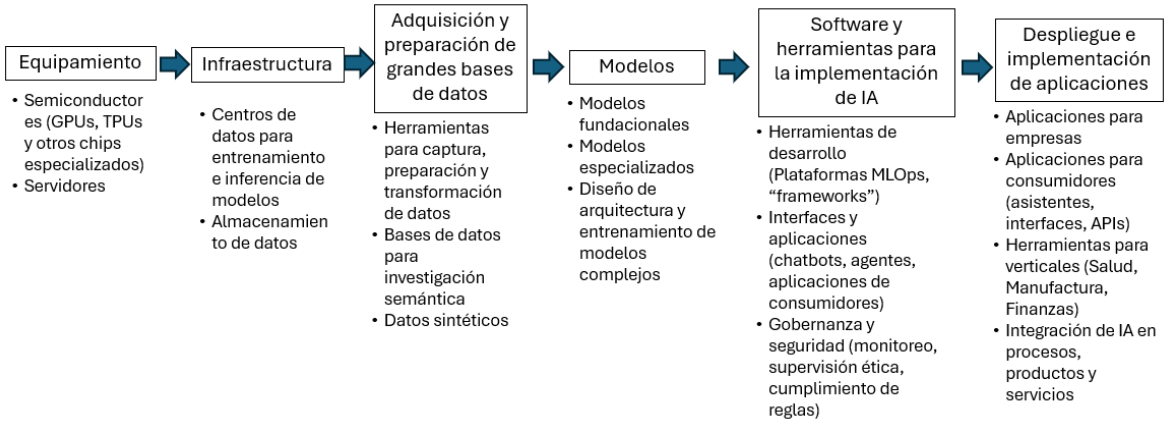


Fuente: Elaboración propia con contribución de CEPAL

C. Estructura de la cadena de valor de la inteligencia artificial

Reconociendo que no existe una visión única de la cadena de valor de IA, sugerimos el siguiente esquema para el tercer período, compuesto por seis eslabones (ver diagrama 9).

Diagrama 9
Estructura de la cadena de valor de IA



Fuente: Elaboración propia.

1. Eslabones de la cadena de valor de inteligencia artificial

Retomamos el esquema de la cadena de valor de IA presentado en el capítulo anterior, el cual está compuesto por seis eslabones.

La cadena comienza identificando al equipamiento como el primer eslabón necesario para desarrollar los modelos de IA. El componente clave dentro de este eslabón es el de los semiconductores, lo que incluye a GPUs (*Graphics Processing Units*, en inglés), TPUs (*Tensor Procesing Units*, en inglés) y otros chips especializados. Algunos esquemas de la cadena de valor de IA incluyen como eslabón independiente el uso de minerales para la manufactura de semiconductores (como ser silicio, cobre, oro, y cobalto). Desde nuestra perspectiva, estos deben ser considerados como un insumo intermedio para la fabricación de semiconductores y, por lo tanto, no deben representar un eslabón diferenciado (ver abajo en el apartado 4.2 una perspectiva de la cadena que incluye insumos intermedios). El segundo componente del primer eslabón es el ensamblaje de procesadores en servidores. Los ensambladores toman el producto de las fundiciones de semiconductores y los ensamblan en servidores funcionales usados para desarrollar y correr los modelos de IA.

La infraestructura de cómputo para la IA, considerada como el segundo eslabón, permite desarrollar y operar las aplicaciones de IA. En algunas versiones de la cadena se identifica a las fuentes energéticas como un eslabón adicional debido a la importancia que estas tienen en las necesidades operacionales de los centros de datos. Sin desconocer este factor, se considera que la energía representa nuevamente un insumo intermedio para sus operaciones y no un eslabón independiente. Por lo tanto, el análisis de la energía como factor determinante del perfil económico y operacional de los centros de datos debe ser tratado dentro del eslabón (aunque de manera independiente dada su importancia). El componente principal del eslabón de infraestructura incluye tres tipos de centros de datos. Por un lado, los hiper-escaladores representan los proveedores de servicios de computación en la nube, apalancando la ventaja de las redes globales de sitios desplegados ya con anterioridad al avance de la IA para responder a las necesidades de procesamiento de datos provenientes de organizaciones públicas y privadas y que en la actualidad representan la infraestructura principal para el desarrollo y aprovisionamiento de IA. Por otro lado, los "*neo-cloud*" son proveedores de computación especializados enfocados en el despliegue de GPUs, ofreciendo servicios más enfocados, y potencialmente más económicos, que los de hiperescaladores. Los neo-clouds ofrecen contratos más flexibles, aprovisionamiento más rápido y configuraciones de infraestructura más ajustadas a las necesidades de organizaciones usuarias (Mazza et al., 2025). El tercer tipo de infraestructura es el denominado "AI factories". Estos son centros de datos pero que han sido diseñados especialmente para correr aplicaciones de inteligencia artificial.

El tercer eslabón corresponde a la adquisición y preparación de datos que alimentan el entrenamiento, ajuste, evaluación y operación de sistemas de inteligencia artificial. Esta etapa incluye la captura, limpieza, estructuración, etiquetado [9], transformación e integración de grandes volúmenes de datos provenientes de fuentes diversas, como texto, imágenes, audio, código, registros administrativos o bases documentales. Los datos preparados en este eslabón pueden organizarse en distintos tipos de repositorios, según su naturaleza y uso previsto. Los "*data lakes*" permiten almacenar grandes volúmenes de datos en bruto o semiestructurados para su procesamiento posterior. Las bases de datos relacionales organizan información estructurada en tablas y registros, facilitando consultas precisas y relaciones definidas entre variables. Las bases vectoriales, en cambio, almacenan representaciones numéricas (*embeddings)*, que capturan propiedades semánticas de textos, imágenes, audio u otros contenidos, permitiendo realizar búsquedas por similitud, en lugar de limitarse a coincidencias exactas de palabras o campos.

[9] En redes neuronales que representan la tecnologia de base de la IA, el "etiquetado" es el proceso de incorporación en cada unidad de datos brutos (imágenes, texto, audio) una "etiqueta" que permite entrenar modelos de IA en el reconocimiento de patrones y realizar predicciones acertadas. Este proceso permite al modelo aprender a partir de la relación entre datos de insumo y generar respuestas.

Este eslabón es especialmente relevante para arquitecturas de RAG (*Retrieval-Augmented Generation*), en las que un sistema no responde únicamente con base en lo aprendido durante el entrenamiento del modelo, sino también recuperando información pertinente desde una base documental externa en el momento de la consulta. En estos sistemas, los *embeddings* y las bases vectoriales cumplen un papel central, ya que permiten identificar documentos o fragmentos semánticamente cercanos a la pregunta formulada. De este modo, la adquisición y preparación de datos no solo provee insumos para entrenar modelos, sino también para mejorar su capacidad de actualización, contextualización y uso efectivo en entornos productivos, por ejemplo.

El cuarto eslabón ocupa una posición central en la cadena de valor de la IA, ya que comprende el desarrollo, entrenamiento y adaptación de modelos fundacionales, entendidos como modelos entrenados sobre grandes volúmenes de datos y reutilizables para múltiples tareas posteriores, incluyendo el desarrollo de aplicaciones de IA (Bommasani et al., 2021). Esta categoría incluye, entre otros, los modelos de lenguaje de gran escala (Large Language Models), así como modelos de visión, audio y multimodales. Ejemplos de estas familias incluyen los Vision Transformers en visión computacional (Dosovitskiy et al., 2021), modelos de reconocimiento y procesamiento de audio como Whisper (Radford et al., 2022), y modelos multimodales como CLIP y Gato (Reed et al., 2022). El desarrollo de este eslabón comprende tanto el entrenamiento inicial del modelo (pretraining) como su adaptación posterior a tareas específicas (fine-tuning), procesos que dependen del acceso a datos, capacidad de cómputo y técnicas de optimización adecuadas. El resultado no es aún la aplicación final, sino un modelo base o especializado que posteriormente puede integrarse en herramientas, agentes, sistemas empresariales o servicios orientados a individuos y organizaciones.

El quinto eslabón comprende el software y las herramientas que permiten adaptar, evaluar, desplegar, monitorear y gobernar sistemas de IA en entornos organizacionales. Su función es actuar como capa intermedia entre los modelos del eslabón anterior y las aplicaciones finales del eslabón siguiente, permitiendo llevar modelos a producción de manera segura, escalable e integrada con procesos de negocio. Este eslabón incluye herramientas de MLOps y LLMOps, orientadas a la gestión del ciclo de vida de los modelos, tales como evaluación, pruebas, versionado, orquestación, despliegue, monitoreo de desempeño, observabilidad y mecanismos de gobernanza. La relevancia de este eslabón radica en que la adopción de IA en organizaciones no depende únicamente de disponer de un modelo, sino de contar con capacidades para integrarlo a sistemas, datos y flujos de trabajo específicos. De acuerdo con una encuesta de OMDIA, 44% de empresas reportan que las barreras principales para la incorporación de IA en los procesos de la organización residen en su complejidad y dificultad en la integración.[10] En este contexto, las herramientas incluidas en este eslabón facilitan tareas como el ajuste fino del modelo, su integración con bases documentales o sistemas internos, las pruebas en producción y el monitoreo continuo de su funcionamiento, reduciendo así la complejidad de implementación y operación.

Finalmente, el sexto eslabón comprende el desarrollo, despliegue e implementación de aplicaciones de IA orientadas tanto a consumidores finales como a empresas y organizaciones. En esta etapa, los modelos y herramientas desarrollados en los eslabones anteriores se traducen en productos, servicios y soluciones concretas, cercanas al usuario final o integradas en procesos de negocio. Es importante mencionar que esta función se diferencia por el mercado que intenta servir: aplicaciones para i) consumidores individuales (por ejemplo, chatbots que integran búsqueda de información para consumidores) o ii) empresas (soluciones diseñadas para optimizar funciones, automatizar procesos o apoyar la toma de decisiones). En el primer caso, la función es asumida por empresas que ya han desarrollado modelos funcionales y que buscan capitalizar su inversión a partir de ingresos en el mercado masivo (ver, por ejemplo, los avances de OpenAI en el mercado de servicios individuales,

[10] OMDIA (2022). *AI maturity study*.

apalancando su posición en ChatGPT). En el segmento empresarial, este eslabón suele estar liderado por integradores de sistemas, consultoras tecnológicas, proveedores sectoriales y, en algunos casos, empresas de telecomunicaciones o servicios digitales, cuya proposición de valor consiste en adaptar la IA a las necesidades específicas de la organización usuaria. Ello supone intervenir no solo en la dimensión tecnológica, sino también en tres ámbitos complementarios:

i) **Datos:** romper los silos departamentales en la organización usuaria para integrar información dispersa y generar una visión unificada del negocio.

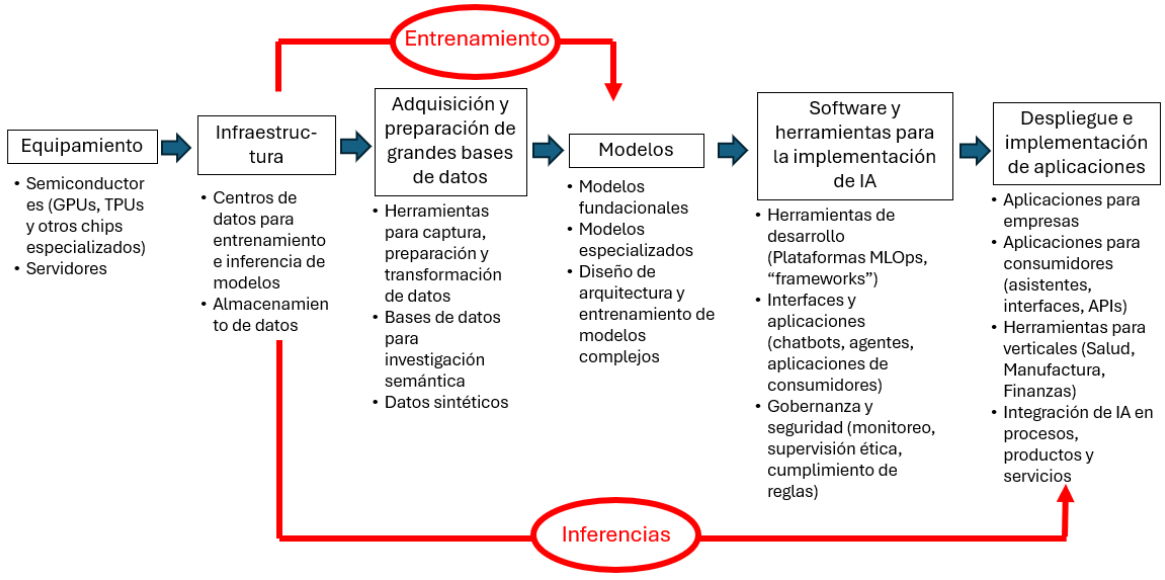
ii) **Talento:** la contratación de roles híbridos que entiendan tanto la estrategia de negocio como las capacidades técnicas de la IA.

iii) **Gestión del Cambio:** rediseñar procesos, funciones y formas de trabajo. No basta con dar una herramienta de IA a un empleado; hay que cambiar cómo trabaja ese empleado. Por ejemplo, en servicio al cliente, el rol del agente humano pasa de responder preguntas simples (ahora automatizadas) a gestionar casos complejos y excepciones emocionales, lo que requiere un perfil de empleado diferente y una capacitación distinta.

De este modo, el sexto eslabón representa la capa en que la IA deja de ser una capacidad tecnológica general y pasa a convertirse en una solución económica y organizacionalmente operativa.

Habiendo presentado las seis etapas de la cadena de valor de IA, es importante resaltar el punto crítico que representan los centros de datos dentro del eslabón de infraestructura. Los centros de datos apoyan dos funciones centrales de la cadena de valor: el entrenamiento y ajuste de modelos (*pre-training* y *fine-tuning*) y, por otro, la inferencia, es decir, la operación de aplicaciones y servicios de IA dirigidos a consumidores individuales y empresas.

Diagrama 10
Funciones del eslabón de infraestructura



Fuente: Elaboración propia.

El entrenamiento de modelos requiere elevados recursos computacionales para llevar adelante su desarrollo (es decir, los modelos fundacionales representan un producto con economías de escala significativas, y por ende altas barreras de entrada). Existen numerosas estimaciones del costo de

entrenamiento de modelos fundacionales. Por ejemplo, *The Economist* (2023) estimó el costo de entrenamiento de ChatGPT 4 en US$ 100 millones (The Economist, 2023). Esta estimación, citada por diferentes autores, es muy conservadora a la luz de la desagregación de componentes requeridos. Por ejemplo, el entrenamiento de ChatGPT 3 requirió un estimado de 6000 GPUs del modelo H100 de NVIDIA, con un costo promedio unitario de US$ 40,000, lo que resultaría en solamente el costo de hardware en US$ 240 millones. Adicionalmente a los GPUs se requiere equipamiento adicional como CPUs, memoria, almacenamiento, sistemas de refrigerado, lo que alcanzaría a su vez US$ 120 millones. A ello se suma el costo del edificio albergando los centros de datos, que alcanzaría US$ 10 millones. En conclusión, el costo total para el entrenamiento del chatbot que precedió al popular ChatGPT4 se estimaría en US$ 370 millones [11].

Los recursos computacionales requeridos para el *fine-tunning* de modelos fundacionales son comparativamente menores (aunque continúan siendo elevados debido al requerimiento de datos específicos y su etiquetado). Por ejemplo, el costo requerido para entregar ChatGPT a usuarios de la plataforma Bing de Microsoft son estimados en US$ 4 mil millones [12].

Las aplicaciones de IA dirigidas a consumidores individuales no se limitan a los chatbots. También incluyen asistentes de búsqueda inteligente, sistemas de generación de texto, imagen, audio y video, asistentes de programación, herramientas de productividad y otros servicios generativos integrados en plataformas digitales. En estos casos, la provisión del servicio requiere recursos de inferencia continuos, ya que cada interacción del usuario demanda capacidad de cómputo, almacenamiento y conectividad. Por ello, a diferencia del software tradicional, el costo marginal por uso no tiende a cero, aun cuando el costo unitario por interacción pueda ser relativamente bajo.

En el caso de las aplicaciones orientadas a empresas y organizaciones, los requerimientos de infraestructura suelen ser aún más exigentes. Ello no depende solo de la inferencia del modelo, sino también de la necesidad de integrarlo con sistemas internos, bases documentales, procesos organizacionales y entornos con mayores exigencias de seguridad, trazabilidad, cumplimiento y disponibilidad. En este segmento, las aplicaciones generativas incluyen asistentes documentales, sistemas de análisis y generación de reportes, copilotos para atención al cliente, herramientas de automatización de procesos y soluciones de apoyo a la toma de decisiones. Por ello, la provisión de aplicaciones de IA en entornos empresariales y públicos tiende a mantener costos elevados y requiere una infraestructura más compleja que la de muchas aplicaciones de consumo.

2. Insumos intermedios de los eslabones de la cadena de valor: aspectos críticos

Retomando la visión estructural de cadenas de valor, pero reflejando ahora el esquema de Williamson (1985) [13], quien incluye los insumos intermedios requeridos para el funcionamiento de cada eslabón, es posible diseñar una cadena de valor más completa (ver diagrama 11).

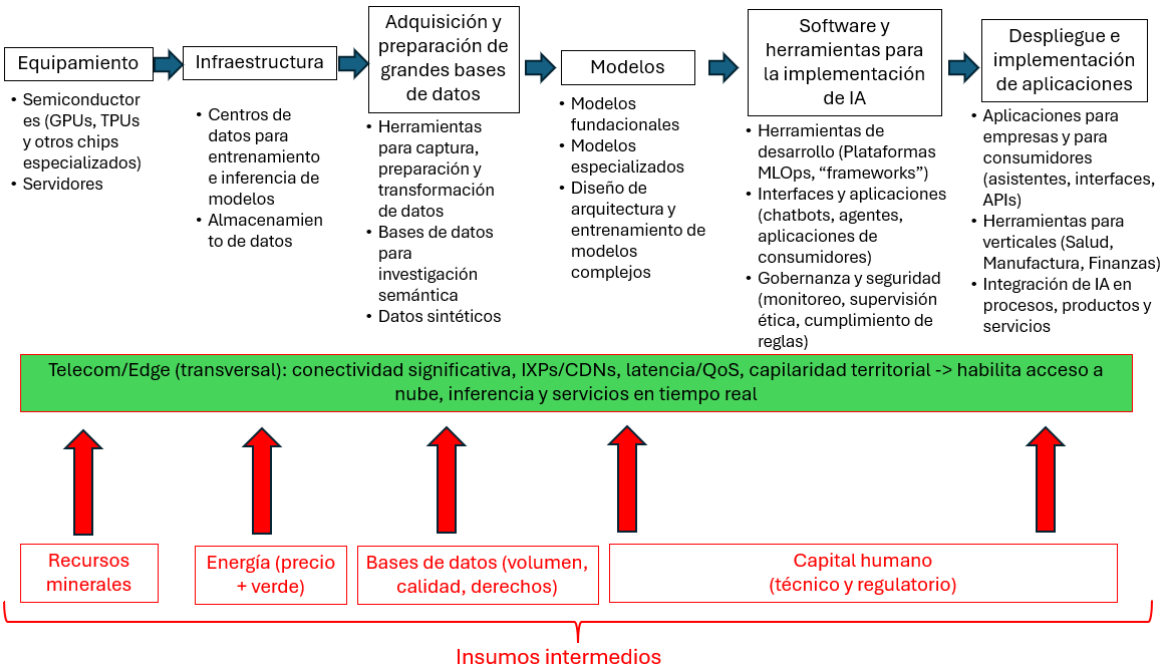
---

[11] Valores y fuentes citados en Ela Glowicka & Jan Malek: "Digital Empires Reinforced? Generative AI value chain". *Network Law Review*. Disponible en: https://www.networklawreview.org/glowicka-malek-generative-ai/.

[12] Vanian, J. and Leswing, K. (2023). "ChatGPT and generative AI are booming, but the costs can be extraordinary" CNBC (March 13), Disponible en: https://www.cnbc.com/2023/03/13/chatgpt-and-generative-ai-are-booming-but-at-a-very-expensive-price.html.

[13] En su análisis de las "fronteras de eficiencia" de la firma, Williamson conceptualiza aquellas funciones e insumos que deben ser externalizados (tercerizados) cuando aun considerando los costos de transacción, resulta más económico adquirirlos de un agente externo que asumir su propiedad. Ver Williamson, O. (1985). *The economic institutions of capitalism*. New York: The Free Prees, p. 98.

Diagrama 11
Estructura de la cadena de valor de IA (incluyendo insumos intermedios)



Fuente: Elaboración propia.

La inclusión de los insumos intermedios permite identificar los cuellos de botella a los que está sometido cada eslabón para acceder a recursos necesarios para llevar adelante su función, y que a su vez condicionan costos y capacidades.

La demanda exponencial de semiconductores para equipar los centros de datos para el entrenamiento e inferencia de modelos de IA genera una presión para la extracción de recursos minerales (en particular, cobre, cobalto, litio, tierras raras y galio). Esto no solo está determinado por la construcción de nuevos centros de datos sino también porque el ciclo de vida de un GPU es extremadamente corto [14], lo que implica rápidos procesos de reemplazo [15]. El "estrangulamiento" en el flujo de minerales para la fabricación de semiconductores está determinado por la concentración de puntos extractivos en pocos estados soberanos (Estado Plurinacional de Bolivia, República Democrática del Congo, Argentina, Australia, Chile, entre otros) y el control geopolítico de la extracción, procesamiento y distribución (enfrentando a China con Estados Unidos).

El insumo energético de los centros de datos, en términos del aumento sin precedentes de la demanda de energía eléctrica de IA, ya ha sido adecuadamente documentado. Los centros de datos de IA consumen una densidad de energía por metro cuadrado muy superior a los centros de datos tradicionales. Se estima que el consumo global de electricidad de los centros de datos podría duplicarse para 2026. La Agencia Internacional de Energía estimó que en el 2024 los centros de datos usados para IA han consumido 15% de la demanda mundial de electricidad, estimándose que, en el 2025, este valor habría alcanzado 20% (IEA, 2025). Así, en ese año la huella de carbono generada por los sistemas de IA habría oscilado entre 32.6 y 79.7 millones de toneladas de CO_2 (de Vries-Gao, 2025). En mercados clave

[14] Un GPU debe ser reemplazado cada 2-3 años; los GPUs retirados de circulación son incinerados dado que la extracción de minerales para un GPU es un proceso muy costoso.

[15] Ver Carayannis, T., Kane, N., Png, M., & Nelson, A. (2025). Summary Report: Workshop on the Geopolitics of Critical Minerals and the AI Supply Chain. Science, Technology, and Social Values Lab, Institute for Advanced Study.

como Estados Unidos, las empresas de servicios públicos deben responder a solicitudes de conexión de carga que superan los 80-90 Gigavatios (GW), una cifra que la infraestructura de generación y transmisión actual no puede absorber sin inversiones de capital masivas, que requieren largos tiempos de despliegue [16]. Esto ha determinado que la disponibilidad de energía haya reemplazado a la latencia o la ubicación geográfica como el criterio principal para la selección de sitios de centros de datos. Esta situación ya ha generado modificaciones en el acceso a fuentes energéticas por parte de grandes operadores de centros de datos: aumento de fuentes basadas en el uso de gas natural para reducir emisiones de carbono, retorno al uso de energía nuclear y el uso de pequeños reactores modulares, e inversión en tecnologías geotermales.

El acceso a bases de datos necesarias para entrenar a los modelos fundacionales también enfrenta límites, lo que genera el tercer punto de presión en el uso de insumos intermedios por parte de la cadena de valor. El primer desafío relacionado con el entrenamiento de modelos fundacionales es el agotamiento de datos de alta calidad en internet. Adicionalmente, sectores regulados como salud y finanzas enfrentan barreras legales para usar datos reales en modelos ajustados a demandas de segmentos específicos. Debido al agotamiento de datos de alta calidad para el entrenamiento de modelos, los desarrolladores de modelos fundacionales deben recurrir a los datos sintéticos [17]. Estos son usados para generar volúmenes masivos de código o razonamiento lógico etiquetado, ayudando a los modelos a aprender patrones complejos donde los datos humanos son escasos. Su aplicación permite a investigadores médicos, por ejemplo, entrenar modelos sobre poblaciones de pacientes virtuales sin violar leyes de privacidad. [18] Adicionalmente, para modelos de conducción autónoma o robótica, es difícil obtener datos de situaciones inusuales (ej. un accidente específico), con lo cual los simuladores sintéticos pueden generar miles de variaciones de estos escenarios para robustecer las predicciones.

El insumo de capital humano difiere en el caso del desarrollo de modelos fundacionales del de la implementación de aplicaciones de IA en organizaciones. El entrenamiento de modelos fundacionales requiere un talento altamente especializado, con una formación académica típica de PhD en informática obtenido en las universidades más importantes del mundo. [19] La escasez en la disponibilidad de talento especializado ha generado una competencia entre los grandes desarrolladores de modelos como Meta, Google, OpenAI, Microsoft y Anthropic por el acceso a este talento, lo que ha resultado en una inflación en la compensación anual y en el otorgamiento de un volumen importante de acciones como mecanismo de retención de recursos humanos [20].

El último punto de presión en el acceso a insumos intermedios en la cadena de valor se refiere al talento requerido en la implementación de IA en organizaciones. Los desafíos que enfrentan las organizaciones para implementar y escalar sistemas de IA han resultado en la especificación de un perfil diferente del típico ingeniero de software: el denominado *forward deployed engineer,* originalmente desarrollado por Palantir para implementar IA en aplicaciones del gobierno estadounidense y extendido a otras empresas, entre ellas OpenAI. Desde el punto de vista del perfil de talento, este individuo presenta capacidades en software, soluciones de IA e ingeniería de plataformas, lo que le permite interactuar de manera alternada entre la comunidad de usuarios e ingenieros de software. Este recurso tiene

---

[16] Nathan, A.; Grimberg, J., Rhodes, A. (2024). GenAI: too much spend, too little benefit? Goldman Sachs Research Top of Mind Issue 129 (June 25). Disponible en: https://www.goldmansachs.com/images/migrated/insights/pages/gs-research/gen-ai--too-much-spend,-too-little-benefit-/TOM_AI%202.0_ForRedaction.pdf

[17] Los datos sintéticos son datos generados artificialmente que imitan las propiedades estadísticas de los datos reales sin contener información de identificación personal (PII).

[18] Shimabukuro, J. (2025). "AI data labelling and processing", Education Technology and Change. (August 27). Disponible en: https://etcjournal.com/2025/08/27/ai-data-labeling-and-processing-update-august-2025/

[19] Las diez mejores universidades del mundo en términos de formación en IA son Stanford (EE. UU.), MIT (EE. UU.), University of California Berkeley (EE. UU.), Carnegie Mellon (EE. UU.), Oxford (UK), Cambridge (UK), Tsinghua University (China), University of Toronto (Canadá), University of Michigan (EE. UU.), Georgia Tech (EE. UU.).

[20] Huang, K. (2025). "Meta hires four more OpenAI researchers". *The Information* (June 28) y Muppidi, S. (2026). "OpenAI reserves $50 billion for stock grants". The information (February 8).

aproximadamente 5 años de experiencia práctica después de los estudios universitarios en computación. Su salario anual en Estados Unidos aproxima US$ 185,000 anuales como compensación básica [21].

3. El papel de las telecomunicaciones

Inicialmente, como fuera mencionado a tiempo de presentar la cadena de valor del ecosistema digital, las telecomunicaciones actúan como insumo intermedio en el funcionamiento de ciertos eslabones de la cadena. Este posicionamiento se extiende a la cadena de valor de IA. Por ejemplo, las redes de banda ancha representan una tecnología complementaria a la computación en la nube para que esta ofrezca acceso a usuarios localizados remotamente. Katz y Jung (2023) han construido modelos de impacto económico de la computación en la nube en los que el desarrollo de las redes de banda ancha representa uno de los insumos esenciales en la creación de valor de la tecnología [22].

La complementariedad de las telecomunicaciones también se extiende al acceso de IA por parte de usuarios individuales. Por ejemplo, ciertos operadores de modelos fundacionales recurren al empaquetamiento de chatbots con operadores de telecomunicaciones. Así, Airtel (India) ofrece suscripciones gratuitas a la aplicación de IA Perplexity Pro a sus usuarios de alto valor, integrando la herramienta de búsqueda por IA como un valor añadido clave de su conectividad [23]. De manera similar, Reliance Jio en India ha establecido alianzas para ofrecer acceso a Google Gemini a su base de usuarios. [24] Este tipo de alianzas replican las ya existentes entre operadores de telecomunicaciones y proveedores de contenido y plataformas, donde el operador provee a la plataforma de IA acceso a una base amplia de usuarios y se beneficia por la diferenciación y la posible reducción de *churn*.

La tercera dimensión del papel de telecomunicaciones en la cadena de valor de IA es esencialmente la misma perseguida por empresas usuarias donde la IA es utilizada para transformar las operaciones de telecomunicaciones en áreas como *self-healing networks* que predicen y corrigen fallos antes de que estas afecten al servicio, la optimización energética de torres 5G (reduciendo una porción de OPEX), y chatbots de atención al cliente de nueva generación con reducción consiguiente de costos operativos.

Por ejemplo, Vodafone ha optado por un modelo de asociación, pero con integración operativa. A través de un acuerdo de diez años con Google y el despliegue de "AI Factories" en sus mercados clave, Vodafone integra capacidades de IA generativa directamente en sus sistemas de gestión de red (*Network Lifecycle*) y atención al cliente (TOBi). [25] En Italia, su alianza con Fastweb y el consorcio de supercomputación Cineca demuestra cómo un operador de telecomunicaciones puede verticalizarse

---

[21] Orosz, G. (2025). "What are Forward Deployed Engineers, and why are they so in demand" (August 12).

[22] En el estudio "The contribution of cloud to economic growth in the Middle East and North Africa" Katz y Jung modelan el impacto económico de la nube mediante un sistema de ecuaciones que endogeniza eficazmente la infraestructura de banda ancha y la penetración de la nube, ya que implica la oferta y la demanda de estas infraestructuras.
La ecuación de la demanda de banda larga establece que la penetración de banda ancha es función del ingreso medio por país, el precio de un servicio estándar para la conexión a la red, el grado de dotación de capital humano y el porcentaje de la población que vive en zonas densamente pobladas. La ecuación de oferta vincula la inversión agregada en banda ancha con los niveles de precios de banda ancha y la intensidad de la competencia en el mercado de telecomunicaciones. Estas variables afectan la dinámica del lado de la oferta en el mercado de banda ancha. La ecuación de infraestructura modela el cambio anual en la banda ancha en función de la inversión en banda ancha. A partir de ahora, imitamos el enfoque seguido en el caso de la banda ancha, para endogeneizar también la penetración de nubes. Esto significa que estamos introduciendo ecuaciones específicas vinculadas a la demanda de la nube, la oferta de la nube y la producción de infraestructura en la nube. En ese sentido, se espera que la demanda de la nube dependa del ingreso medio, de los precios de la nube, del grado de capital humano y del grado de urbanización. En cuanto a la ecuación de suministro de nube, vincula la inversión en la nube en función de los precios de la nube y la intensidad competitiva en el sector local de la nube. Finalmente, la variación en la penetración de la nube se modela para depender de la inversión en la nube.

[23] Digwatch (2025). *Perplexity AI overtakes ChatGPT on India's App Store*. (July 18). Disponible en: https://dig.watch/updates/perplexity-ai-overtakes-chatgpt-on-indias-app-store

[24] Kadia, H. *(2025). "AI in India: Why Airtel-Perplexity and Jio-Gemini offer free premium access".* Tecknexus (December 3). Disponible en: https://tecknexus.com/ai-in-india-why-airtel-perplexity-and-jio-gemini-offer-free-premium-access/

[25] Said, S. and Guido, M. (2024). *How Vodafone is using gen AI to enhance network life cycle.* Google Clkoud. Disponible en: https://cloud.google.com/blog/topics/telecommunications/vodafone-gen-ai-enhances-network-lifecycle

hacia el sector B2B de alto rendimiento, ofreciendo sus superordenadores para investigación y desarrollo industrial [26].

La cuarta área de intervención de los operadores de telecomunicaciones en la cadena de valor de IA se refiere a la transformación del modelo de negocio de empresas de telecomunicaciones para competir directamente con los proveedores de servicios de nube. Esta diversificación representa la respuesta de los operadores de telecomunicaciones a los cambios en la composición del mercado digital de empresas. Este último incluye la conectividad, la provisión de servicios en la nube, el mercado de software para el desarrollo de aplicaciones, y la IA. En el marco de la recomposición del mercado de empresas, los principales negocios que generan valor económico no forman parte del entorno de la conectividad, aunque dependen directamente de esta. Es por ello por lo que los operadores de telecomunicaciones buscan posicionarse como plataforma para la entrega de servicios de IA en tres áreas [27]:

i) **Edge Computing:** Desplegar capacidad de cómputo en el borde de la red, cerca de las torres y centrales locales para apoyar aplicaciones de IA que requieren latencia ultrabaja, como vehículos autónomos, robótica industrial o realidad aumentada [28].

ii) **GPU-as-a-Service (GPUaaS):** Algunos operadores están comenzando a ofrecer capacidad de procesamiento basada en GPU directamente a empresas, aprovechando su infraestructura distribuida, compitiendo o complementando a las nubes públicas.

iii) **Aplicaciones verticales:** reflejando la intención de consolidar su posición en ciertos segmentos del mercado de empresas, operadores de telecomunicaciones ofrecen aplicaciones de IA para servir mercados verticales como la salud y la educación.

Por ejemplo, SK Telecom (Corea del Sur) ha articulado una estrategia denominada "Pirámide de IA". En la base de esta, invierte en infraestructura propia, incluyendo centros de datos y diseño de semiconductores de IA (*Sapeon*). En la capa media (AIX), utiliza la IA para transformar sus propias operaciones y reducir costos. En la capa superior, lanza servicios directos al consumidor como su asistente personal "A. dot". El objetivo financiero de esta estrategia es triplicar las inversiones relacionadas con IA para 2028 y convertirse en una compañía global de IA, no solo un operador de telecomunicaciones [29].

Otro ejemplo es el de Deutsche Telekom quien ha adoptado un rol de "agregador de confianza". En lugar de intentar construir sus propios grandes modelos de lenguaje desde cero, se asocia estratégicamente con líderes como OpenAI y Perplexity. Con base en ello, ha lanzado un "Teléfono IA" que integra estos servicios en el sistema operativo, eliminando la fricción usuaria generada por las apps. Su modelo de negocio es el empaquetamiento para ofrecer suscripciones premium de IA (como Perplexity Pro) incluidas en las tarifas móviles, actuando como el canal de distribución masivo para el mercado europeo y aumentando el ARPU (Ingreso Promedio por Usuario) [30].

---

[26] Fastweb and Vodafone (2025). *Fastweb+Vodafone and Cineca: an alliance to accelerate the adoption of Generative AI in businesses*. Disponible en: https://www.fastwebvodafone.it/var/storage_feeds/cms20/media/fwb_wbce_cms20_fv_comun/cf1/cf1cab42c5a5841f5a7513e4262c3622/comunicato_en/2025-02-18-cs-eng-cineca.pdf

[27] Shrivastava, A., Gaffey, B., Grundin, G. Cubela, S. and Lajous, T. (2025). *AI infrastructure: a new growth avenue for telco operators*. Mckinsey and Co. Disponible en: https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/ai-infrastructure-a-new-growth-avenue-for-telco-operators

[28] Douglas, S. (2025). *Advancing AI opportunities at the Telecom edge*. Spirent, August 15. Disponible en: https://www.spirent.com/blogs/advancing-ai-opportunities-at-the-telecom-edge

[29] SK telecom (2023). SK Telecom announces AI Pyramid strategy to become a global AI company (September 29). Disponible en: https://www.sktelecom.com/en/press/press_detail.do?idx=1582

[30] Hema, K. (2025). Deutsche Telekom AI phone: Bundled AI for everyone. *Tecknexus* (August 15). Disponible en: https://tecknexus.com/ deutsche-telekom-ai-phone-bundled-ai-for-everyone/

El camino hacia la integración con IA para los operadores de telecomunicaciones no está exento de obstáculos. El más crítico es el de la brecha de capacidades de software. Las empresas de telecomunicaciones son expertas en gestionar redes físicas estáticas, no en orquestar cargas de trabajo de software dinámicas como las generadas por el uso de herramientas de IA. Asimismo, carecen de los ecosistemas de desarrolladores y las herramientas de MLOps maduras que poseen los hiperescaladores [31]. Es por ello que muchas optan por un modelo híbrido: proveen la infraestructura física y se asocian con NVIDIA o los hiperescaladores para la capa de software, convirtiéndose en canales de distribución verticalizados. Sin embargo, para capturar el margen real, deben superar este déficit de capacidades y aprender a vender "resultados de IA" y no solo "capacidad de red" [32]. Este es el mismo desafío que enfrentaban ya en la cadena de valor del ecosistema digital debido a la comoditización del transporte.

[31] Nadlall V. (2025). Telecom GPU-as-a-Service: Beyond the hype, RCR Wireless News. (October 30). Disponible en: https://www.rcrwireless.com/20251030/analyst-angle/telecom-gpu-as-a-service

[32] Shrivastava, A., Gaffey, B., Grundin, G. Cubela, S. and Lajous, T. (2025). *AI infrastructure: a new growth avenue for telco operators*. Mckinsey and Co. Disponible en: https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/ai-infrastructure-a-new-growth-avenue-for-telco-operators

II. La cadena de valor de la inteligencia artificial en Estados Unidos: estructura, dinámicas y dimensionamiento

Este capítulo retoma la estructura de la cadena de valor detallada en el capítulo anterior y la analiza con ejemplos ilustrativos de Estados Unidos. Para ello, en la primera sección se presentan los actores principales a partir de su posición original en la cadena y se analizan los movimientos de integración vertical de jugadores para ilustrar las dinámicas convergentes e intentos de control de puntos estratégicos en la creación de valor. Asimismo, considerando que la cadena de valor de la IA se encuentra en un estadio de transición y cambios continuos, el capítulo aborda los grandes interrogantes y puntos de presión que podrían ocasionar cambios en la cadena de valor en el futuro. Sobre la base del relevamiento de actores operando en cada eslabón, la siguiente sección estima el tamaño económico de cada uno, proyectándose también hacia donde fluyen los márgenes de ganancia. Para reiterar, todos los ejemplos y actores incluidos en este documento se refieren a Estados Unidos para después comparar este líder global en el desarrollo de la IA con China y Europa, analizados en el siguiente capítulo.

A. La cadena de valor de la IA en Estados Unidos

Esta sección se enfoca en el análisis de movimientos de integración vertical encarados por participantes de la cadena de valor de la IA, buscando capturar o defender posiciones dominantes, o minimizar los costos de transacción incurridos en la adquisición de insumos.

El estudio de la organización industrial en términos de fragmentación e integración de cadenas productivas originalmente planteado por George Stigler en el artículo citado en el capítulo I, postuló que, en sus orígenes, las industrias tienden a estar servidas por firmas integradas verticalmente a lo largo de cadenas productivas. Progresivamente, con el desarrollo las cadenas productivas tienden a fragmentarse en múltiples eslabones dado que firmas presentes en ciertas funciones específicas pueden operar de manera más eficiente que si continúan integradas verticalmente. Uno de los factores que

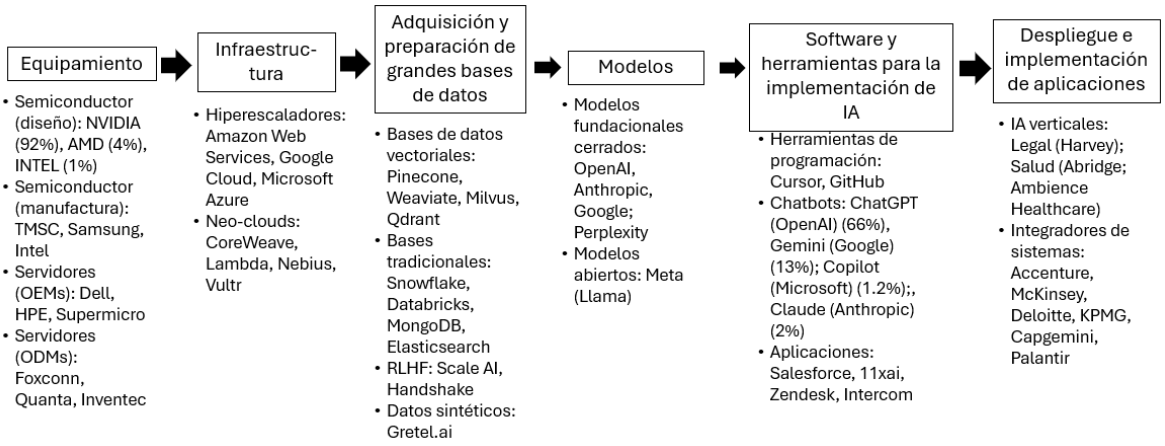
determina la fragmentación de la cadena productiva en firmas consolidadas horizontalmente son las economías de escala o de especialización. Al consolidar toda la producción de un eslabón en particular, sus costos unitarios tienden a ser más bajos y por lo tanto resulta más eficiente para una firma adquirir el insumo de un productor operando a escala que producirlos el mismo. Después de ese periodo caracterizado por la fragmentación vertical, cuando la industria alcanza un nivel de desarrollo maduro, las cadenas vuelven a integrarse para poder responder a una ralentización de la demanda. Otras motivaciones que guían la re-integración serían el control de la experiencia del usuario, o el acceso más eficiente a materias primas.

La siguiente sección va a examinar los movimientos de fragmentación e integración vertical de la cadena de valor de la IA a la luz de la teoría de Stigler, aunque incorporando los conceptos de Williamson de costos de transacción y dinámica competitiva, como factor determinante.

1. Posiciones originales en la cadena de valor

Si bien el desarrollo de la cadena de valor de la IA no ha ocurrido en términos tales como los articulados por Stigler de integración vertical inicial donde una sola empresa controlaba todos los eslabones de la cadena de valor, el hecho es que, en sus orígenes la cadena estaba constituida por firmas que ocupaban posiciones específicas, debido a que ninguna de ellas poseía los atributos técnicos y económicos para integrarse a lo largo de la cadena (ver diagrama 12).

Diagrama 12
Posicionamiento original de firmas en la cadena de valor de la IA



Fuente: Elaboración propia.

Es así como el nacimiento y desarrollo inicial de empresas del ecosistema de IA estaban vinculados a un eslabón especifico de la cadena. Si bien las maniobras de integración vertical ocurrieron casi simultáneamente con el desarrollo inicial de cada firma, sus modelos de negocio y características económicas originales estaban condicionados por el eslabón donde cada una nace. De hecho, la organización industrial y el nivel de intensidad competitiva varían sustancialmente en cada eslabón, requiriendo un análisis detallado para cada uno.

Equipamiento

El eslabón de equipamiento incluye tres segmentos: el diseño de semiconductores, su manufactura, y el ensamblaje de semiconductores en servidores a ser instalados en centros de datos.

Diseño de semiconductores

Basado en un mercado de US$ 142 mil millones, NVIDIA controla una posición dominante en el segmento de diseño de GPUs (*Graphics Processing Units*) para centros de datos para IA, estimada en 92% de cuota, mientras que el resto del mercado se divide entre AMD (4%), Huawei (2%), Intel (1%), Groq (1%), y Cerebras (1%) [33].

El aspecto más sobresaliente del modelo de negocio de NVIDIA es su capacidad para extraer márgenes brutos típicos de una empresa de *Software as a Service*, vendiendo bienes físicos. Los reportes financieros para el año fiscal 2025 indican ingresos récord de $130.5 mil millones de dólares, con márgenes brutos GAAP del 75.0% [34]. Esta cifra es anómala en la industria del hardware, donde los márgenes suelen oscilar entre el 30% y el 50% en los mejores casos. La razón de esta anomalía reside en que el precio de venta de sus sistemas (como los servidores DGX o las tarjetas H100/Blackwell) no se calcula en función del costo de los materiales sumado a un margen estándar, sino en función del valor aportado al Costo Total de Propiedad (TCO) del centro de datos del cliente. Al reducir drásticamente el tiempo de entrenamiento de modelos y el consumo energético relativo por operación, NVIDIA captura una porción significativa del valor económico generado por sus productos.

NVIDIA ha logrado trascender el modelo de negocio tradicional de un diseñador de GPUs sin fábrica propia, para convertirse en una plataforma de computación integrada. Es así como su dominio del mercado esta apalancado en una arquitectura de ecosistema cerrado que genera costos de cambio extremadamente altos para los usuarios. La posición dominante de NVIDIA está sustentada por la arquitectura de CUDA (*Compute Unified Device Architecture*), que provee la capa de software donde operan la mayoría de las bibliotecas de *Deep learning*, como PyTorch y TensorFlow. Esto genera un ciclo virtuoso de retroalimentación: i) los desarrolladores de modelos de IA construyen en NVIDIA porque las herramientas funcionan mejor allí; ii) las herramientas de desarrollo se optimizan para NVIDIA porque es la plataforma usada por los desarrolladores; y iii) las empresas compran GPUs de NVIDIA para garantizar la compatibilidad y el rendimiento de la infraestructura existente. Romper este ciclo requiere no solo un GPU de mejor desempeño, sino recrear un ecosistema de software que ya se beneficie de dos décadas de inversión realizada.

Adicionalmente, NVIDIA ha evolucionado de la oferta de GPUs a la venta de sistemas completos, ofreciendo DGX, una "supercomputadora en una caja" que integra procesadores, redes e interconexiones propietarias (NVLink).

Más allá de la venta transaccional de GPUs, NVIDIA está migrando agresivamente hacia modelos de ingresos recurrentes. La empresa ha comenzado a desglosar ingresos de "Software y Servicios". Servicios como NVIDIA AI Enterprise, que es esencialmente un sistema operativo para la IA empresarial, y DGX Cloud, permiten a la empresa monetizar continuamente su base instalada. Esta estrategia transforma el gasto de capital de sus clientes, que es cíclico y episódico, en gasto operativo, mitigando la volatilidad histórica inherente al mercado de semiconductores [35]. Este modelo simplifica el despliegue para los clientes empresariales, quienes prefieren comprar una solución llave en mano que integrar componentes dispares, reforzando aún más la dependencia del proveedor [36].

---

[33] IOT Analytics (2025). *The leading generative AI companies*. (March 4). Disponible en: https://iot-analytics.com/leading-generative-ai-companies/

[34] NVIDIA (2025). *NVIDIA announces financial results for fourth quarter and fiscal 2025* (February 26). Disponible en: https://nvidianews.nvidia.com/news/nvidia-announces-financial-results-for-fourth-quarter-and-fiscal-2025

[35] NVIDIA (2026). NVIDIA Corporation Annual Review. Form 10-K

[36] Future Bridge Americas (2025). The Untold Story of NVIDIA. Disponible en: https://future-bridge.us/the-rise-of-nvidia-how-ai-hardware-became-the-hottest-sector-in-us-semiconductors/

Un punto que podría amenazar la posición dominante de NVIDIA en el segmento de GPUs es si el desarrollo de modelos fundacionales puede incorporar aprendizaje continuo (*continual learning*), lo que permitiría a estos a requerir menos datos y capacidad de procesamiento, lo que resultaría en menos demanda de GPUs. Todos los participantes en el eslabón de modelos fundacionales están encarando investigación en modelos de aprendizaje continuo, aunque todavía no han generado resultados significativos [37].

Manufactura de semiconductores

De la misma manera que NVIDIA lidera el segmento de "fabless", la fabricación de GPUs está concentrada en TSMC, de Taiwan. TSMC opera bajo el modelo de *Pure-Play Foundry*, fabricando chips diseñados por terceros sin competir con ellos. A diferencia de NVIDIA, su negocio es intensamente dependiente de la inversión de capital para construir fábricas de semiconductores (denominado "fabs") de vanguardia. TSMC mantiene márgenes brutos extraordinariamente altos para una empresa manufacturera, oscilando entre el 53% y el 57% [38]. Esto se debe a su monopolio efectivo en los nodos de proceso más avanzados (3nm y 5nm), necesarios para la eficiencia energética que exige la IA.

La posición competitiva de TSMC está sustentada en su control de las tecnologías de empaquetado avanzado, denominada *Chip-on-Wafer-on-Substrate* (CoWoS). El CoWoS permite integrar múltiples componentes (como la GPU y la memoria de alto ancho de banda HBM) en un solo sustrato, maximizando la velocidad de transferencia de datos. De hecho, la capacidad de TSMC para escalar sus líneas de producción de CoWoS determina, en última instancia, el ritmo de adopción de la IA a nivel mundial. Esta dependencia confiere a TSMC un poder económico inmenso, reflejado en sus retornos consistentes superiores al 40% y un crecimiento de ingresos impulsado por la demanda de chips de IA [39].

Manufactura de servidores

Aguas abajo de los diseñadores de semiconductores y la manufactura, se ubican las empresas que toman las GPUs, las CPUs, la memoria y el almacenamiento, y los ensamblan en racks de servidores funcionales. Aquí, la organización industrial y las variables económicas cambian drásticamente de márgenes cuasi-monopolísticos de los segmentos precedentes a una competencia perfecta, desatando una "guerra de márgenes" acelerada. La manufactura de servidores destinados a ser instalados en los centros de datos está estructurada de acuerdo con dos modelos clásicos de manufactura de hardware: los *Original Equipment Manufacturers* (OEMs) y los *Original Design Manufacturers* (ODMs).

Los OEMs de servidores de AI incluyen a empresas como Dell Technologies y Hewlett Packard Enterprise (HPE), las que operan bajo un modelo de marca y servicio empresarial. Su propuesta de valor no es solo el equipamiento, sino también el soporte, la financiación y la integración con sistemas corporativos legados (ERP, bases de datos). Este modelo se caracteriza por una presión en márgenes importante. A pesar de reportar ingresos importantes ($96 mil millones para Dell en 2025), la construcción de servidores de IA es un negocio de margen reducido. Por ejemplo, los márgenes brutos para estos servidores específicos se sitúan en torno al 5-10% [40]. La razón de ello es estructural: la gran parte del costo del servidor está representado por los GPU de NVIDIA, el cual es transferido al adquirente. En este contexto, Dell o HPE no pueden añadir mucho margen sobre un componente que tiene un precio de mercado transparente y elevado. En este contexto, el beneficio real para los OEMs

---

37 The Information Staff (2025). "Deep Research Report: AI 2026 Predictions". The Information (September 17).

38 Pradhan, M. (2025). NVIDIA vs. TSMC: What's the better buy? *The Motley Fool* (December 30). Disponible en: https://www.fool.com/investing/2025/12/30/nvda-vs-tsm-whats-the-better-buy/

39 Miller, L. (2026). "Why an NVIDIA chip could supercharge TSMC's next rally". *Market B*eat (January 6). Disponible en: https://www.marketbeat.com/originals/why-an-nvidia-chip-could-supercharge-tsmcs-next-rally/

40 Morgan, T.P. "The only thing worse than selling AI servers is not selling them", The Next Platform (February 28). Disponible en: https://www.nextplatform.com/2025/02/28/the-only-thing-worse-than-selling-ai-servers-is-not-selling-them/

proviene de la venta combinada de almacenamiento, redes y servicios de consultoría, donde los márgenes son tres veces superiores.

Super Micro Computer (Supermicro) intentó romper este paradigma mediante una estrategia de "Time-to-Market" acelerado y especialización en ingeniería térmica (refrigeración líquida o DLC). Inicialmente, esto les permitió capturar una cuota de mercado del 10% y márgenes del 17%. Sin embargo, la ventaja tecnológica en hardware es de corto plazo. A medida que competidores como Dell entraron agresivamente en el mercado con precios competitivos para ganar cuota, y los costos de implementación de la refrigeración líquida aumentaron, los márgenes brutos de Supermicro se colapsaron al 11% hacia finales de 2025 [41]. Esto demuestra la fragilidad de la ventaja competitiva en el ensamblaje de servidores: sin un control sobre la propiedad intelectual del semiconductor o el software, la innovación en el segmento de ensamblaje de servidores se comoditiza rápidamente.

En el otro extremo del espectro están los *Original Design Manufacturers* como Foxconn, Quanta e Inventec. El modelo de negocio en este caso se basa en la generación de altos volúmenes, aunque a márgenes reducidos, trabajando directamente con los hiperescaladores para entregar productos ajustados a sus necesidades. Por ejemplo, Foxconn reporta márgenes brutos globales de apenas el 6.35%, incluso con la expansión de la IA [42]. Los hiperescaladores en muchos casos evitan adquirir servidores de los OEMs como Dell, y contratan directamente a los ODMs para ensamblar sus propios diseños de servidores, eliminando así intermediarios. De acuerdo con este modelo, el cliente compra las GPUs directamente a NVIDIA y las envía al ODM solo para el ensamblaje, reduciendo así al ODM como un proveedor de mano de obra y logística de ensamblaje, limitando aún más su capacidad de capturar valor [43].

1.2 Infraestructura de IA

En 2026, Estados Unidos dispone de una capacidad instalada de 2,528 centros de datos, de los cuales 40% de poder de cómputo (o sea 10.1 GW) está dedicado a IA. En el 2031, el número de centros aumentará a 3,226 y la capacidad dedicada a IA alcanzará 26.4 GW [44]. Este recurso es provisto por los centros de datos enfocados en la provisión de servicios de computación en la nube, originalmente controlado por los hiperescaladores, aunque ahora segmentado entre estos y las neo-clouds.

El mercado de hiperescaladores está estructurado alrededor de un oligopolio estable compuesto por AWS (Amazon Web Services), Microsoft Azure, y Google Cloud, y minoritariamente Oracle, quienes poseen la ventaja de la escala masiva, redes globales y ecosistemas de software empresarial profundamente arraigados. Sin embargo, su infraestructura heredada y las capas de virtualización pueden introducir latencia e ineficiencias en el entrenamiento de modelos masivos. Sus infraestructuras, diseñadas originalmente para cargas de trabajo web generales y microservicios, a menudo carecen de las optimizaciones necesarias para el entrenamiento de modelos masivos como los requeridos por IA, especialmente en su etapa de entrenamiento. Por ejemplo, la falta de soporte nativo generalizado para interconexiones de baja latencia como InfiniBand por parte de hiperescaladores en todas sus zonas de disponibilidad es una desventaja técnica frente a soluciones más modernas [45]. Es por ello que, si bien Meta no ofrece servicios de computación en la nube, esta opera grandes centros de datos para el desarrollo de sus modelos en lugar de depender de otros competidores.

---

[41] "The liquid cooling King: How super Micro Computer defied the skpetics to lead the AI infrastructure race into 2026". *Financial Content*. Disponible en: https://markets.financialcontent.com/stocks/article/marketminute-2026-1-1-the-liquid-cooling-king-how-super-micro-computer-defied-the-skeptics-to-lead-the-ai-infrastructure-race-into-2026

[42] Foxconn revenue hits $27 billion in November, driven by AI server growth (5 december, 2025). Disponible en: https://www.techinasia.com/news/foxconn-revenue-hits-27b-in-nov-driven-by-ai-server-growth

[43] Morgan, T.P. "The only thing worse than selling AI servers is not selling them", The Next Platform (February 28). Disponible en: https://www.nextplatform.com/2025/02/28/the-only-thing-worse-than-selling-ai-servers-is-not-selling-them/

[44] Gergs, L. (2026). *Where is AI data center demand growth being driven?* ABI Research (February 23).

[45] "GPU cloud providers in 2026". *Live Docs* (Nov, 24, 2025). Disponible en: https://livedocs.com/blog/cloud-gpu-providers-analysis

En competencia con los hiperescaladores, han surgido las "Neo-clouds" (CoreWeave, Lambda Labs, Nebius, Vultr), proveedores especializados que ofrecen acceso básico a clústeres masivos de GPUs. Estas empresas, de las cuales se cuentan 179 en la actualidad, en algunos casos se originaron inicialmente en el procesamiento de criptomonedas, han diseñado sus centros de datos específicamente para la densidad térmica y de red que requiere la IA [46]. Las "neo-clouds" comenzaron a crecer en el mercado de IA como respuesta a la demanda insatisfecha de GPUs proveniente de start-ups, laboratorios y empresas que no conseguían acceder al servicio, dado que los hiperescaladores controlaban un porcentaje importante de la oferta. Este tipo de jugadores ofrece contratos flexibles, aprovisionamiento rápido y configuraciones especializadas, facturando el precio de los GPUs a un 85% menos que el de los hiperescaladores. El segmento de neo-clouds está más fragmentado que el de los hiperescaladores: se estima que existen más de 100 empresas de este tipo en el mundo, aunque alrededor de 15 han alcanzado escala significativa en Estados Unidos [47].

Empresas como CoreWeave y Lambda Labs ofrecen precios por hora de GPU que son consistentemente un 50-70% más bajos que los de AWS o Google[1,3] Logran esto mediante una estructura de costos radicalmente diferente: no poseen la inmensa carga de servicios legados, operan con gastos generales corporativos mínimos y diseñan sus centros de datos específicamente para la densidad térmica extrema de las GPUs, maximizando así la eficiencia por metro cuadrado. Por ejemplo, CoreWeave ha pasado de ser un proveedor clave de infraestructura para laboratorios como OpenAI y Microsoft, gestionando una cartera de pedidos que supera los $30 mil millones. Su ventaja competitiva radica en la disponibilidad inmediata de hardware de última generación (GPUs H100/H200 de NVIDIA) y una arquitectura de red optimizada que maximiza su utilización, reduciendo así el tiempo y costo de entrenamiento.

Los "neo-clouds" ofrecen acceso a servidores físicos sin capas de virtualización que consumen recursos y redes InfiniBand de alto ancho de banda como estándar, lo cual es crítico para clústeres de entrenamiento distribuidos en más de 10,000 GPUs. Esta superioridad técnica les ha permitido ganar clientes como Microsoft (que paradójicamente usa CoreWeave para complementar su propia capacidad) y desarrolladores de modelos fundacionales como Mistral [48].

La respuesta de los hiperescaladores a la competencia de los "neo-clouds" ha sido pragmática: Microsoft, por ejemplo, ha firmado acuerdos para utilizar la infraestructura de CoreWeave, reconociendo implícitamente que necesitan capacidad externa para gestionar los picos de demanda mientras readecuan sus propias instalaciones.

De todas maneras, el principal riesgo del modelo de "neo-cloud" es la sostenibilidad de su capacidad para fijar precios. Si la escasez mundial de GPUs se reduce, su ventaja principal (disponibilidad inmediata) se erosiona. Para mitigar esto, CoreWeave se ha posicionado como un socio premium de NVIDIA, obteniendo acceso prioritario a los chips de última generación (como los GB200) antes que el mercado general, manteniéndose así en la frontera de la eficiencia [49]. Alternativamente, los neoclouds podrían ocupar posiciones de nicho en las que los hiperescaladores son menos efectivos (por ejemplo, apoyo a operadores regionales o laboratorios de IA soberanos que son más renuentes a utilizar servicios de hiperescaladores multinacionales). Otra opción podría ser apalancar su flexibilidad de precio y servicio para seguir apoyando a startups.

---

[46] Markets and Markets (2025). *AI Agents Market size, share and growth and latest trends* (April).

[47] Mazza, M., Sachdeva, P., Aruntyunyan, S. and Alatovik, T. (2025). *The evolution of neoclouds and their next moves*. McKinsey & Co. (November 19).

[48] Wheeler, K. (2025). "Top 10: GPU platforms for Deep learning". AI Magazine (October 8). Disponible en: https://aimagazine.com/top10/top-10-gpu-platforms-for-deep-learning

[49] Salanki, P. (2025). CoreWeave Ranks as #1 AI Cloud, Backed by SemiAnalysis's Platinum ClusterMAX™ Rating. Disponible en: https://www.coreweave.com/blog/coreweave-ranks-as-1-ai-cloud-backed-by-semianalysiss-platinum-clustermax-tm-rating

Adquisición y preparación de bases de datos

El eslabón de bases de datos está compuesto en primer lugar por propietarios de datos que controlan los derechos de propiedad intelectual y los ponen a disposición de los desarrolladores de modelos de manera gratuita o por un pago de derechos (Common Crawl para datos de la web, Wikipedia, repositorios de software como Github, contenido propietario, como Reddit, periódicos académicos como arXiv). Asimismo, las empresas editoriales se han asociado para ofrecer "data pools" relativamente desestructurados.

Prosiguiendo en el eslabón de datos, existen jugadores concentrados en la estructuración y el etiquetado de bases de datos. Por ejemplo, ScaleAI, una empresa recientemente adquirida por Meta, opera un modelo basado en proporcionar *Reinforcement Learning from Human Feedback* (RLHF), el proceso mediante el cual seres humanos evalúan y corrigen las respuestas de la IA para hacerlas seguras y útiles. Aunque sus márgenes brutos (~50%) son inferiores a los de una empresa de software puro debido al costo de la mano de obra global, su posicionamiento competitivo es importante. ScaleAI ha acumulado series de datos propietarios de correcciones humanas que son difíciles de replicar, convirtiéndose en un socio indispensable para OpenAI, Meta y el gobierno de EE. UU. [50]. Otro ejemplo del crecimiento explosivo de este eslabón es la empresa Handshake, que también depende de una red de contratistas para el etiquetado de datos. Por otra parte, algunos laboratorios adquirentes de datos, como xAI, han comenzado a internalizar servicios de etiquetado [51].

Ante el inminente agotamiento de los datos públicos de alta calidad en internet y las crecientes preocupaciones sobre la privacidad y los derechos de autor, ha surgido un mercado para los "Datos Sintéticos". Empresas como Gretel.ai venden la capacidad de generar series de datos artificiales que imitan estadísticamente los datos reales sin exponer información privada. Su modelo de negocio es más cercano al SaaS puro, cobrando por licencias de plataforma o volumen de datos generados, y se está convirtiendo en una pieza clave para entrenar modelos en sectores como finanzas y salud donde los datos reales son escasos o no pueden ser usados por razones de privacidad [52].

Como se describe en el capítulo anterior, para que los modelos puedan ser usados en empresas, necesitan acceso a información privada actualizada y una "memoria" a largo plazo. Aquí es donde son aplicadas las bases de datos vectoriales y la arquitectura RAG (*Retrieval-Augmented Generation*). La función de estas bases de datos es almacenar la información en forma de vectores matemáticos (embeddings) para permitir búsquedas semánticas rápidas. Este mercado esta segmentado entre firmas especializadas y los proveedores de bases de datos tradicionales.

Entre los especialistas, Pinecone ha desarrollado un modelo de *Database-as-a-Service* (DBaaS) totalmente gestionado y cerrado. Su propuesta de valor es la simplicidad y el tiempo reducido para su uso por los desarrolladores de modelos. Si bien Pinecone cobra por uso y almacenamiento de datos, este modelo puede volverse prohibitivo para grandes volúmenes de datos dado que los costos escalan linealmente. Por otro lado, Weaviate ofrece un modelo más flexible de "Open Core". Este permite a los usuarios alojar la versión de código abierto (ideal para necesidades de soberanía de datos) o utilizar su servicio gestionado en la nube. Esta flexibilidad atrae a desarrolladores que necesitan control granular sobre la infraestructura y los costos escalables [53].

---

[50] Sacra (2025). ScaleAI. Disponible en: https://sacra.com/c/scale-ai/

[51] Hornstein, J. and Weinberg, C. (2026). "Handshake, Mercor revenue surges on demand for human contractors to train AI", The Information (April 13).

[52] Gretel (2026). What is synthetic data creation? Disponible en: https://www.gretel.ai/technical-glossary/what-is-synthetic-data-creation

[53] Ahuja, H. (2025). "Pinecone vs. Weaviate: The Trade-offs you only discover in production". DZone (November,18). Disponible en: https://dzone.com/articles/pinecone-vs-weaviate-the-trade-offs-you-only-disco

El mayor riesgo competitivo para los proveedores de datos especializados es que las bases de datos tradicionales incorporen capacidades vectoriales como una mera "funcionalidad" adicional. Esto puede observarse en PostgreSQL con su extensión proveedor, MongoDB y Elasticsearch, las cuales ya han añadido soporte vectorial. Para la gran mayoría de las aplicaciones empresariales, el rendimiento de proveedor integrado en su base de datos existente es preferible a la complejidad y costo de contratar un proveedor especializado nuevo. Esta dinámica presiona a los especialistas a innovar en características avanzadas (como búsquedas híbridas o gestión de grafos) para justificar su viabilidad competitiva [54].

Desarrollo de modelos fundacionales

Los modelos fundacionales son redes neuronales de gran escala entrenados en bases de datos que pueden ser texto (LLMs), multimodales (texto, audio, imágenes), visión de computadora, y modelos de programación. En términos generales, existen dos tipos de modelos fundacionales: de código cerrado y abierto. Los modelos abiertos permiten su adaptación a diferentes usos a partir del acceso público.

Los modelos cerrados (por ejemplo, GPT 4o de OpenAI, RLHF de Anthropic, Gemini 2.0 de Google [55]) mantienen su liderazgo en tareas de "inteligencia de frontera", razonamiento complejo y multimodalidad avanzada. Son la opción elegida para prototipado rápido y aplicaciones donde la precisión extrema justifica el costo. Su modelo de negocio está basado en el concepto de *API-as-a-Service* (cobro por millón de tokens [56] procesados) y suscripciones por empresa. La propuesta de valor de los modelos cerrados está basada en la calidad superior del razonamiento, la seguridad, el control ético y, la indemnización legal por derechos de autor. Estos modelos son particularmente adecuados para ser implementados en grandes corporaciones funcionando en sectores regulados, como servicios financieros, salud, y defensa, que priorizan la estabilidad y la responsabilidad sobre el costo [57].

Los modelos de código abierto, liderado por Llama (Meta), Pixtral Large (Mistral), y DeepSeek-V3 (DeepSeek), han demostrado que es posible obtener capacidades de razonamiento, codificación y razonamiento sin incurrir en los costos de licencia perpetuos de los modelos cerrados, lo que ha alterado la economía de la implementación de IA. Los modelos abiertos están siendo adoptados en entornos corporativos para tareas de alto volumen y menor complejidad. Estos permiten a empresas mantener el control sobre sus datos, realizar alineamiento profundo (*fine-tuning*) sin violar la propiedad intelectual y controlar sus costos unitarios al alojar los modelos en su propia infraestructura [58].

Mistral representa un enfoque híbrido y pragmático. Lanza modelos abiertos de alta eficiencia (como Mixtral 8x7B) para ganar cuota de conocimiento entre los desarrolladores y establecer un estándar técnico. Sin embargo, el startup francés monetiza a través de modelos "frontera" más grandes disponibles solo vía API comercial y mediante servicios de consultoría y personalización para empresas que requieren despliegues privados. Su modelo de negocio se asemeja al de "Open Core" tradicional del software empresarial [59].

En el caso de Meta, el modelo abierto es una maniobra estratégica defensiva ya que, al establecer su modelo fundacional como el estándar para la industria, erosiona el poder de mercado de OpenAi y Google. Al ofrecer Llama gratuitamente, Meta comoditiza el complemento de su negocio principal (la

---

54 Sverdlik, D. (2025). "Vector database selection guide: Pinecone vs. Weaviate vs. Qdrant for enterprises". Xenoss (July 11). Disponible en: https://xenoss.io/blog/vector-database-comparison-pinecone-qdrant-weaviate

55 Med-Gemini es completamente cerrado mientras que Claude 3.5 y ChatGPT 4 proveen acceso limitado mediante APIs.

56 En IA, el token es uno de los bloques fundamentales de datos (texto, imagen, audio) necesario para procesar el modelo y generar respuestas basadas en la interrelación entre unidades. Tokens son usados para gestionar costo de computación.

57 Gordon, D. (2025). "Open-source vs. closed-source AI: which model should your enterprise trust", VKTR (October 27). Disponible en: https://www.vktr.com/ai-platforms/open-source-vs-closed-source-ai-which-model-should-your-enterprise-trust/

58 Or, B. (2025). "Open vs. Closed LLMs in 2025: Strategic Tradeoffs for enterprise AI". *Data Science Collective* (June 6). Disponible en: https://medium.com/data-science-collective/open-vs-closed-llms-in-2025-strategic-tradeoffs-for-enterprise-ai-668af30bffao

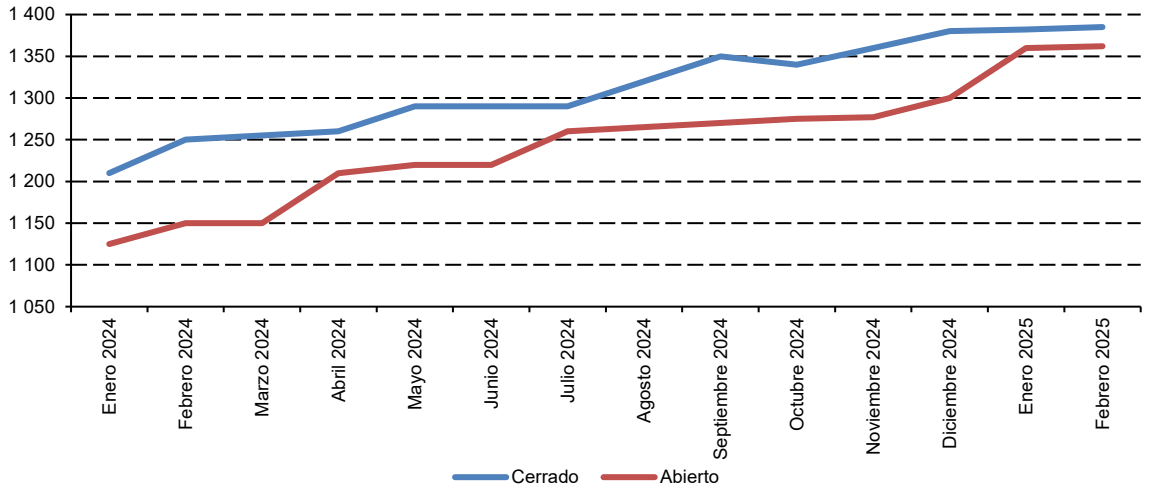
59 "Mistral AI: The OpenAI competitor you need to know about", *Just Think* (May 24). Disponible en: https://www.justthink.ai/blog/mistral-ai-the-openai-competitor-you-need-to-know-about

inteligencia artificial) para evitar que competidores como Google o Apple controlen la plataforma subyacente. Esto es así dado que Meta no necesita generar fondos vendiendo el modelo. Su modelo de negocio se basa en el mejoramiento de la eficacia de sus anuncios publicitarios, manteniendo a los usuarios en sus plataformas gracias a una IA mejorada por la comunidad global de desarrolladores [60].

Una categoría diferenciada de modelos es la denominada modelos derivados, o modelos basados en Llama. Estos usan el modelo de Meta como punto de partida, pero refinan los pesos decisorios en base a una serie de datos específicos para cumplir ciertas tareas. Por ejemplo, LatamGPT, el modelo desarrollado por el Centro Nacional de Inteligencia Artificial de Chile (CENIA) junto con instituciones latinoamericanas, es un caso de modelo derivado de Llama, adaptado a cultura y lenguajes latinoamericanos.

La dinámica competitiva de los modelos fundacionales ha experimentado un cambio radical. Hasta 2023, existía una brecha de rendimiento clara entre los modelos propietarios de frontera (como GPT-4) y los modelos de código abierto (como Llama 3.1 o DeepSeek). Sin embargo, en el 2025 esta brecha de desempeño se ha reducido significativamente. Análisis recientes indican que la diferencia de rendimiento en benchmarks técnicos entre los modelos cerrados líderes y los mejores modelos abiertos se ha reducido de un 8% en el 2023-4 a apenas un 1.7% (ver gráfico 1) [61].

Gráfico 1
Desempeño de modelos cerrados vs. abiertos



Fuente: LMSYS, AI Index 2025.

Finalmente, corresponde mencionar que un estudio reciente sobre la plataforma Hugging Face proporciona evidencia sobre el efecto competitivo de los modelos abiertos [62]. El lanzamiento de modelos abiertos de alta calidad reduce drásticamente las barreras de entrada para nuevos desarrolladores, pero destruye el valor de las capacidades especializadas previas. Los desarrolladores incumbentes en áreas ahora "resueltas" por el modelo abierto se ven obligados a retirarse hacia nichos más complejos o a evolucionar hacia aplicaciones verticales para mantener su diferenciación.

---

[60] Green, T. (2025). "The Great AI subsidy squeeze". Rawveg.substack. Disponible en: https://dev.to/rawveg/the-great-ai-subsidy-squeeze-4pjf

[61] HAI (2025). Artificial Intelligence Index Report 2025.

[62] Wu, D., L. Z. (2025) Foundation Models and AI Innovation: Evidence from the Hugging Face Platform Face Platform. ResearchGate (October). Disponible en: https://www.researchgate.net/publication/396561257_Foundation_Models_and_AI_Innovation_Evidence_from_the_Hugging_Face_Platform_Face_Platform

Software y herramientas

El primer componente clave de este eslabón son los chatbots de propósito general. Al momento, estos incluyen ChatGPT (OpenAI), Gemini (Google), y Claude (Anthropic), entre otros. El número de usuarios semanales de ChatGPT ha superado los 800 millones, de los cuales 90% provienen de países al exterior de EE. UU. y Canadá, mientras que Gemini ya ha alcanzado 350 millones. La dinámica competitiva entre estos se está acelerando, particularmente entre Gemini avanzando sobre ChatGPT. Esta competencia se basa en la habilidad de combinar el Chatbot con negocios clásicas como búsqueda, publicidad y ventas. Obviamente, Goggle posee la ventaja que le otorga el posicionamiento dominante en el mercado de búsqueda. Su apalancamiento, combinado con los avances en el desempeño de Gemini, dan a Google una oportunidad competitiva importante. A juzgar por las tendencias de tráfico, esta posibilidad podría ocurrir: la participación de Gemini en el tráfico total de chatbots en el último mes ha crecido 14.36% mientras que la de ChatGPT ha disminuido 5.21% (ver cuadro 5).

Cuadro 5
Participación en tráfico total de Chatbots, diciembre de 2025
(En porcentajes)

Chatbot	Cuota de tráfico	Cambio mes a mes
ChatGPT	65,97	-5,21
Gemini	13,23	+14,36
Deepseek	3,64	-2,95
Perplexity	2,13	-13,64
Claude	1,97	-8,47
Copilot	1,15	-1,95

Fuente: Jacobson, A. LLM adoption is flat. The information.

Si 2023 fue el año de entrada de los Chatbots en el mercado, el ciclo 2025-2026 marca el advenimiento de la IA Agéntica. La distinción entre ambos es fundamental: un agente no solo conversa, sino que también actúa. Los agentes poseen capacidades de planificación, memoria a largo plazo y uso de herramientas. Estos pueden navegar por la web, interactuar con APIs de software empresarial (ERPs, CRMs), y ejecutar flujos de trabajo de múltiples pasos sin supervisión humana constante.

Se proyecta que el mercado de agentes de IA explote desde los $7.6 mil millones en 2025 hasta más de $180 mil millones en 2033[63]. Este cambio redefine la propuesta de valor del software: se pasa de herramientas que "ayudan" al usuario a herramientas que "desempeñan" el trabajo del usuario. El termino clave a considerar es cuantas tareas pueden llegar a ser desempeñadas por el agente, en la medida de que esto implica consecuencias importantes en el mercado laboral.

Desarrollo e implementación de aplicaciones

Como se detalla en el diagrama que esquematiza la cadena de valor de la IA, el último eslabón está ocupado por dos tipos de firmas: i) los desarrolladores de aplicaciones de IA orientados a sectores verticales (como, por ejemplo, salud, y servicios legales), y ii) integradores de sistemas, es decir firmas que proveen el servicio de adaptación de plataformas de IA a las necesidades del negocio de clientes.

[63] Grandview Research (2025). AI Agents Market (2026-2033).

El modelo de negocio tradicional de SaaS horizontal (*Software as a Service*) —es decir aplicaciones genéricas para todos los sectores económicos sin distinción— está siendo desafiado por la IA Vertical. Estas consisten en aplicaciones construidas sobre los modelos fundacionales del cuarto eslabón, pero entrenadas en la ontología, flujos de trabajo y regulaciones de una industria específica, con bases de datos específicos de una industria. En muchos casos, los modelos ajustados o en arquitecturas más eficientes que apoyan estas aplicaciones, y presentan un desempeño de mayor precisión que los grandes modelos[64]. En consecuencia, estos modelos requieren menos memoria, pueden servir dispositivos individuales y demuestran menores costos para entrenar. En muchos casos, estos modelos más pequeños dependen de modelos generales accedidos vía API como modelos ajustados, o están basados en arquitecturas multi-modelo o sistemas complementados con mecanismos de recuperación. Estos modelos pueden ser más precisos en tareas específicas, acotadas y estructuradas, y bajo condiciones de ajuste fino y disponibilidad de datos de calidad. En particular, el mercado de aplicaciones verticales de IA se ha desarrollado en tres sectores:

i) **Sector legal:** herramientas como Harvey asisten a estudios de abogados en la revisión de contratos, litigios y diligencia debida, integrando conocimientos jurídicos que un modelo generalista como ChatGPT no posee con la precisión necesaria. Desde su lanzamiento, Harvey ha alcanzado a ofrecer servicio a 42% de los 100 estudios de abogados más importantes de Estados Unidos. Su valor está basado en el hecho de que su alineamiento se basa en documentos legales confidenciales, propiedad del estudio que lo adopta[65].

ii) **Salud:** plataformas como Abridge o Ambience Healthcare escuchan las consultas médico-paciente y generan automáticamente la documentación clínica estructurada en el registro médico electrónico (EHR), reduciendo directamente el problema de la sobrecarga de trabajo del profesional de la salud[66].

iii) **Programación:** herramientas como Cursor y GitHub Copilot están evolucionando de ser auto-completadores de código apoyando a ingenieros de software autónomos que pueden generar bases de código completas o escribir módulos enteros a partir de especificaciones en lenguaje natural.

Por otro lado, el mercado de integradores de sistemas para implementar soluciones de IA está extremadamente fragmentado, constituido por firmas de consultoría estratégica como McKinsey y Boston Consulting Group, así como integradores de sistemas. Accenture, el jugador más importante controla tan solo el 7.2%, aunque potenciado con asociaciones con OpenAI, Microsoft, NVIDIA, y Google. Los otros jugadores importantes incluyen Deloitte (3.4%), IBM (2.2%), McKinsey (2.2%), CapGemini (1.9%), BCG (1.9%), Cognizant (1.0%), EY (1.0%), Bain (1.0%), y otros con menor cuota. Todos estos jugadores están invirtiendo sumas importantes en el segmento de servicios profesionales para capturar posiciones de mercado.

2. Movimientos de integración vertical en la cadena de valor de la inteligencia artificial de Estados Unidos

Casi simultáneamente con la estructuración de la cadena de valor fragmentada, las empresas que ocupaban posiciones únicas comienzan a buscar oportunidades de diversificación en otros eslabones. Pero esta reintegración no fue una respuesta a la madurez de los mercados, o las condiciones de integración vertical "clásicas" como había conceptualizado Stigler originalmente. Oliver Williamson consideró que la reintegración de la cadena de valor es ocasionada por dos tipos de factores. Por un

[64] Chat GPT 4 está basado en 1.8 trillones de parámetros.
[65] Bull, A. (2025). The rise of the AI application layer.
[66] Bull, A. (2025). The rise of the AI applications Layer. Medium (November 25). Disponible en: https://medium.com/b8125-fall2025/the-rise-of-the-ai-application-2a17aa712e33

lado, las empresas tienden a integrarse verticalmente buscando el control estratégico de ciertos estadios de la cadena (por ejemplo, la relación con el usuario final), o buscando reducir los costos de transacción entre estadios de la cadena[67]. Otra motivación para la entrada en otros estadios de la cadena de valor es la búsqueda de nuevas posiciones de rentabilidad superior, o para preservar posiciones de dominancia.

Tendencias a la integración vertical en la cadena de la IA

Como se demostró en el análisis de la cadena de valor de la IA, el paradigma dominante de organización industrial del ecosistema digital original fue la especialización horizontal y la modularidad: los desarrolladores de contenido proveían el insumo que las plataformas de agregación usaban para capturar posiciones en el mercado y entregar valor a usuarios mediante servicios de telecomunicaciones. Si bien se manifestaban intentos de integración vertical (por ejemplo, cuando las empresas de telecomunicaciones intentaban entrar en el negocio de distribución de contenidos), los esfuerzos fracasaron.

Al mismo tiempo, la cadena de sistemas de información estaba compuesta por empresas que desarrollaban semiconductores (Intel, por ejemplo), fabricantes de equipamiento (como discos duros), Microsoft que desarrollaba el sistema operativo, y un universo de proveedores de software independientes que construían las aplicaciones. Este modelo permitía una innovación rápida a través de capas estandarizadas donde cada actor optimizaba su componente específico, sin intenciones de integración vertical.

La irrupción de la Inteligencia Artificial (IA) ha transformado estas fronteras modulares, resultando en un proceso de integración vertical agresivo. En el contexto actual, la eficiencia computacional y energética se ha convertido en la variable crítica de competitividad. Los modelos de IA, que superan el billón y hasta alcanzan el trillón de parámetros, imponen demandas de latencia, ancho de banda de memoria y consumo energético que la arquitectura modular tradicional —con sus cuellos de botella inherentes entre componentes discretos— ya no puede satisfacer de manera rentable. Como resultado, las fronteras entre los diseñadores de GPUs, los hiperescaladores, y los desarrolladores de modelos fundacionales se están diluyendo. Las empresas que anteriormente eran clientes y proveedoras en una cadena de suministro lineal de capacidad técnica ahora compiten y colaboran simultáneamente en múltiples eslabones, buscando controlar la totalidad de la cadena de valor.

Reiterando la teoría de Williamson, las empresas tienden a integrarse verticalmente cuando el mecanismo de mercado se vuelve ineficiente debido a la alta especificidad de los activos y la incertidumbre contractual. Bajo las nuevas condiciones tecnológicas de la IA, los activos se han vuelto hiperespecíficos en cada eslabón de la cadena, pero dependientes de otros eslabones. Por ejemplo, el entrenamiento de un modelo fundacional que corre en equipamiento controlado por una firma ubicada en un eslabón distinto de la cadena genera ineficiencias críticas causadas por la fricción transaccional implícita en tener que adaptar ese modelo a un hardware genérico. Estas ineficiencias se manifiestan en latencia en la transferencia de datos, consumo excesivo de energía por operaciones innecesarias y la incapacidad de optimizar el compilador de software para la arquitectura del GPU que está siendo usado. La integración vertical surge, entonces, como un mecanismo para resolver estas externalidades técnicas, eliminando la necesidad de contratos complejos y permitiendo una coordinación mejorada entre el modelo desarrollado y el GPU[68]. Este enfoque implica que la arquitectura del GPU se diseña específicamente para acelerar el algoritmo dominante, y viceversa. La modularización de la cadena de valor original, que permitía mezclar y combinar hardware y software de diferentes proveedores,

---

[67] Williamson, O. (1985). The Economic Institutions of Capitalism. New York: The Free Press, pp. 85-130.

[68] Colle, A., Donaldson, K., and Dragone, M. (2025) "Co-design methodology for rapid prototyping of modular robots in care settings", *Front Robot AI* (May 22); 12 :1581506. Disponible en: https://pmc.ncbi.nlm.nih.gov/articles/PMC12137090/

introduce sobrecargas antieconómicas. La investigación académica reciente corrobora que los sistemas embebidos y de alto rendimiento logran eficiencias superiores cuando se rompen las barreras entre eslabones, permitiendo, por ejemplo, que el software gestione directamente la memoria o la energía consumida [69]. En estos términos, la organizacional industrial tiende a "reflejar" la arquitectura técnica de sus productos. A medida que la arquitectura técnica de la IA se vuelve más integrada y menos modular, la estructura industrial tiende a integrarse en monopolios verticales o ecosistemas cerrados, abandonando la estructura desagregada del pasado.

La integración vertical de hiperescaladores hacia el eslabón de equipamiento

Ante los márgenes brutos de NVIDIA, que superan el 70%, y la escasez crónica de suministro de GPUs, los hiperescaladores, como Amazon Web Services (AWS), Google Cloud (GCP) y Microsoft Azure, están ejecutando una estrategia de integración vertical agresiva. Para ello, están diseñando sus propios semiconductores personalizados o ASICs (*Application-Specific Integrated Circuits*) con el objetivo de reducir la dependencia de NVIDIA y optimizar sus costos operativos.

En primer lugar, esta tendencia hacia la diversificación de la provisión de GPUs por parte de los hiperescaladores es una respuesta económica racional. Para cargas de trabajo masivas y estables, como la inferencia de modelos que sirven a miles de millones de usuarios (ej. ChatGPT, búsquedas de Google potenciada por Gemini), los ASICs ofrecen una eficiencia por watt consumido y por dólar superior a las GPUs de propósito general. Es por ello que los hiperescaladores han desplegado maniobras de integración hacia atrás en la cadena:

- **Google:** ha liderado este camino con sus TPUs (*Tensor Processing Units*), llegando a la sexta generación (*Trillium*), diseñada específicamente para acelerar sus modelos de *Transformer*.
- **AWS:** ha desplegado las familias *Trainium* (para entrenamiento) e *Inferentia* (para inferencia) para la IA, ofreciendo instancias de menor costo en su nube.
- **Microsoft:** ha entrado en la competencia con sus GPUs Maia, optimizados para la infraestructura de Azure y los modelos de OpenAI.

Asimismo, la entrada en el eslabón de semiconductores por parte de los hiperescaladores se justifica desde el punto de vista competitivo. Al optimizar las herramientas de software (como TensorFlow o PyTorch) para su hardware propietario, los hiperescaladores crean ecosistemas cerrados donde el rendimiento máximo solo se logra permaneciendo dentro del contexto del proveedor [70].

De hecho, Google estableció el modelo de integración vertical en IA mucho antes del auge actual de la IA generativa. Con el desarrollo de sus *Tensor Processing Units* (TPU), ahora en su sexta generación conocida como *Trillium*, Google ha logrado una sinergia hardware-software casi perfecta. A diferencia de sus competidores que inicialmente dependían de capas de abstracción como CUDA de NVIDIA, Google co-diseñó su marco de aprendizaje profundo (*TensorFlow* y posteriormente JAX) junto con su arquitectura de TPU. Esta integración permite a Google optimizar operaciones a nivel de compilador (XLA) que serían imposibles a obtener en hardware de terceros. El chip *Trillium* no solo ofrece mejoras en rendimiento bruto, sino que está diseñado para la eficiencia energética y la escalabilidad dentro de los grupos de TPUs de Google, lo que de hecho es una arquitectura de red propietaria que evita los cuellos de botella de Ethernet tradicionales. La colaboración continua con Broadcom para el diseño físico de estos GPUs demuestra cómo la integración vertical no significa

[69] Wali, K. (2024). Hardware-software Co-design for power efficient Edge-AI Systems", *Artificial Intelligence Machine Learning Data Science* 2024 2(4), 2754-2760.

[70] Narechania, T. and Sitaraman, G. (2024). "An antimonopoly approach to governing artificial intelligence". *Yale Law & Policy Review* 43, Issue (Fall). Disponible en: https://yalelawandpolicy.org/antimonopoly-approach-governing-artificial-intelligence

necesariamente "hacerlo todo solo", sino controlar la propiedad intelectual (IP) crítica mientras se subcontrata la implementación física y la fabricación [71].

Amazon Web Services ha adoptado un enfoque diferente, centrado en ofrecer opciones de GPUs personalizadas directamente a sus clientes de nube bajo el modelo de "*Utility Computing*". Sus líneas de semiconductores *Trainium* e *Inferentia* están diseñadas para ofrecer el menor costo por inferencia en la nube, atacando directamente el posicionamiento de NVIDIA. La estrategia de AWS es particularmente notable por su asociación con Marvell Technology y Alchip para el desarrollo de la plataforma Trainium. Al ofrecer un desempeño similar al de H100 de NVIDIA a un costo significativamente menor, AWS incentiva a los desarrolladores a migrar fuera del ecosistema CUDA, utilizando su propio SDK para IA generativa (Neuron) como plataforma. Esta es una jugada clásica de integración vertical para capturar el excedente del consumidor sensible al precio y clientes dentro del ecosistema de AWS.

Microsoft llegó más tarde a las maniobras del GPU personalizado, pero ha acelerado rápidamente con el lanzamiento de Azure Maia y Azure Cobalt (CPU basada en arquitectura Arm) [72]. La motivación de Microsoft es la optimización de sus cargas de trabajo de IA generativa masivas, específicamente los servicios de OpenAI (ChatGPT) y Microsoft 365 Copilot. El GPU Maia 100 está diseñado para responder a las necesidades específicas de los modelos GPT de OpenAI, optimizando la precisión numérica y el ancho de banda de memoria para los patrones de acceso de estos modelos transformadores. Al controlar también la CPU (Cobalt), Microsoft elimina el "sobrecargo" que debía pagar a Intel o AMD, logrando así una integración vertical completa en el rack de servidores. Esto no solo mejora los márgenes brutos de los productos de IA de Microsoft, sino que también le da un control sobre la hoja de ruta de rendimiento de su infraestructura crítica, reduciendo la vulnerabilidad a los ciclos de productos de proveedores externos.

A pesar de todos estos movimientos por parte de hiperescaladores, NVIDIA sigue manteniendo la primacía en el entrenamiento de modelos de frontera y en el mercado empresarial general, mientras que los ASICs capturan volumen en las cargas de trabajo internas de los hiperescaladores.

La integración vertical de hiperescaladores hacia el eslabón de laboratorios de modelos fundacionales

En paralelo con la integración vertical hacia atrás en la cadena, los hiperescaladores entran en el eslabón de modelos fundacionales, sea por una posición propietaria (completa o mediante el control de acciones) o mediante alianzas estratégicas. El caso paradigmático de control completo es el de Google, quien ha adquirido una posición en el desarrollo de modelos fundacionales mediante la fusión de dos unidades bajo su control: *Deep Mind* y *Google Mind* que lideran el desarrollo de Gemini.

Por otra parte, las estrategias de inversión accionaria se registran en la inversión de Amazon en Anthropic y la de Microsoft en OpenAI. Estas alianzas aseguran que, independientemente de qué modelo triunfe, el valor económico fluya hacia el proveedor de servicios de la nube subyacente a través de compromisos de gasto en cómputo [73]. Sin embargo, el modelo que guía esta estrategia de inversión de hiperescaladores en desarrolladores de modelos fundacionales no es homogéneo.

Por ejemplo, la relación entre Microsoft y OpenAI, inicialmente definida por una exclusividad contractual y una visión unificada, ha entrado en una fase de mayor complejidad jurídica y financiera hacia finales de 2025. Aunque Microsoft mantiene su posición como inversor principal, la reestructuración de OpenAI hacia una *Public Benefit Corporation* (PBC) ha alterado fundamentalmente

---

[71] Taw Y. (2025). "The Golden age of customized chips". *Neuberger Berman* (February 5). Disponible en: https://www.nb.com/en/global/insights/the-golden-age-of-customized-ai-chips

[72] Siegel, J. (2023). With a systems approach to chips, Microsoft aims to tailor everything "from silicon to service" to meet AI demand (November 15). Disponible en: https://news.microsoft.com/source/features/ai/in-house-chips-silicon-to-service-to-meet-ai-demand/

[73] Ela Glowicka & Jan Malek: "Digital Empires Reinforced? Generative AI value chain". *Network Law Review*. Disponible en: https://www.networklawreview.org/glowicka-malek-generative-ai/

los términos de la relación. En ese marco, la participación accionaria de Microsoft en OpenAI se ha diluido del 32.5% al 27% en la nueva entidad. Asimismo, se ha establecido una cláusula de terminación de la propiedad intelectual de OpenAI para Microsoft cuando la primera alcance el estadio de Inteligencia artificial general, y la pérdida del derecho de preferencia de Microsoft en la provisión de recursos computacionales[74]. Como puede verse, esta dilución no es meramente financiera. La redefinición refleja una maniobra de OpenAI para diversificar su base de capital y reducir la influencia de un solo actor corporativo. Por otro lado, el objetivo estratégico de Microsoft en este caso es mitigar el riesgo de concentrar todo su desarrollo de modelos fundacionales en OpenAI.

A diferencia del modelo de inversión de Microsoft con OpenAI, la coordinacion entre Amazon y Anthropic incluye elementos clave del "AI stack". Anthropic se ha comprometido a utilizar los chips AWS *Trainium* e *Inferentia* definidos arriba, así como AWS en tanto plataforma principal para el entrenamiento y despliegue de modelos futuros, alejándose deliberadamente de la dependencia exclusiva de las GPUs de NVIDIA[75]. Esta colaboración implica una ingeniería conjunta profunda: los equipos técnicos de Anthropic están trabajando directamente con Annapurna Labs (la división de diseño de chips de Amazon) para escribir *kernels* de bajo nivel que optimicen el hardware desde el GPU hacia arriba en las capas[76]. Para el mercado empresarial, esto posiciona a los modelos Claude como "infraestructura central" dentro de Amazon Bedrock, un servicio de AWS que provee el acceso a varios modelos fundacionales via APIs. Al mismo tiempo, la inversión financiera de Amazon en Anthropic ya alcanza los US$ 8 mil millones.

Una tercera avenida de integración vertical de hiperescaladores en los desarrolladores de modelos funcionales son las alianzas estratégicas. Por ejemplo, Microsoft y Anthropic sellaron una alianza estratégica mediante la cual esta última se compromete a usar US$ 30 mil millones en la adquisición de capacidad de computación de Azure, mientras que Microsoft investirá US$ 5 mil millones en Anthropic y se compromete a ofrecer los modelos Claude de Anthropic a sus clientes.

Las alianzas estratégicas crean una red de interrelaciones múltiples de los grandes participantes de la cadena de valor (ver cuadro 6).

Cuadro 6
Alianzas estratégicas de los principales participantes en la cadena valor para entrar en negocios adyacentes o complementarios

Empresa	Servidores de GPUs	Clusteres de entrenamiento de IA	Alquiler de servidores de IA	API para desarrolladores de IA	Estado del arte LLM	Apps de IA para empresas	Apps para consumidsores individuales	Dispositivos de IA	Robots humanoides
Google	X	X	X	X	X	X	X	X	X
Microsoft	X	X	X	X	X	X	X		
Amazon	X	X	X	X	X	X	X	X	X
OpenAI	X	X		X	X	X	X	X	X
Meta	X	X		X	X		X	X	X
xAI	X	X		X	X	X	X		X
NVIDIA	X	X	X	X	X				X
Apple	X	X			X		X	X	X
Anthropic		X		X	X	X	X		

Fuente: Amir Efrati (2025). OpenAI, Meta and Their AI Rivals Ramp Up Turf Wars and Partnerships, in Three Charts, The Information.

[74] Waters, J. (2025). Microsoft, OpenAI rewrite partnership rules ahead of AGI race", Pure AI (November 3). Disponible en: https://pureai.com/articles/2025/11/04/microsoft-openai-rewrite-partnership-rules-ahead-of-agi-race.aspx

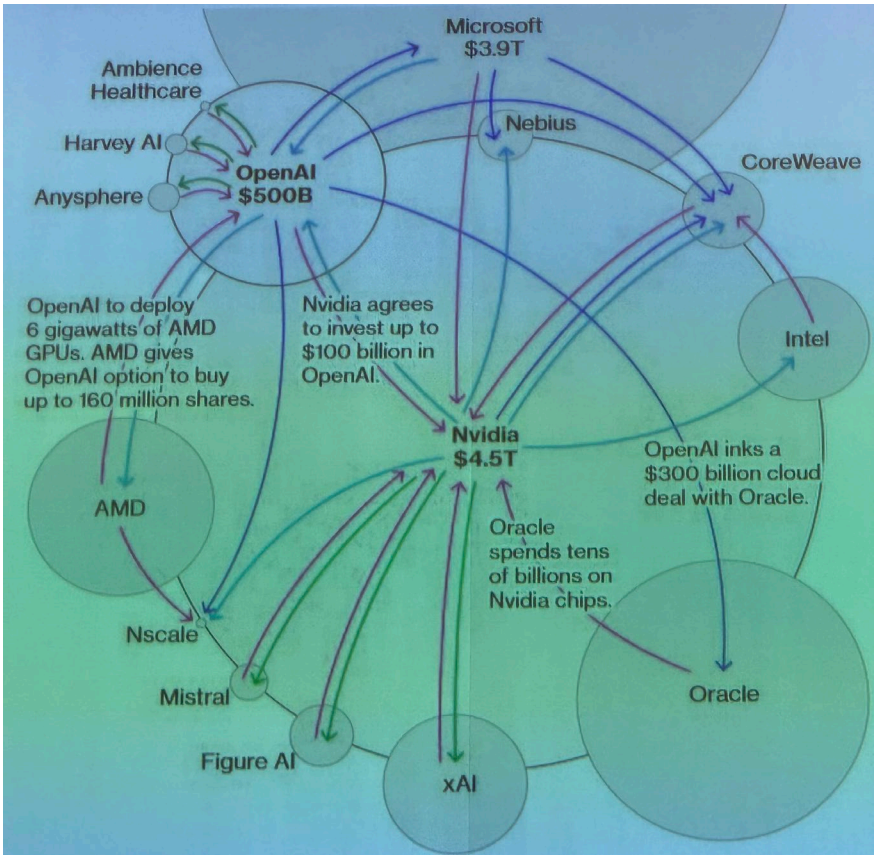
[75] AWS (2023). *Amazon and Anthropic announce strategic collaboration to advance generative AI* (September 25). Disponible en: https://www.aboutamazon.com/news/company-news/amazon-aws-anthropic-ai

[76] Anthropic (2024). *Powering the next generation of AI development with AWS* (November 22). Disponible en: https://www.anthropic.com/news/anthropic-amazon-trainium

La integración vertical de proveedores de GPUs hacia infraestructura y modelos fundacionales

NVIDIA, el principal proveedor de GPUs, también ha comenzado a integrarse hacia adelante en la cadena de valor. En primer lugar, ha firmado acuerdos de colaboración con Foxconn para la manufactura de GPUs, con el objetivo de resolver el cuello de botella de TSMC resultante de sus límites en la producción de semiconductores. Asimismo, en una vinculación con el eslabón siguiente, NVIDIA ha invertido en CoreWeave, una neo-cloud [77]. Sin embargo, el avance más importante de proveedores de GPUs en infraestructura es la inversión de US$ 100 millones realizada por NVIDIA en OpenAI para acelerar su despliegue en capacidad de procesamiento. Esta inversión está orientada a satisfacer la necesidad de GPUs por parte de OpenAI, a su vez ser provistos por NVIDIA, por lo cual este movimiento ha sido calificado de "circular" (ver diagrama 13).

Diagrama 13
Inversiones cruzadas de NVIDIA y OpenAI



Fuente: Bloomberg.

La integración de laboratorios de modelos fundacionales en el suministro de datos

El caso paradigmático de la integración vertical de modelos fundacionales hacia atrás en la provisión de datos es la adquisición de ScaleAI por parte de Meta. En junio del 2025, Meta adquirió 49% de las acciones de ScaleAI por aproximadamente US$ 14.3 mil millones. Con ello, Meta se asegura el

[77] NVIDIA (2025). *NVIDIA and partners build America's AI infrastructure and create blueprint to power the next industrial revolution* (October 28). Disponible en: https://nvidianews.nvidia.com/news/nvidia-partners-ai-infrastructure-america

acceso exclusivo de datos de alta calidad para el entrenamiento de sus modelos fundacionales. ScaleAI opera una fuerza global de contratistas en Kenia, Filipinas y Venezuela para el etiquetado manual de imágenes, texto y video para aplicaciones de machine learning. Asimismo, ScaleAI contiene una capacidad de generación de datos sintéticos. A partir de la adquisición, debido a preocupaciones de índole competitiva, Google interrumpió proyectos con ScaleAI, mientras que OpenAI comenzó a disminuir su involucramiento, de la misma manera que xAI[78].

La entrada de laboratorios de modelos fundacionales siguiendo el modelo de Meta no es una tendencia universal. Por ejemplo, OpenAI ha construido un equipo centralizado para gestionar sus relaciones con firmas de etiquetado. Anthropic, por otro lado, prosigue un modelo descentralizado donde diferentes equipos firman contratos específicos con empresas de etiquetado[79]. De todas maneras, en paralelo con la presentación de ofertas públicas de laboratorios, es de esperar que el mercado ponga presión en los laboratorios para que estos centralicen la adquisición de datos de una manera más disciplinada. En este contexto, la posibilidad de que otros laboratorios imiten la estrategia de Meta es probable, excepto que estos no consideran la opción de tener que gestionar una fuerza de trabajo compuesta por miles de empleados como muy atractiva. Por otro lado, un modelo descentralizado compuesto por múltiples proveedores permite a los laboratorios optimizar sus costos de adquisición haciéndolos competir entre sí.

La integración de laboratorios de modelos fundacionales en el diseño de GPUs

Quizás el movimiento más disruptivo y revelador de la profundidad de esta tendencia es la decisión de OpenAI de diseñar sus propios chips de inferencia (denominado internamente "XPU") en colaboración con Broadcom y TSMC[80]. Históricamente, un laboratorio de modelos como OpenAI dependería completamente de proveedores de nube y chips. Sin embargo, la escala económica de sus operaciones ha cambiado la ecuación.

El acuerdo de OpenAI con Broadcom busca desarrollar un ASIC (Circuito Integrado de Aplicación Específica) personalizado para inferencia, fabricado en el nodo de 3nm de TSMC. Este proyecto implica para OpenAI abandonar los planes iniciales de construcción de una red manufactura propia para enfocarse en el diseño de GPUs y el aprovechamiento de la propiedad intelectual de Broadcom en interconexiones de alta velocidad[81]. La lógica económica es que, a medida que OpenAI escala hacia agentes autónomos que requieren un razonamiento continuo, el costo marginal por token se convierte en la métrica operativa fundamental. El objetivo estratégico es reducir el costo de inferencia para hacer económicamente viables los modelos de razonamiento avanzado (como las series o1/o2 de OpenAI) y eliminar el cuello de botella de suministro de GPUs. Las GPUs de propósito general de NVIDIA, aunque potentes, tienen una sobrecarga de área y energía para funciones que no son estrictamente necesarias para la inferencia de modelos GPT. De esta manera, un GPU diseñado por OpenAI y Broadcom puede eliminar toda la lógica superflua (como los núcleos de gráficos o de computación científica de doble precisión no utilizados) y maximizar el ancho de banda de memoria y la lógica de matriz sistólica específica para sus pesos de modelo. Se estima que, con base en este acuerdo, OpenAI podría desplegar chips con una capacidad total a 10 GW de capacidad para finales del 2026.

---

78 MSV J. (2025). Meta invests $14 billion in ScaleAI to strengthen model training. *Forbes* (June 23). Disponible en: https://www.forbes.com/sites/janakirammsv/2025/06/23/meta-invests-14-billion-in-scale-ai-to-strengthen-model-training/

79 Palazzolo, S. (2026). "The enigma of the data labeling startup explosion". *The Information* (April 14).

80 OpenAI (2025). *OpenAI and Broadcom announce strategic collaboration to deploy 10 gigawatts of OpenAI-designed AI accelerators* (October 13). Disponible en: https://openai.com/index/openai-and-broadcom-announce-strategic-collaboration/

81 Khosla, P. (2024). "OpenAI to build its first AI chip with Broadcom and TSMC", *AI Harware & Edge AI Summit* (October 30). Disponible en: https://www.kisacoresearch.com/blog/openai-build-its-first-ai-chip-broadcom-and-tsmc

Este acuerdo convierte a OpenAI efectivamente en una empresa de hardware *fabless*, integrándose verticalmente para asegurar su viabilidad económica a largo plazo, reduciendo su dependencia de las asignaciones de GPUs de Microsoft y NVIDIA. Para ello, OpenAI ha reservado capacidad en el proceso de fabricación de 3 nanómetros (3nm) de TSMC [82], lo que indica que sus chips competirán en rendimiento con los productos más avanzados de NVIDIA.

La integración de chatbots hacia el despliegue e implementación de aplicaciones

Stigler menciona en su trabajo que, en etapas avanzadas de desarrollo de una industria, firmas operando en eslabones de manufactura de un producto tienden a integrarse hacia adelante en la cadena productiva para controlar una reducción en el poder de marca del producto final, lo que podría conllevar una disminución en ingresos de la manufactura. El caso de integración hacia delante de los operadores de modelos fundacionales entrando en el eslabón de despliegue de aplicaciones representa una analogía al concepto teórico de Stigler. El ejemplo de integración vertical en este caso es el de OpenAI.

Al mismo tiempo que vende IA a empresas y gobierno, OpenAI está ofreciendo ingenieros y consultores para apoyar la adaptación a las necesidades de clientes, para lo cual ofrece servicios de *fine-tunning* por un valor aproximado de US$ 10 millones [83]. Por este costo, OpenAI accede a los datos del usuario para desarrollar aplicaciones adaptadas a sus necesidades. Este tipo de servicio posiciona a OpenAI como competidor directo a Accenture y Palantir, para lo cual está reclutando talento de *forward deployed engineers* [84].

Otro anuncio reciente de este tipo de integración es la asociación de Anthropic con Blackstone, Hellman & Friedman, dos firmas de capital de inversión y otros inversores para crear el "McKinsey de AI" [85].

La integración vertical de modelos fundacionales a modelos clásicos de búsqueda, publicidad y servicios adyacentes

Como se mencionó arriba respecto a la competencia entre chatbots, esta está generando una integración vertical hacia delante a los negocios de publicidad, y búsqueda. Como es obvio, Google ya lo ha realizado de hecho, dado que su entrada en el negocio de los chatbots está destinada a apalancar su negocio madre. En una iniciativa reciente, Google anunció que vincularía Gemini con Google Pay para completar transacciones de compra de productos [86].

El nuevo intento de integración vertical de desarrolladores de modelos fundacionales puede ser identificado en la modificación del plan de negocio original de OpenAI. En su nueva formulación, OpenAI menciona que generará US$ 112 mil millones en ingresos provenientes de la venta de publicidad y el cobro de comisiones por la venta de productos en línea. En el 2030, OpenAI estima que generaría ya US$ 46 mil millones, aunque dado que 90% de la base de usuarios de ChatGPT está basada al exterior de Norteamérica, y considerando que los ingresos por usuario de esta región son mucho más altos que en el resto del mundo, la concreción de este plan debe considerar como podría OpenAI crecer en el uso de su chatbot en EE.UU. y Canadá [87]. Al mismo tiempo, se debe considerar el obstáculo que enfrentará OpenAI a Google, Meta y Amazon que ocupan posiciones dominantes en publicidad digital, controlando aproximadamente 60% del mercado.

---

[82] Wheeler, K. (2025). "What's behind OpenAI's first custom chip design with TSMC," AI Magazine (February 10. Disponible en: https://aimagazine.com/articles/whats-behind-openais-first-custom-chip-design-with-tsmc

[83] Holmes, A. and Muppidi, S. (2025). "OpenAI takes a page from Palantir, doubles down on consulting services". *The Information* (June 25).

[84] Estos son definidos como ingenieros con experiencia en aplicaciones, perfiles de producto, despliegue e integración, capaces de adaptar modelos y herramientas a procesos reales de negocio. Ver Orosz, G. (2025). "What are Forward Deployed Engineers, and why are they so in demand" (August 12).

[85] Nicoll, A. and Alexander, R. (2026)." Wall Street's $1.5 billion plan to build the "McKinsey of AI". *Business Insider* (May 4).

[86] Gehan, A. (2026). "OpenAI's shopping ambitions hit messy reality". The Information (January 8).

[87] Muppidi, S. (2026). "OpenAI's International Conundrum", *The Information* (January 4).

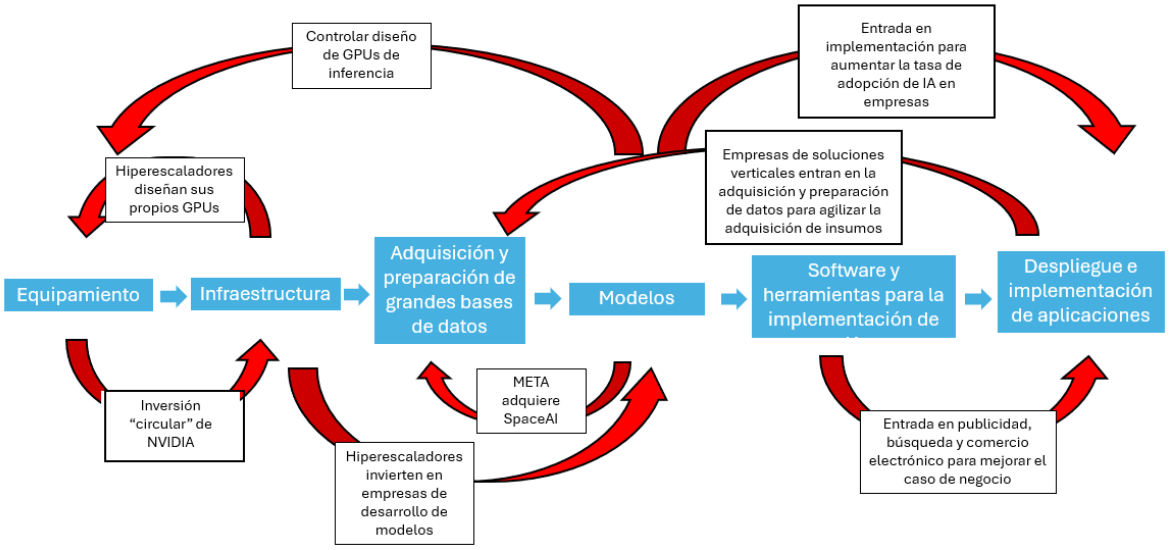
El segundo movimiento de integración vertical de OpenAI consiste en diversificarse hacia la venta digital de productos digitales. En este campo, OpenAI ha introducido un servicio denominado *Instant Checkout* que vincula a usuarios de ChatGPT con firmas vendedoras para el pago de productos. Asimismo, OpenAI ya ha firmado alianzas estratégicas con Walmart, Shopify, Zillow y DoorDash, aunque el apalancamiento de IA para entrar en el negocio de ventas digitales plantea obstáculos relacionados de cuál sería la reacción de usuarios de IA en relación con la integración de búsqueda información y la promoción de productos y servicios [88].

La tercera iniciativa de integración vertical se refiere a la entrada de chatbots en el negocio de salud. Anthropic ha anunciado el despliegue de *Claude for Healthcare,* un conjunto de herramientas y herramientas a ser usados por proveedores de servicio, empresas de seguros y consumidores de servicios de salud. Esta iniciativa es similar a la formulada por OpenAI para la oferta de servicios de salud enfocados en consumidores individuales [89]. Este movimiento fue sustanciado con la compra de Torch, una app basada en IA que permite a usuarios individuales analizar información sobre su salud y monitorear el desempeño de actividades físicas [90].

Los movimientos simultáneos de integración vertical a lo largo de la cadena de valor de IA

El desarrollo de todos los movimientos de integración vertical hacia atrás y adelante en la cadena de valor analizados en los apartados precedentes permiten visualizar la intensidad de maniobras llevadas adelante por los participantes del ecosistema de IA (ver diagrama 14).

Diagrama 14
Movimientos de integración vertical simultáneos en la cadena de valor de IA



Fuente: Elaboración propia.

Si bien es importante mencionar que no todos estos movimientos estratégicos van a resultar en éxitos, el propósito es demostrar como las cadenas de valor originales del ecosistema digital han devenido obsoletas en el contexto de la IA.

---

88 Muppidi, S. and Palazzolo, S. (2025). "OpenAI's ads push starts taking shape", *The Information* (December 24).
89 Peers, M. and Osawa, J. (2026). "Anthropic unveils New healthcare features for Claude", *The Information* (January 12).
90 Mandaro, L. (2026). "OpenAI agrees to buy AI healthcare App for about US$ 100 million." *The Information* (January 12).

3. Los puntos de presión en la cadena de valor de inteligencia artificial

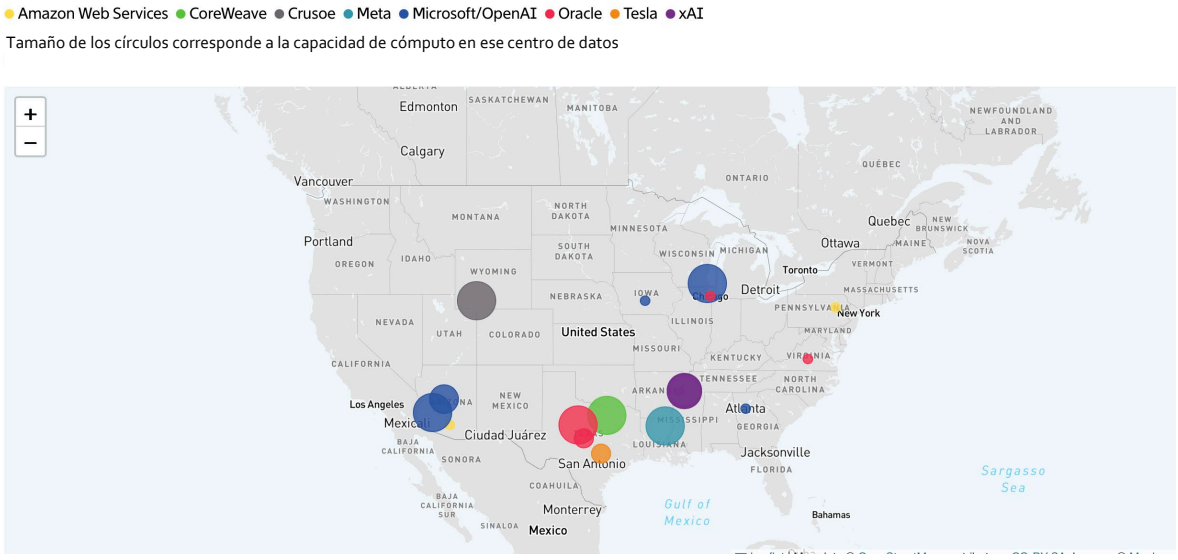
Los movimientos de integración vertical y planes de inversión en la cadena de valor de IA plantean una serie de interrogantes sobre su viabilidad económica y estratégica o inestabilidad potencial. Sin desconocer que estos cuestionamientos pueden llegar a ser resueltos de manera positiva para el ecosistema, el propósito de la siguiente sección es resumir una serie de interrogantes planteados relacionados con la robustez económica y financiera de la cadena.

Sobreinversión en centros de datos?

La integración vertical de OpenAI, donde esta combina su inversión con Softbank y Oracle para el despliegue de centros de datos masivos en Stargate, la inversión sostenida de hiperescaladores y el despliegue de neoclouds, así como los compromisos de inversión circular de NVIDIA han generado al momento un salto en la inversión actual y futura en el despliegue de infraestructura. La inversión masiva en despliegue de centros de datos está determinada por el supuesto estratégico de que, en la medida de que estos apalancan un acceso preferencial y prioritario a recursos escasos (por ejemplo, GPUs) resultan en una ventaja competitiva sustancial.

Solamente en Estados Unidos, el despliegue de centros de datos para el procesamiento de IA que han sido completados, o están en proceso de construcción al momento alcanzan cuatro para AWS, cuatro para Meta, cinco para Microsoft/OpenAI, cuatro para Oracle, uno para CoreWeave, uno para OpenAI autónomo, dos para xAI, uno para Crusoe y uno para TeraWulf, por un total de veintitrés, sumando para los que se dispone de información pública 5,233,000 GPUs, la mayoría de ellos perteneciente al modelo H100 de NVIDIA.

Mapa 1
Estados Unidos: mapa de centros de datos desplegados o por desplegar para procesamiento de IA

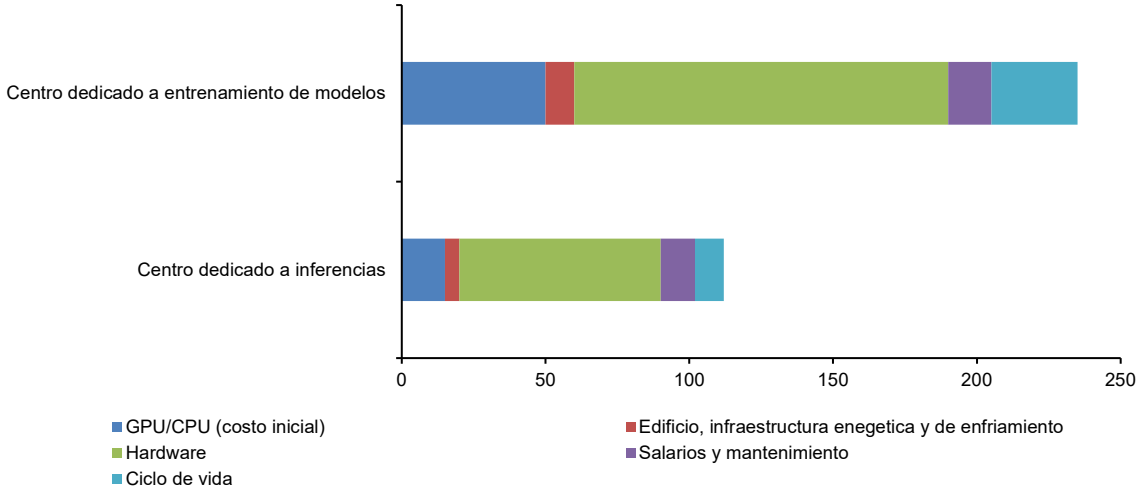


Fuente: The Information's AI Data Center database.
Nota: Los límites y los nombres que figuran en este mapa no implican su apoyo o aceptación oficial por las Naciones Unidas.

Esto representa la inversión en infraestructura de centros de datos más agresiva registrada históricamente. Parte de la inversión en despliegue de centros de datos está potenciada por la "circularidad" de acuerdos mencionada arriba, donde, por ejemplo, NVIDIA invierte US$ 100 mil millones en OpenAI para que este despliegue centros de datos que requerirán, a su vez, la compra de GPUs a ser provistos por NVIDIA. Así, NVIDIA se transforma en inversor y vendedor de tecnología al mismo tiempo.

El volumen de inversión de capital incrementa el perfil de riesgo del despliegue de data centers, lo que incluye la escala, el tiempo asignado a la construcción, y la sofisticación técnica, así como el número de proveedores de insumos. El costo de inversión de un centro de datos de 1 GW de potencia supera los US$ 100 mil millones (ver gráfico 2).

Gráfico 2
Costo promedio de construcción y operación de un centro de cómputo por un ciclo de operaciones de 20 años



Fuente: Standard & Poor' Global ratings.

El interrogante planteado a partir de este despliegue de capacidad de cómputo es en términos del riesgo que existe para generar una tasa de retorno adecuada para este monto de inversión:

i) ¿Cuál es la utilización de capacidad existente y futura en términos de entrenamiento y entrega de servicio (es decir, inferencia) de modelos fundacionales y aplicaciones? Dos indicadores por monitorear respecto de la sobrecapacidad de centros de datos son la tasa de ocupación de capacidad instalada, y los incentivos a ser ofrecidos por operadores para la venta de capacidad (en otras palabras, cuanto más incentivos, más capacidad ociosa). Entre la evidencia a considerar en este terreno es la disrupción ocasionada ya en Fermi, una startup con planes de despliegue de centros en Amarillo, Texas. La empresa salió al mercado con una valoración de US$ 19 mil millones, pero sin ingresos comprometidos. A fines de abril, la empresa no pudo contratar un solo cliente, con lo cual el valor de la acción cayó un 84% y el CEO y fundador fue despedido [91].

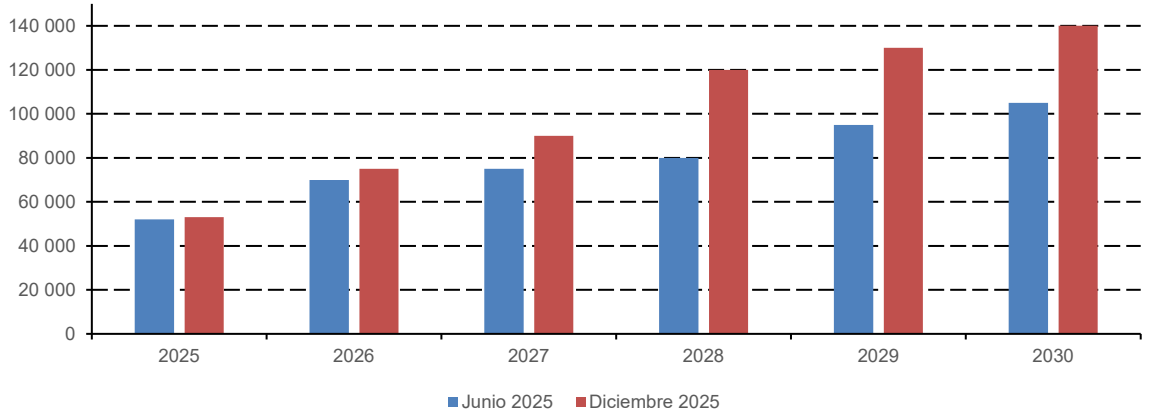
ii) ¿Cuál será la evolución tecnológica que podría disrumpir las arquitecturas actuales? Cambios en el diseño y estructura de modelos podrían indicar la existencia de sobre capacidad en el despliegue de infraestructura. Por ejemplo, la evolución a modelos fundacionales basados en aprendizaje continuo (*continual learning*) donde estos dependen menos en entrenamiento requiriendo uso masivo de datos, la demanda de capacidad de cómputo estaría generada principalmente por inferencias, que requieren menor uso de recursos.

iii) Adicionalmente, considerando que el ciclo de vida de un GPU es extremadamente corto, esto implica rápidos procesos de reemplazo, con lo cual la inversión en centros de datos debe mantenerse relativamente constante y va más allá del mantenimiento normal.

[91] Ma, M. and Malik, N. (2026). "Data center startup Fermi ousts CEO after struggling to sign up clients". *Bloomberg*, May 4.

Generado ya por la construcción excesiva de centros de datos, se está creando un retardo en la conclusión de proyectos, con lo cual se demora la posibilidad de generar ingresos. Al perfil de riesgo económico se suma la complejidad en ejecución. De acuerdo con Jones Lang LaSalle, un grupo inmobiliario, 57% de los proyectos de construcción de centros de datos en el 2025 han experimentado demoras en su finalización de tres meses o más. A esto se suma, el incremento proyectado en el consumo energético (ver gráfico 3).

Gráfico 3
Estados Unidos: consumo energético de centros de datos (2025-2030)
(En megavatios)



Fuente: Standard & Poor' Global ratings.

La demora también se produce no solamente en la construcción del sitio sino también en el desarrollo de las fuentes de energía necesarias para operarlos [92]. Esto no considera las demoras regulatorias provocadas por la preocupación en el impacto en la grilla eléctrica [93].

¿Considerando el control que NVIDIA y TSMC ejercen en el segmento de GPUs de última generación, qué puede ocurrir si estos jugadores no pueden satisfacer la demanda en los plazos determinados por el despliegue de centros? De hecho, investigación de EpochAI indica que uno de los cuellos de botella más importantes en la capacidad de escalamiento de IA es la producción de GPUs [94].

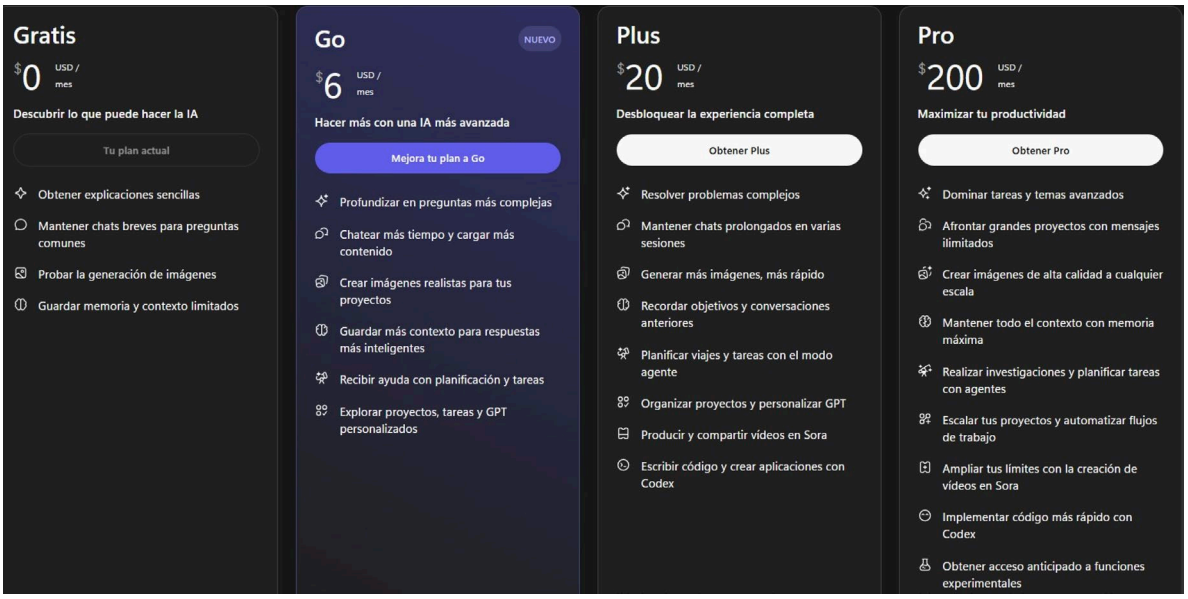
La demanda acelerada de insumos ocasionada por la construcción de centros de datos ha generado la irrupción de proveedores enfocados en la venta de sistemas de refrigeración especializados, equipos de comunicación de alta velocidad, y de gestión de fuentes energéticas. Si bien estas empresas están creciendo de manera acelerada, su foco en la provisión de este tipo de insumos crea riesgos de concentración si la demanda no se materializa de acuerdo con las proyecciones actuales.

¿Qué ocurre si la entrega de servicios adicionales ocasionando procesos de inferencia elevados no puede generar los ingresos necesarios para rentabilizar la inversión? Contrariamente al software tradicional (Microsoft Windows, por ejemplo) donde el costo marginal tiende a cero, la entrega de servicios de IA requiere recursos computacionales constantes. Por lo tanto, las economías de escala tienden a estabilizarse a un cierto punto del volumen de usuarios. En este contexto, el proveedor de

[92] Hasta septiembre del 2025, los centros planeados consumirán 25.3 Giga watts, de acuerdo con DatacenterHawh, un aumento de 38% desde finales de 2024.
[93] Vaughan, Ann (2025). "AI Infrastructure", *The Information*.
[94] EPOCHAI (2025). *Can AI scaling continue through 2030?*

servicios de IA debe cobrar un sobre precio premium por ciertos servicios (ver diagrama 15 con estructura de precios de ChatGPT). Por otra parte, Google cobra US$ 249 mensuales por acceso a Gemini bajo el plan Ultra de GoogleOne, mientras que Grok cobra US$ 300 por el modelo SuperGrok.

Diagrama 15
Precios de ChatGPT (2026)



Fuente: Sitio de ChatGPT.

¿Qué ocurre si la voluntad de pago es inferior al precio demandado? La respuesta a muchos de estos interrogantes está condicionada por la evolución de la demanda futura de servicios de IA y la necesidad de recursos de computación desplegados.

¿Viabilidad de "pure plays"?

La cadena de valor de IA incluye ciertos jugadores dominantes que pueden ser categorizados como "pure play", o sea empresas enfocadas exclusivamente en IA. Ejemplos de ello son OpenAI y Anthropic, mientras que los jugadores en IA que apalancan posiciones en otros sectores incluyen a Google (publicidad y búsqueda), Amazon (computación en la nube, ventas digitales y publicidad, entre otros), y Meta (redes sociales y publicidad). La diferencia entre el primer grupo de jugadores y el segundo es el método de financiamiento en el desarrollo de negocios basado en IA (ver el volumen de ventas y CAPEX del segundo sector de jugadores).

Cuadro 7
Ventas y CAPEX de jugadores provenientes de otras cadenas adyacentes (2025)
(En miles de millones de dólares)

Empresa	Ventas	EBITDA	CAPEX
Alphabet (Google)	385,47	127,7 (2024)	93
Amazon	700,8	157,00	125
Meta	192,46	102,76	72
Microsoft	281,7	162,7	88,2

Fuente: Elaboración propia sobre la base de informes anuales.

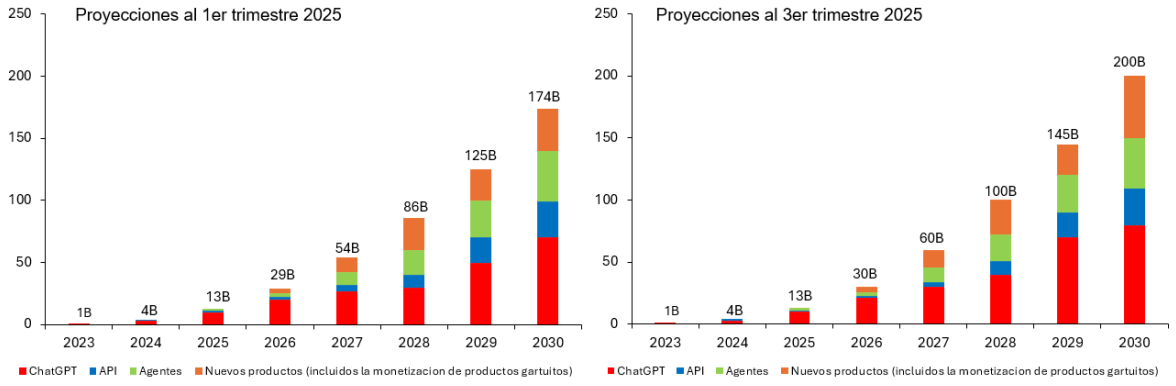
Los datos del cuadro 7 indican la capacidad de apalancamiento financiero que poseen estos cuatro jugadores para financiar sus emprendimientos en IA que, por otra parte, son clave en su consolidación en cadenas de valor adyacentes.

Ahora bien, en el caso de los "pure plays" la situación plantea ciertos interrogantes. Como ejemplo, de acuerdo con la última actualización de su proyección financiera, OpenAI estima un gasto de efectivo (cash burn) de US$ 115 mil millones hasta el 2029, lo que incluye US$ 35 mil millones en 2027 y US$ 45 mil millones en 2028 respectivamente (en total US$ 80 mil millones más altos que en la proyección del plan de negocio previamente formulada por la empresa). Uno de los factores clave en esta proyección es el alquiler de servidores de la nube para entrenar y apoyar el funcionamiento de sus modelos fundacionales. Para controlar el gasto, OpenAI está desplegando esfuerzos para desarrollar sus GPUs propios y construir centros de datos bajo su control, aunque todavía estos esfuerzos están lejos de generar frutos. Al mismo tiempo, OpenAI está enfrentando la necesidad de retener talento en respuesta al ataque de competidores, con lo cual proyecta la necesidad de reservar US$ 20 mil millones en compensación por acciones de retención hasta el 2030.

Esto explica la urgencia de la empresa para continuar su búsqueda de inversión de capital, aunque por el momento inversionistas siguen interesados en continuar financiado a OpenAI con la adquisición de acciones en el mercado privado, lo que estaría generando una capitalización de mercado de US$ 500 mil millones, o sea 20% del valor total de Google. Es en ese marco que dos de los "pure plays" más importantes están considerando salir al mercado próximamente [95].

En ese contexto, la proyección de gastos mencionada arriba requiere incrementar los ingresos lo que determina la entrada de "pure plays" en negocios adyacentes a la cadena de valor de IA. A partir de ello, la proyección de ingresos de OpenAI actualizada en los últimos seis meses demuestra un aumento en 2025 de US$ 174 mil millones a US$ 200 mil millones (ver gráfico 4).

Gráfico 4
OpenAI: proyección de ingresos (1T2025 vs. 3T2025)
(En dólares americanos)



Fuente: Muppidi, S. (2025). OpenAI says its business Will Brun $155 billion through 2029. The Information (September 5).

Estas proyecciones de ingresos plantean el interrogante de si OpenAI será capaz de monetizar servicios en cadenas adyacentes como publicidad y ventas, o el cobro de servicios para los usuarios de ChatGPT que acceden al chatbot de manera gratuita. Por otro lado, las proyecciones de ingresos están asumiendo que OpenAI estaría generando ingresos por abonado y márgenes similares a los de

95 Pau, V. (2026). "Nine tech IPO contenders to watch this year." *The Information* (January 16).

Facebook. Esta estrategia asume que OpenAI podría entrar en ese segmento sin desencadenar una respuesta por parte de Google, lo que no es viable. De hecho, la pretendida entrada de OpenAI en el segmento de e-commerce desencadena la respuesta de no sólo Google sino también de Amazon.

Este interrogante también puede ser extendido a los "neoclouds". Como se verá en el capítulo siguiente de dimensionamiento económico, CoreWeave, una neocloud está proyectando invertir US$ 35 mil millones en capital contra ingresos de US$ 12 mil millones. Este ratio 3/1 está generando ya presión en el mercado de valores, a lo que se suma una deuda de US$ 24.8 mil millones y tan solo US$ 3 mil millones en efectivo, con lo cual la acción ha disminuido en aproximadamente 9%[96].

El volumen de deuda en participantes de la cadena de valor

Un estudio de Morgan Stanley estima que entre el 2025 y el 2028, la construcción de infraestructura de IA (esencialmente centros de datos) requerirá a nivel mundial US$ 2.9 trillones, de los cuales $ 1,400 mil millones van a ser financiado por CAPEX de hiperescaladores, US$ 350 mil millones por fondos de private equity, capital de riesgo y fondos soberanos, US$ 800 millones en crédito privado, US$ 150 mil millones en valores de colateralizados (*Asset-based securities*), y préstamos colateralizados (*Collateralized Mortgage Based Securities*) [97], y US$ 200 millones en deuda corporativa. Estas cifras no consideran el gasto en energía necesaria para apoyar los centros de datos.

La construcción de centros de datos y la inversión concomitante en GPUs requiere fondos que están siendo parcialmente accedidos mediante instrumentos de deuda. Algunos participantes de la cadena, como Google, no han necesitado, hasta el momento, recurrir a deuda para financiar la inversión. Otros lo han hecho. Por ejemplo, Meta ha negociado un paquete de 27 mil millones de dólares en deuda para financiar su centro de dato en Luisiana. En otro caso, CoreWeave ha vendido deuda por US$ 13 mil millones para financiar su adquisición de GPUs en centros de datos. Los intereses en el 2025 solo para CoreWeave suman US$ 1.25 mil millones[98].

En el curso del 2025, las empresas que estaban financiando la construcción de centros de datos se han comprometido en deuda de hasta US$ 100 mil millones[99]. La mayor parte de este monto proviene de operadores como AWS quien ha obtenido US$ 15 mil millones. Sin embargo, existe también deuda contraída por empresas más pequeñas. Por ejemplo, Wulf Compute, una subsidiaria de Terawulf contrajo deuda por US$ 3.2 mil millones en bonos a una tasa de 7.75%, comparada con 5.5% en casos similares. La misma tasa ha sido comprometida por Cipher Compute por US$ 1.7 mil millones. Applied Digital, el locatario principal de CoreWeave, contrajo deuda por US$ 2.35 mil millones con una calificación crediticia de BB.

Oracle ha decidido recurrir a bonos de deuda para adquirir GPUs de NVIDIA para operar los centros ofrecidos a OpenAI [100]. Los fondos generados por la venta de bonos por USD$ 18 mil millones serán usados no solo para adquirir equipamiento sino también para la adquisición de terrenos, construcción de edificios, y equipamiento especializado [101]. Esto permite a Oracle evitar la dilución de capital accionario o usar fondos propios y mantener flexibilidad operativa.

---

96 Peers, M. (2026) "CoreWeave and Risk", *The Information* (May 7).

97 Tirupattur, V. (2025). "How credit markets could finance AI's trillion-dollar gap". *Morgan Stanley Thoughts on the Market* (August 6). Disponible en: https://www.morganstanley.com/insights/podcasts/thoughts-on-the-market/credit-markets-ai-financing-gap-vishy-tirupattur-vishwas-patkar

98 Shi, M (2025). "Data center pile in shows little concern about exits" *PitchBook* (November 21).

99 Rennison, J. (2025). "As AI companies borrow billions, debt investors grow wary" *New York Times* (December 26).

100 Kruppa, M. (2025). 2026 "Predictions: Cash-strapped Oracle issues chip-backed Debt". *The Information* (December 31).

101 El bono fue sobre vendido por un valor de US$ 50 mil millones, indicando el interes de inversionistas.

Más allá del volumen de deuda, las tasas de interés demuestran cierta preocupación por parte de los proveedores de capital. Por ejemplo, Applied Digital, un constructor de centros de datos debe pagar una tasa de 3.75 puntos porcentuales más altos que empresas de la misma categoría.

Estas preocupaciones se potencian con la probabilidad de "exit" por parte de inversores institucionales que han adquirido este tipo de bonos. Sin una posibilidad de salida, la deuda podría transformarse en pérdidas para inversionistas. Una encuesta de Alix Partners entre 400 empresas de private equity, fondos de inversión y firmas relacionadas con centros de datos, indica que aquellas instituciones que han adquirido deuda contemplan una venta entre 5 y 10 años. Los horizontes de salida son más complejos en la medida de que existen dudas sobre la tasa de retorno de la deuda adquirida. Esto podría anunciar un escenario posible de consolidación [102].

Problemas en la adopción de IA en organizaciones

Información cuantitativa y anecdótica indica que la adopción de IA en organizaciones todavía presenta limitaciones. Challapally et al. (2025) reportan en una investigación realizada en organizaciones de EE. UU. que 95% de las encuestadas reportan un retorno nulo en la inversión en IA. Si bien ChatGPT y Copilot han sido adoptadas masivamente por usuarios individuales en empresas, solamente 5% de 153 empresas encuestadas reportan que la adopción de IA ya está generando un retorno económico. [103] Corresponde mencionar, sin embargo, que el estudio presenta algunas limitaciones. Por ejemplo, está basado en entrevistas con 52 empresas y 153 encuestas sin especificar la dimensión de la empresa que responde a la encuesta o su distribución sectorial. De todas maneras, esta evidencia es parcialmente coincidente con resultados de otra encuesta en EE. UU. y Europa, donde se reporta que 42% de empresas encuestadas han abandonado iniciativas de implementación de IA, y 46% cancelan proyectos antes de alcanzar la "prueba de concepto" [104].

Una investigación con más rigurosidad estadística es la realizada por el Bureau del Censo de Estados Unidos en 2022 sobre una muestra estratificada de 300,000 empresas con una tasa de respuesta de 63.3% [105]. En ella, 3.3% de las empresas censadas reportan adopción de IA, alcanzando 12.8% en el sector de TIC (ver gráfico 5).

A pesar de que esta última encuesta fue realizada en el 2022, los resultados son parcialmente coincidentes con los de las investigaciones anteriormente citadas.

Sin embargo, corresponde mencionar que la evidencia respecto a la adopción en empresas no es totalmente consistente. Otras investigaciones muestran resultados parcialmente contradictorios con los mencionados arriba. Por ejemplo, McKinsey realizó un estudio sobre el uso de IA generativa basado en encuestas a 1,993 participantes en 105 países. Los resultados concluyen que, si bien el uso de IA continúa creciendo a nivel global, tan solo 7% de las organizaciones reportan uso generalizado de la tecnología, mientras que 62% todavía están en un estadio de experimentación y proyectos piloto, y 31% están en una etapa de escalamiento. En particular, el escalamiento de agentes basados en modelos fundacionales no excede el 10%, mientras que 30% de los encuestados reportan problemas por efectos negativos en recomendaciones generadas por IA. En América del Norte, en particular, McKinsey reporta que en el 2024, el 80% de las organizaciones reportan el uso de IA. En parte, la diferencia entre los resultados de McKinsey y los del Censo de ABS pueden ser explicados por la diferencia temporal, dado que los valores de este último fueron generados en el 2022. Otra razón puede ser deducida en términos de que la investigación de McKinsey no establece diferencia entre adopción individual (es decir un

---

[102] Ver (2025). All chips in! Would a fall in AI-related asset valuations have financial stability consequence? Bank of England (October 31).

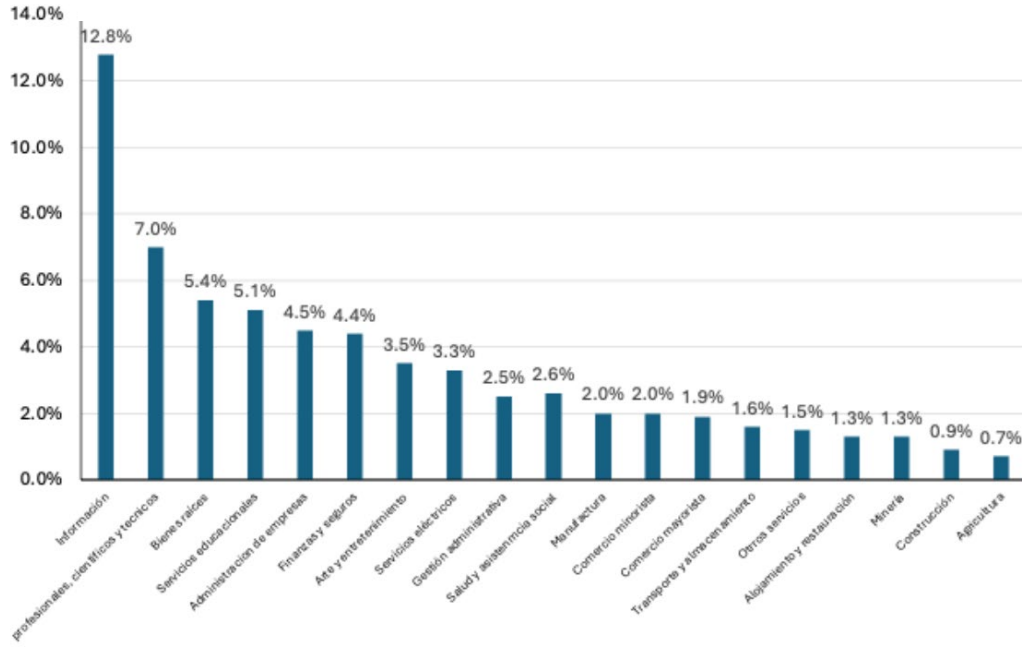
[103] Challapally, A., Pease, C., Raskar, R., and Chari, P. (2025). The state of AI in business 2025. MIT Project Nanda.

[104] (2015). "Generative AI experiences rapid adoption, but with mixed outcomes – Highlights from Vote: AI & machine learning", S&P Global (May 30). Disponible en: https://www.spglobal.com/market-intelligence/en/news-insights/research/ai-experiences-rapid-adoption-but-with-mixed-outcomes-highlights-from-vote-ai-machine-learning

[105] Census Bureau (2023). Annual Business Survey 2022.

usuario individual de la empresa utilizando Copilot) y el uso de IA para eficientizar un proceso de negocio. De hecho, un estudio realizado por la Universidad de Pennsylvania [106] reportando una encuesta realizada entre 800 tomadores de decisiones en empresas de más de 1,000 empleados y más de US$ 50 millones en ingresos de Estados Unidos indica que la adopción de IA se registra en tareas como análisis de datos (73%), resumen de reuniones (705), y edición de documentos (68%), y en menor adopción en programación y creación de informes, todas tareas individuales.

Gráfico 5
Estados Unidos: adopción de IA por empresas, 2022
(En porcentajes)



Fuente: Census Bureau. Annual Business Survey (ABS).

Por otra parte, corresponde mencionar que, en términos cualitativos, Microsoft reporta que el uso de 365 Copilot no es tan intenso como lo esperado y que muchos usuarios prefieren acceder a ChatGPT o Gemini de manera gratuita. Por ello, Microsoft está financiado el entrenamiento de usuarios de Copilot para mejorar el nivel de adopción [107].

En resumen, aun considerando que la información de adopción en empresas no es consistente, corresponde plantear la pregunta de si una demora en asimilación de la tecnología en sectores productivos puede resultar en un punto de presión adicional en la cadena de valor de la IA.

Conclusión sobre los puntos de presión

Los cuatro puntos de presión descritos arriba deben ser considerados en el marco de la sostenibilidad de largo plazo de la cadena de valor. Sin dejar de plantear el valor económico de la IA en términos de su impacto en empresas y el sector público, la pregunta que corresponde hacerse es sobre la viabilidad y sostenibilidad de la cadena de valor en su configuración actual en el largo plazo, o si estas cuestiones anticipan una consolidación.

---

[106] Korst, J., Puntoni, S. and Tambe, P. (2025). Accountable acceleration: Gen AI Fast track into the enterprise (October).
[107] Jacobson, A. (2025). "Microsoft, OpenAI admit to core problems internally and with users", The Information.

B. Dimensionamiento económico de la cadena de valor de la inteligencia artificial en Estados Unidos

En el capítulo I de este estudio enfocado en la organización industrial que estudia las cadenas de valor se plantean cuatro conceptos clave que explican su evolución:

- El concepto de cadena de valor fue introducido para entender el funcionamiento de industrias, como la IA, en términos de las dinámicas de integración y fragmentación vertical a lo largo de su ciclo de vida (Stigler, 1951).
- Esta dinámica permite entender modelos de negocio, dinámicas competitivas, alianzas estratégicas y modelos de integración vertical y horizontal.
- Adicionalmente, de acuerdo con el concepto de costos de transacción, empresas operando en eslabones específicos de cadenas de valor tienden a integrarse verticalmente cuando el mecanismo de mercado se vuelve ineficiente debido a la alta especificidad de los activos y la incertidumbre contractual (Coase, 1937).
- Complementariamente, empresas tienden a integrarse verticalmente buscando el control estratégico de ciertos estadios de la cadena (por ejemplo, la relación con el usuario final), o buscando reducir los costos de transacción entre estadios de la cadena (Williamson, 1985).

El estudio de dimensionamiento económico de actores de la cadena de valor de la IA de esta sección se enfoca en identificar las empresas que ocupan eslabones específicos de la cadena y medir sus ingresos, costos operativos, márgenes de rentabilidad, inversión, y capitalización de mercado. Al integrar la medición de estas variables en cada eslabón de la cadena se pueden identificar cuáles son los centros de control estratégico. Una vez identificados dichos centros, es posible anticipar movimientos de integración vertical, a partir de los cuales los jugadores ubicados en dichos eslabones pueden expandir sus operaciones en otros como estrategia defensiva. Otros movimientos de integración vertical permiten entender como jugadores ubicados en eslabones secundarios donde no pueden apalancar posiciones dominantes tienden a integrarse verticalmente buscando los centros de control estratégico de la cadena. En este contexto, el dimensionamiento económico permite cuantificar el valor estratégico de cada eslabón y anticipar movimientos futuros.

Habiendo dicho esto, el análisis de dimensionamiento económico enfrenta una complejidad analítica. Si bien existen jugadores que ocupan una posición única en uno de los eslabones de la cadena (los denominados "pure play"), existen otros cuya integración vertical y la opacidad en instrumentos de medición no permiten visualizar, por ejemplo, ingresos y rentabilidad por eslabón para estos. El caso paradigmático de esta complejidad es Google cuya presencia en desarrollo de GPUs, infraestructura de centros de datos, modelos fundacionales, y herramientas de IA no permite entender las dinámicas económicas específicas por eslabón para este jugador. En este contexto, reconociendo esta complejidad, optamos por mencionar aquellos jugadores, ubicar su posicionamiento en los correspondientes eslabones y resaltar la complejidad de medir las dimensiones económicas por eslabón para ellos.

Una pegunta típicamente hecha en el marco del dimensionamiento de una cadena de valor es "¿Dónde se crea valor?" En realidad, la respuesta es compleja dado que el valor económico puede ser medido con base en más de una métrica (por ejemplo, retorno a inversores o en posicionamiento estratégico). Dado que la respuesta es multidimensional, el siguiente análisis se enfoca en cinco variables económicas clave:

i) **Ingresos:** mide el flujo que percibe una empresa por la comercialización de bienes, licencias o la prestación de servicios vinculados a su eslabón específico dentro de la cadena de valor de la IA, antes de deducir cualquier costo.

ii) **Costos operativos:** mide los gastos recurrentes para mantener el funcionamiento continuo de las operaciones en cada eslabón de la cadena de valor de la IA. Esta métrica captura los flujos de caja constantes destinados a rubros críticos como la investigación y desarrollo (I&D), el consumo energético masivo, el arrendamiento y mantenimiento de infraestructura en la nube, los costos algorítmicos (entrenamiento e inferencia), y la captación y retención de talento humano especializado. En start-ups esta métrica está asociada al "cash burn". [108]

iii) **Rentabilidad:** refleja la capacidad de las empresas en cada eslabón para generar valor económico real una vez deducidos los costos y gastos asociados a su operación. En este análisis, la rentabilidad es medida a través del margen EBITDA (*Earnings Before Income Taxes, Depreciation and Amortization*), el cual permite ilustrar la asimetría estructural de la cadena de valor: mientras ciertos eslabones críticos respaldados por la exclusividad tecnológica y el poder de fijación de precios gozan de rentas elevadas, otros enfrentan una fuerte compresión de márgenes, déficits estructurales o "cash burn" justificada por la fase de expansión y los altos costos operativos.

iv) **Inversión de capital:** mide los fondos que requiere una empresa en términos de propiedades, edificios, maquinaria y equipamiento para entrar en un mercado, operar y crecer**.** Su medición anual calcula la diferencia entre la inversión al final del ciclo contable menos la inicial a los que se suma la amortización. La inversión de capital provee típicamente una perspectiva sobre las barreras de entrada. El análisis de la inversión de capital en la cadena de valor de la IA permite identificar la magnitud de los recursos destinados a activos físicos e infraestructura crítica necesarios para viabilizar cada eslabón, y que pueden actuar como barreras de entrada para nuevos actores.

v) **Capitalización de mercado**: métrica financiera calculada en función del valor de la acción y el número de acciones en el mercado. Si bien esta asume que la empresa cotiza en bolsa, el concepto también es usado para medir la inversión en función de la participación accionaria en operaciones privadas. En última instancia, esta métrica refleja la perspectiva de inversores sobre el valor de una empresa.

Este capítulo presenta el análisis de dimensionamiento económico de la cadena para Estados Unidos. Todos los valores presentados corresponden al año 2025. El estudio de métricas es presentado de manera gráfica y en resúmenes en las secciones siguientes, mientras que las fuentes y cuantificación detallada es incluida en apéndices. Si bien el análisis se enfoca en empresas cuya casa matriz se encuentra en Estados Unidos, en muchos casos su peso económico incluye una cuantificación de variables generadas en el exterior de este país debido a su expansión multinacional o porque funcionan en cadenas productivas globales. Por ejemplo, los ingresos de NVIDIA incluyen no solo aquellos originados en Estados Unidos sino los provenientes de clientes en el exterior.

[108] "Cash burn", término usado principalmente en startups, define la velocidad con la cual una empresa usa sus reservas de efectivo para cubrir gastos antes de generar flujos de caja positivos. Esto representa flujos negativos, y se reporta mensualmente o por trimestre para determinar por cuanto tiempo puede una empresa emergente continuar operando antes de necesitar financiamiento adicional.

1. Ingresos

La distribución de los ingresos en la cadena de valor de la IA revela una asimetría estructural, donde la magnitud de facturación depende fuertemente de la exclusividad tecnológica, y control de mercados. En este entorno, coexisten ingresos masivos concentrados en nodos críticos frente a otros eslabones de la cadena que presentan ingresos comparativamente inferiores.

- ***Eslabón 1, Equipamiento*:** incluye: i) el diseño de GPUs (empresas como NVIDIA y AMD), ii) la fabricación de GPUs (donde TSMC controla 92% del mercado mundial), y iii) los ensambladores de servidores (como Dell Technologies). En el área de diseño de GPUs, NVIDIA alcanzó ingresos de US$ 130,497 millones, mientras que AMD registró US$ 34,639 millones. En cuanto a la fabricación de GPUs, TSMC reportó ingresos por US$ 122,424 millones. Por otra parte, en el segmento de ensambladores de servidores, Foxconn (ODM) lideró con US$ 215,000 millones (correspondientes a su año fiscal 2024), Dell Technologies reportó US$ 95,567 millones y Supermicro alcanzó los US$ 21,972 millones[109]. Es importante mencionar que en este último segmento, los ingresos representan valores totales lo que incluyen componentes no relacionados con la IA, como teléfonos móviles para Foxconn y PCs para Dell.
- ***Eslabón 2, Infraestructura de IA:*** en el eslabón de infraestructura computacional, las empresas proveedoras de servicios de nube y centros de datos registraron volúmenes de facturación elevados en relación con el resto de las etapas de la cadena de valor[110]. AWS lideró este segmento con ingresos de US$ 128,700 millones, mientras que Microsoft Azure reportó US$ 106,265 millones en su segmento Intelligent Cloud. Google Cloud, por su parte, alcanzó ingresos de US$ 58,705 millones. Oracle reportó ingresos de US$ 57,399 millones totales —de los cuales US$ 49,230 millones correspondieron a infraestructura de nube—. Finalmente, la "neo-cloud" CoreWeave registró un destacado crecimiento interanual del 168%, alcanzando ingresos por US$ 5,131 millones[111].
- ***Eslabón 3, Preparación y gestión de datos***: en la preparación y gestión de datos, Scale AI proyectó ingresos por US$ 2,000 millones para el año fiscal 2025. Handshake, operando en el servicio de etiquetado por contratistas, ha alcanzado ingresos anualizados de US$ 1,000 millones[112]. En el segmento de bases de datos vectoriales, Pinecone registró ingresos de US$ 26.6 millones en 2024, destacando un crecimiento del 66.6%. Por su parte, Gretel.ai, antes de su venta a NVIDIA, reportó ingresos por US$ 14.6 millones[113].
- ***Eslabón 4, Modelos fundacionales:*** en los laboratorios de modelos fundacionales, los niveles de facturación reflejaron las distintas estrategias de monetización y adopción del mercado. OpenAI alcanzó una tasa de ingresos anualizados de US$ 25,000 millones a fines de febrero de 2026, tras haber superado previamente los US$ 20,000 millones durante el año 2025. Anthropic reportó ingresos por US$ 9,000 millones en el mismo año. Por su parte, los ingresos específicos de Google DeepMind no se encuentran disponibles, ya que la empresa no contabiliza de modo discriminado las ventas de su modelo Gemini. En el caso de xAI, se registraron ingresos por US$ 107 millones en el tercer trimestre de 2025, lo que equivale a un monto anualizado de US$ 428 millones. Finalmente, Meta no generó ingresos directos con sus modelos Llama al tratarse de un modelo de código abierto,

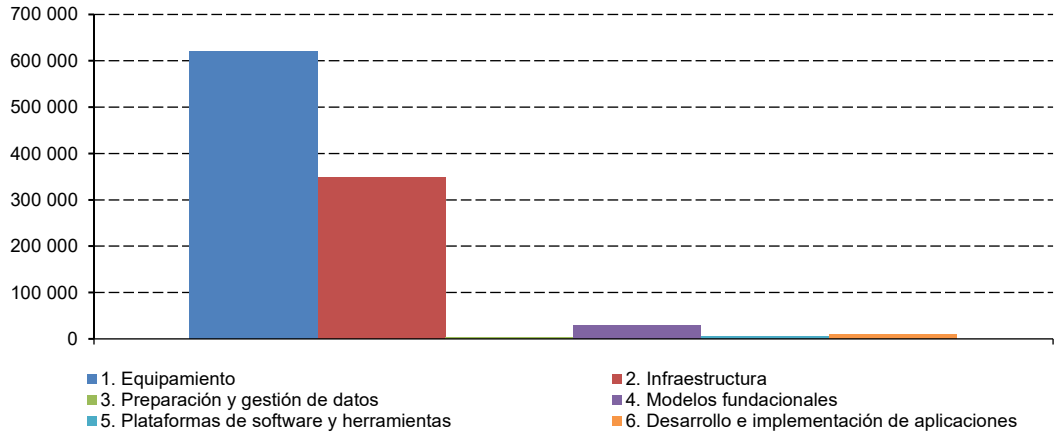
[109] Fuentes detalladas en anexo 1.
[110] Los valores de ingresos corresponden a las ventas globales.
[111] Fuentes detalladas en anexo 1.
[112] Hornstein, J. and Weinberg, C. (2026). "Handshake, Mercor revenue surges on demand for human contractors to train AI", *The Information* (April 13).
[113] Fuentes detalladas en anexo 1.

enfocándose en los beneficios indirectos para la mejora de rendimiento de publicidad de su modelo de negocio original [114].

- ***Eslabón 5, Plataformas de software y herramientas:*** en el eslabón de plataformas de software y herramientas de IA, los ingresos se basaron en suscripciones y licencias. Databricks reportó US$ 5,400 millones con un crecimiento superior al 65%. Hugging Face proyectó ingresos por US$ 130 millones para 2025, mientras que Suno lo hizo por US$ 300 millones. Por su parte, LangChain reportó ingresos de entre US$ 12 y US$ 16 millones [115].
- ***Eslabón 6, Desarrollo e implementación de aplicaciones:*** en la implementación de aplicaciones, los niveles de facturación estuvieron determinados por la capacidad de implementar soluciones de IA en segmentos verticales específicos. Accenture registró ingresos por US$ 69,673 millones —de los cuales $5,900 millones son atribuibles directamente a la IA, mientras que Palantir reportó US$ 4,475 millones generados por ventas de aplicaciones de IA en el mercado de defensa. Por su parte, los startups de aplicaciones de IA en verticales específicos demostraron ingresos comparativamente más reducidos: Harvey alcanzó los US$ 190 millones y Abridge US$ 100 millones [116].

En conclusión, la distribución por etapa de la cadena de valor de la IA confirma que los eslabones iniciales de la cadena concentran la mayor parte de la facturación consolidada. El eslabón de equipamiento lidera con US$ 620,099 millones, seguido por infraestructura, que registró ingresos por US$ 348,031 millones. En el extremo opuesto, los niveles posteriores de la cadena presentan una facturación comparativamente inferior: los laboratorios de modelos fundacionales totalizaron US$ 29,428 millones (no se consideran los ingresos de Google por no tener el detalle de lo que corresponde a Gemini), mientras que el desarrollo e implementación de aplicaciones alcanzó los US$ 10,665 millones. Por su parte, las plataformas de software y herramientas de IA, junto con la preparación y gestión de datos, reportaron US$ 5,844 millones y US$ 2,041 millones, respectivamente. Esta estructura evidencia que, a pesar del impacto tecnológico de los modelos de frontera, la monetización actual se localiza primordialmente en la provisión de hardware y servicios de cómputo (ver gráfico 6).

Gráfico 6
Ingresos por eslabón de la cadena de valor de la IA en Estados Unidos, 2025
(En millones de dólares)



Fuente: Elaboración propia sobre la base de fuentes públicas.
Nota: Excluye ingresos de Google en la etapa 4 de Modelos fundacionales, ya que la empresa no reporta de modo separado ese valor.

[114] Fuentes detalladas en anexo 1.
[115] Fuentes detalladas en anexo 1.
[116] Fuentes detalladas en anexo 1.

2. Costos operativos

La cadena de valor de la IA impone una estructura de OPEX recurrente y significativa, que altera el paradigma de costo marginal cercano a cero de los modelos de software tradicionales. Esta tecnología requiere flujos de caja constantes para el procesamiento de inferencias y la retención de talento especializado.

- ***Eslabón 1 -Equipamiento*:** en el eslabón de equipamiento, el OPEX reflejó la división del trabajo de la industria de semiconductores. Como es lógico, las empresas *fabless* concentraron sus gastos en I&D: NVIDIA reportó un OPEX de US$ 16,405 millones, destinando el 78.7% (US$ 12,914 millones) a I&D, mientras que AMD alcanzó los US$ 13,458 millones. Por otra parte, TSMC, concentrada en la fabricación de GPUs registró un OPEX de US$ 11,109 millones para las operaciones de manufactura. En contraste, los ensambladores de servidores operaron bajo una lógica de costos directos de fabricación y adquisición de componentes (*cost of goods* en ingles): Dell Technologies reportó costos de US$ 74,317 millones con un OPEX de US$ 15,013 millones, y Supermicro registró un costo de ventas de US$ 19,542 millones frente a un OPEX de US$ 1,177 millones. Foxconn reportó un OPEX de US$ 7,043 millones para gestionar una nómina de 633,167 empleados [117]. Nuevamente, los costos operativos de los ensambladores de servidores representan costos totales incluyendo equipamiento como dispositivos móviles para Foxconn y PCs para Dell.
- ***Eslabón 2 Infraestructura de IA***: el eslabón de infraestructura presentó una estructura de OPEX dominada por el consumo energético y gastos de personal. AWS registró gastos operativos de US$ 83,119 millones destinados principalmente al gasto energético y sistemas de enfriamiento. Google Cloud reportó US$ 44,795 millones, incluyendo personal técnico y red global, mientras que Oracle reportó costos por US$ 39,721 millones. Microsoft reportó en su segmento de Intelligent Cloud un OPEX de US$ 21,505 millones, complementado por un costo de ventas de US$ 40,171 millones, totalizando un OPEX de $ 61,676 millones. La "neo-cloud" CoreWeave registró un OPEX de US$ 3,724 millones, enfocado en tecnología y administración [118].
- ***Eslabón 3, Preparación y gestión de datos***: en la preparación y gestión de datos, el costo operativo fue intensivo en capital humano para procesos de RLHF (*Reinforcement Learning for Human Feedback*) [119]. Scale AI basó su estructura de costos en el uso de una red de más de 700,000 colaboradores. En términos generales, las firmas de etiquetado deben pagar entre 60% y 70% de sus ventas a mano de obra de etiquetado [120]. No obstante, el sector inició una transición hacia la generación de datos sintéticos, desplazando el costo desde las nóminas hacia el procesamiento algorítmico y el arrendamiento de servidores. Firmas como Pinecone y Gretel.ai mantuvieron estructuras de costos centradas en talento especializado y consumo de infraestructura de nube [121].
- ***Eslabón 4, Modelos fundacionales:*** los laboratorios de modelos fundacionales registraron un consumo de efectivo continuo debido al alquiler de capacidad de cómputo para entrenamiento e inferencia diaria. OpenAI proyectó un déficit operativo de US$ 115,000 millones entre 2025 y 2029, con un gasto estimado en inferencia de US$ 150,000 millones en dicho lustro. Anthropic estimó OPEX de US$ 5,200 millones en 2025. Además de la infraestructura, la retención de talento de los modelos fundacionales

[117] Fuentes detalladas en Anexo 2.
[118] Fuentes detalladas en Anexo 2.
[119] RLHF es el proceso donde analistas proveen retroalimentación para mejorar el desempeño de modelos fundacionales.
[120] Hornstein, J. and Weinberg, C. (2026). "Handshake, Mercor revenue surges on demand for human contractors to train AI", The Information (April 13).
[121] Fuentes detalladas en Anexo 2.

representó un costo crítico. Así también, Google DeepMind reportó una pérdida operativa de US$ 16,760 millones [122].

- ***Eslabón 5, Plataformas de software y herramientas:*** las plataformas de software y herramientas de IA operaron bajo el modelo *asset-light*. El OPEX se concentró en nóminas de ingeniería, alojamiento en nubes de terceros y marketing. Databricks gestionó un gasto de entre US$ 4,200 y US$ 4,800 millones para optimización de fuerza de ventas. Hugging Face reportó un OPEX de entre US$ 180 y US$ 260 millones para el mantenimiento de su repositorio y subsidio de capas gratuitas. Suno reportó costos operativos de entre US$ 280 y US$ 380 millones debido a la alta demanda de inferencia [123].
- ***Eslabón 6, Desarrollo e implementación de aplicaciones****:* en el eslabón de implementación de aplicaciones, el OPEX estuvo definido por el pago de tokens a modelos de terceros [124] y el gasto en recursos humanos (especialistas en las áreas de desarrollo de soluciones de IA). Palantir reportó un OPEX de US$ 2,272 millones, destinando US$ 1,056 millones a ventas y marketing. Startups como Harvey y Abridge reportaron gastos estimados de entre US$ 230 y US$ 290 millones y US$ 180 y US$ 240 millones respectivamente para financiar su integración en los sectores legal y de salud [125].

En resumen, la estructura de costos operativos de la IA impone cargos recurrentes que eliminan el paradigma del costo marginal cero. En la cadena de valor de la IA, el OPEX está dominado por el consumo eléctrico de data centers, el pago de personal especializado y el costo de inferencia diaria, proyectando déficits como los US$ 115,000 millones acumulados previstos por OpenAI hasta 2029. En los eslabones superiores, a pesar de que el gasto físico se externaliza a la nube, este es reemplazado por costos en recursos humanos y por el consumo de tokens. Esta dinámica se refleja en los US$ 2,272 millones de Palantir en 2025, así como en el uso de efectivo de startups como Harvey (US$ 230-290 millones) y Abridge (US$ 180-240 millones).

3. Rentabilidad

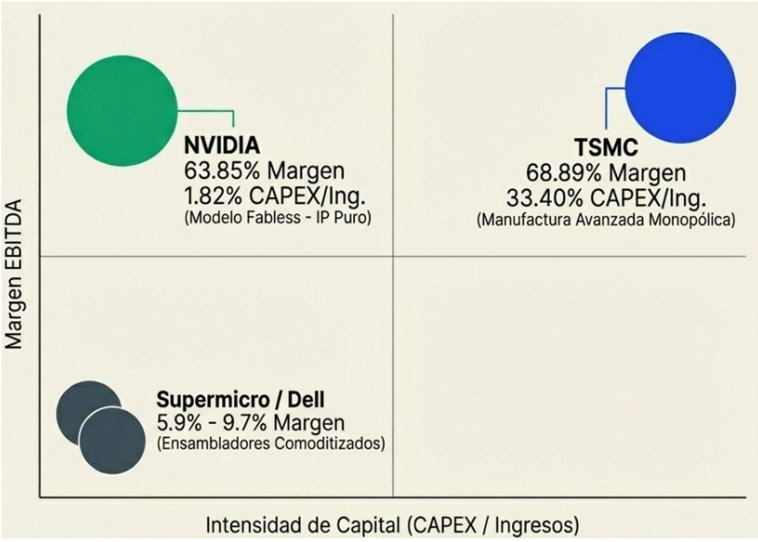
La rentabilidad en la cadena de valor de la IA revela una asimetría estructural, donde la captura de valor económico depende de la exclusividad tecnológica y el poder de fijación de precios. En este entorno, coexisten rentas elevadas en nodos críticos como el equipamiento, frente a una comoditización acelerada en otros eslabones.

- ***Eslabón 1, Equipamiento*****:** en el eslabón de equipamiento, las empresas *fabless* y de fundición registraron márgenes elevados. NVIDIA operó con un margen EBITDA del 63.85%, característico de una empresa que detenta un poder de mercado cercano al monopolio. De manera similar, TSMC reportó un margen EBITDA del 68.89%, mientras que AMD registró un margen del 19.34%. Por el contrario, los ensambladores enfrentaron una fuerte compresión de su rentabilidad: Dell Technologies reportó un margen EBITDA del 9.79%, Foxconn registró un 6.25% y Supermicro operó con apenas un 5.97% [126]. En conclusión, el eslabón de equipamiento revela una asimetría estructural extrema en términos de rentabilidad. El diseño "*fabless*" y la fundición avanzada gozan de rentas extraordinarias (>60% EBITDA), mientras que los ensambladores de servidores enfrentan una comprensión de márgenes típica de mercados más comoditizados (ver diagrama 16).

---

[122] Fuentes detalladas en Anexo 2.
[123] Fuentes detalladas en Anexo 2.
[124] En inteligencia artificial, los tokens son unidades de datos que los modelos fundacionales procesan, analizan y generan. Estos modelos descomponen el texto en tokens para comprender las relaciones entre ellos, en lugar de leer palabras completas.
[125] Fuentes detalladas en Anexo 2.
[126] Fuentes detalladas en Anexo 3.

Diagrama 16
Equipamiento: propiedad intelectual versus ensamblaje



Fuente: Elaboración propia.

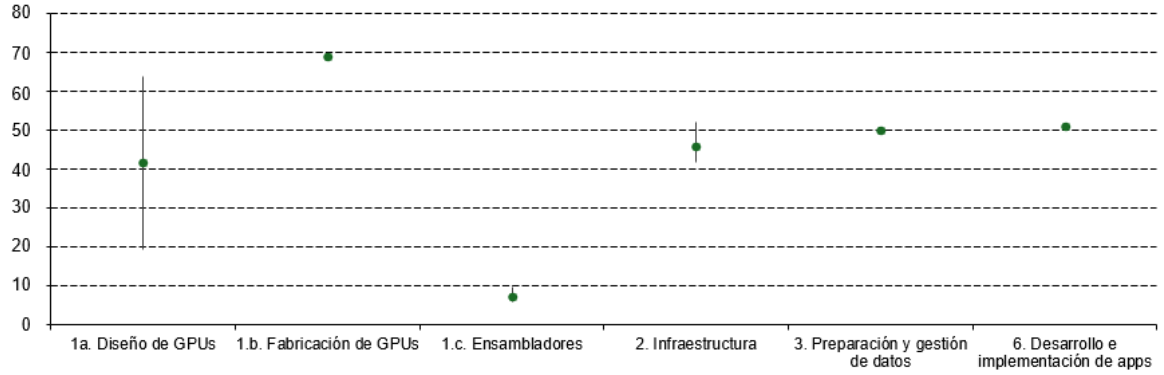
- ***Eslabón 2, Infraestructura de IA***: en el eslabón de infraestructura computacional, los márgenes EBITDA fueron elevados para todas las empresas del eslabón, aunque los mismos no consideran la poca vida útil de la infraestructura necesaria en esta etapa de la cadena de valor.[127] AWS lideró con un margen EBITDA del 52.09%, mientras que Microsoft Azure reportó en su segmento Intelligent Cloud un margen operativo del 41.96%. Google Cloud alcanzó un margen del 23.69% (que no es comparable con el resto de las empresas, pues ese margen considera la amortización, la cual es excluida del EBITDA). Por su parte, Oracle reportó un margen consolidado del 41.55%. La "neo-cloud" CoreWeave registró un margen de EBITDA del 46.93%[128].
- ***Eslabón 3, Preparación y gestión de datos:*** en la preparación y gestión de datos, los márgenes son elevados para la empresa que reportó datos. Scale AI mantuvo un margen de EBITDA cercano al 50% en 2024. Por otro lado, en el segmento de bases de datos vectoriales y generación de datos sintéticos, empresas como Pinecone y Gretel.ai no cuentan con información pública disponible para evaluar su rentabilidad operativa[129].
- ***Eslabón 4, Modelos fundacionales:*** los laboratorios de modelos fundacionales presentaron un déficit estructural debido a los elevados costos de entrenamiento e inferencia. OpenAI proyectó que no alcanzará la rentabilidad hasta 2030. Anthropic, por su parte, se mantiene operando sin alcanzar el equilibrio financiero. Google DeepMind registró una pérdida de US$ 16,760 millones en el año fiscal 2025, mientras que xAI acumuló pérdidas por US$ 7,800 millones en los primeros nueve meses del mismo año. Finalmente, Meta opera con márgenes negativos respecto a la monetización directa de sus modelos Llama (al ser un modelo abierto), enfocándose en la mejora del rendimiento de los anuncios de su modelo de negocio original[130].

[127] De acuerdo con estimaciones, los GPUs deben ser reemplazados cada 18 meses para acomodar el aumento en necesidad de poder de cómputo.
[128] Fuentes detalladas en Anexo 3.2.
[129] Fuentes detalladas en Anexo 3.3.
[130] Fuentes detalladas en Anexo 3.4.

- ***Eslabón 5, Plataformas de software y herramientas:*** en el eslabón de plataformas de software y herramientas de IA, la rentabilidad presentó resultados mixtos. Databricks operó con un flujo de caja positivo. Hugging Face logró alcanzar el punto de equilibrio financiero (breakeven). En contraste, la rentabilidad de Suno fue baja o negativa debido a los altos costos operativos requeridos para procesar siete millones de canciones diarias. Por su parte, LangChain operó con márgenes negativos, situación esperable al encontrarse aún en una fuerte fase de inversión [131].
- ***Eslabón 6, Desarrollo e implementación de aplicaciones:*** finalmente, en la implementación de aplicaciones, la rentabilidad estuvo determinada por la capacidad de resolver problemas verticales específicos. Accenture reportó un margen de EBITDA del 18.18%, (para el conjunto de la empresa) mientras que Palantir operó con un margen del 50.95%. Por su parte, los startups de aplicaciones de IA vertical operaron con márgenes negativos, ya que empresas como Harvey y Abridge priorizaron la fase de expansión y el escalamiento sobre la rentabilidad neta [132].

La dinámica de rentabilidad de la cadena de valor de la IA expone un panorama económico asimétrico a través de sus eslabones. En el primer eslabón, el diseño y manufactura avanzada registran márgenes de EBITDA elevados, con NVIDIA (Diseño de GPUs) alcanzando un 63.85% y TSMC (Fabricación de GPUs) un 68.89%, mientras que los ensambladores enfrentan una fuerte compresión que reduce la rentabilidad de Supermicro al 5.97%. El eslabón de infraestructura computacional mantiene rentas elevadas lideradas por el 52.09% de margen EBITDA de AWS. Posteriormente, la preparación y gestión de datos presenta márgenes del 50% en Scale AI, mientras que los laboratorios de modelos fundacionales operan con déficits estructurales masivos para subsidiar la adopción global (no incluidos en el gráfico por falta de datos exactos). En los eslabones posteriores, las plataformas de software no presentan datos exactos, pero reportan promedios cercanos al margen 0, contrastando con el eslabón de implementación de aplicaciones, donde la especialización permite a Palantir consolidar un margen del 50.95%, aunque otros actores de ese eslabón presentan márgenes negativos de startups en fase de expansión (Harvey y Abridge) (Ver gráfico 7).

Gráfico 7
Margen EBITDA por eslabón la cadena de valor de la IA en Estados Unidos, mínimos, máximos y promedio
(En porcentajes)



Fuente: Elaboración propia sobre la base de fuentes públicas.
Nota: Se incluye al eslabón de Modelos Fundacionales y de Plataformas con un valor de 0%, aunque ambos están en una etapa de *cash burn* (margen negativo).

[131] Fuentes detalladas en Anexo 3.5.
[132] Fuentes detalladas en Anexo 3.6.

4. Inversión de capital

En conclusión, la cadena de valor de la IA presenta una estructura de inversión de capital asimétrica, donde la mayor concentración de activos físicos se localiza en la provisión de infraestructura de nube para el entrenamiento de modelos y el apoyo a la inferencia operativa.

- ***Eslabón 1, Equipamiento*:** en el eslabón de equipamiento, la inversión de capital está concentrada en la manufactura física de semiconductores a cargo de TSMC, que registró un CAPEX de US$ 40,895 millones para el año fiscal 2025, respondiendo a la necesidad de financiar la infraestructura necesaria para la fabricación de chips. En contraste, las empresas enfocadas en el diseño de GPUs, reportan un Capex significativamente menor (NVIDIA con US$ 2,369 millones y AMD con US$ 1,000 millones). Finalmente, los ensambladores de servidores reportan volúmenes de inversión comparables (Dell Technologies reportó un CAPEX de US$ 2,652 millones, y Supermicro US$ 127 millones).
- La conclusión previa también se refleja en la relación CAPEX/ingresos que provee una perspectiva respecto de la intensidad de capital para cada segmento de este eslabón. Esta ratio refleja que la inversión se encuentra concentrada dentro del segmento de fabricación de GPUs[133]. La relación CAPEX/ingresos del segmento de fabricación de GPUs indica las elevadas barreras de entrada, lo que sustenta la posición cuasi monopólica de TSMC. Por otro lado, las relaciones en los otros dos segmentos denotan que la entrada en estos no está condicionada por la inversión de capital sino por otros factores, como la propiedad intelectual y los recursos humanos en el caso del diseño de GPUs. Adicionalmente, corresponde mencionar que las firmas enfocadas en el diseño de GPUs reportan el gasto operativo en capital humano e investigación en las categorías de OPEX e I+D.
- ***Eslabón 2, Infraestructura de IA:*** el eslabón de infraestructura computacional (nube y centros de datos) concentró el mayor volumen de inversión física de la cadena de valor, impulsado por la necesidad de hardware. Los hiperescaladores lideran esta tendencia: Amazon (AWS) registró un CAPEX de US$ 96,496 millones, Google Cloud anunció US$ 75,000 millones y Microsoft Azure totalizó US$ 64,550 millones para el año fiscal 2025. Estas inversiones se destinaron primordialmente a la adquisición de clústeres masivos de GPUs, sistemas de refrigeración y la construcción de instalaciones.
- La expansión de la capacidad de cómputo también fue impulsada por actores especializados y proveedores tradicionales en transición. Oracle reportó una inversión en activos fijos de US$ 21,215 millones, lo que representó un crecimiento del 209% respecto al ejercicio anterior. Por su parte, la "neo-cloud" CoreWeave registró desembolsos por US$ 10,309 millones en propiedad y equipamiento. Este despliegue de capital consolidó la transición de los centros de datos hacia infraestructura dedicada al entrenamiento de modelos y procesamiento de inferencias.[134] La inversión de capital de CoreWeave está creciendo exponencialmente, siendo estimada a US$ 35 mil millones para el 2026, lo que debe ser considerado con relación a ingresos proyectados de entre US$ 12 mil millones y US$ 13 mil millones[135].

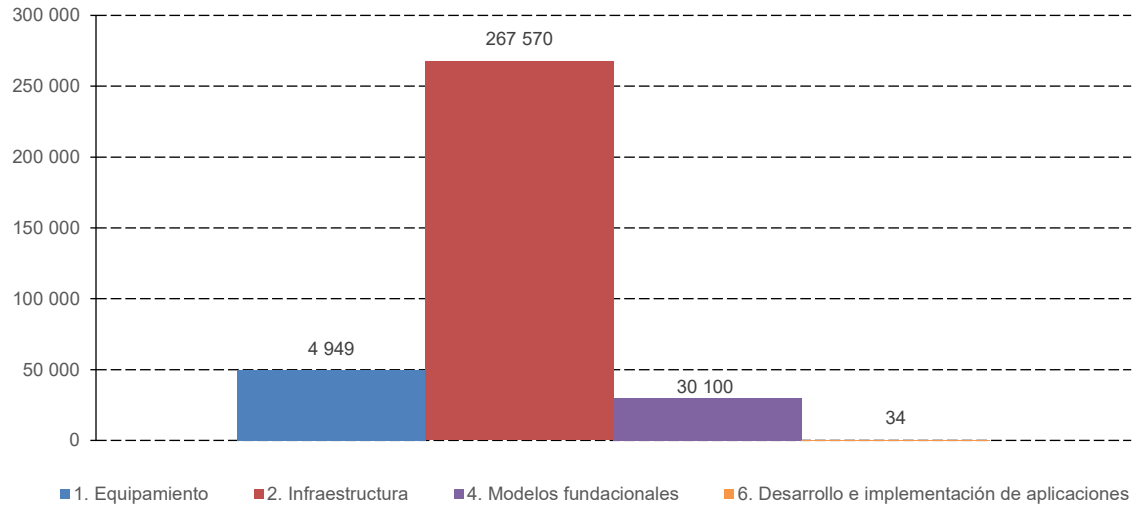
[133] Fuentes detalladas en el Anexo 4.
[134] Fuentes detalladas en el Anexo 4.
[135] Peers, M. (2026). "CoreWeave and risk". *The Information* (May 7).

- ***Eslabón 3, Preparación y gestión de datos***: en el eslabón de preparación y gestión de datos, la inversión se orientó a la estructuración de plataformas capaces de procesar grandes volúmenes de información, pero no se cuenta con información específica de las inversiones realizadas [136].
- A pesar de la falta de información, se reportó que Meta Platforms lideró una inversión estratégica de US$ 14,300 millones en Scale AI en junio de 2025, tras una ronda previa de US$ 1,000 millones en mayo de 2024. Por su parte, NVIDIA adquirió Gretel.ai en marzo de 2025 por un monto aproximado de US$ 320 millones, mientras que Pinecone captó US$ 100 millones en una ronda Serie C en octubre de 2025. Estos movimientos de capital financiaron la capacidad de generación de datos necesarios para el entrenamiento de modelos de frontera.
- **Eslabón 4, Modelos fundacionales:** en el centro de la industria, los laboratorios de modelos fundacionales operaron con una intensidad de capital enfocada en el entrenamiento inicial (pre-training) y en buscar el control de la infraestructura requerida. Un análisis de la inversión de este segmento requiere diferenciar los "pure plays" (OpenAI y Anthropic) de los jugadores integrados verticalmente (Google y Meta).
- En la primera categoría, considerando que las principales empresas no cotizan en bolsa, la estimación depende de anuncios públicos. Por ejemplo, OpenAI comunicó proyecciones de gasto en servidores de cómputo e infraestructura de entre US$ 600,000 y US$ 665,000 millones hasta el final de la década, de los cuales US$ 100,000 millones se destinarían al control directo de centros de datos. Anthropic anunció un plan de US$ 50,000 millones para el desarrollo de infraestructura propia. xAI invirtió US$ 18,000 millones en la adquisición de 555,000 GPUs NVIDIA en 2025, proyectando otros US$ 20,000 millones para la construcción de un centro de datos en 2026. De esta manera, para este eslabón se consideran los siguientes valores anualizados para el 2025: OpenAI $ 8 mil millones, Anthropic (inversión a través de terceros), y xAI$18 mil millones [137].
- ***Eslabón 5, Plataformas de software y herramientas:*** en el eslabón de plataformas de software y herramientas de IA, la inversión de capital presentó una asimetría respecto a las capas de infraestructura física. Estas entidades operaron bajo un modelo de *asset-light*, donde el CAPEX en propiedad, planta y equipamiento fue prácticamente nulo al externalizar la infraestructura pesada hacia los proveedores de nube. Bajo esta lógica, las empresas centraron su capital en el desarrollo de software en lugar de la construcción de complejos industriales o la adquisición de clústeres de GPUs. Plataformas como Hugging Face y Databricks delegaron el procesamiento y mantenimiento de sistemas a terceros, por lo que no tuvieron un requerimiento significativo de CAPEX. Incluso firmas con alta demanda de procesamiento, como Suno, mantuvieron su CAPEX en niveles mínimos al utilizar proveedores externos para el alojamiento y soporte de su infraestructura de información [138].
- ***Eslabón 6, Desarrollo e implementación de aplicaciones****:* en el último eslabón de la cadena, la estructura de inversión reforzó un modelo orientado a servicios, manteniendo el gasto en activos físicos en niveles mínimos. Startups de IA en segmentos verticales como Harvey y Abridge operaron con una inversión en activos fijos cercana a cero al delegar la infraestructura en hiperescaladores. Por su parte, actores de mayor escala como Palantir reportaron un CAPEX de US$ 33.9 millones en 2025 [139].

[136] Ver fuentes detalladas en Anexo 4.
[137] Ver fuentes detalladas en Anexo 4.
[138] Ver fuentes detalladas en Anexo 4.
[139] Ver fuentes detalladas en Anexo 4.6.

En conclusión, la cadena de valor de la IA presenta una estructura de inversión de capital asimétrica, donde la mayor concentración de activos físicos se localiza en la provisión de infraestructura de nube, con desembolsos anuales que superan los US$ 90,000 millones en casos como AWS. En un estrato intermedio se sitúa la fabricación de GPUs (con el caso de TSMC) y el diseño de semiconductores; con firmas como NVIDIA y AMD que invirtieron US$ 2,369 millones y US$ 1,000 millones en CAPEX respectivamente. En un eslabón intermedio similar se encuentra los modelos fundacionales, que en su conjunto reflejaron un CAPEX de US$ 26,000 millones al 2025 pero con anuncios de un fuerte aumento hacia el 2030. Por el contrario, los eslabones de software, orquestación e IA vertical operaron bajo un modelo *asset-light* con un CAPEX muy reducido, como se observó en Hugging Face, Databricks y Harvey, trasladando los requerimientos de capital físico hacia el gasto operativo por servicios de nube y la captación de talento especializado (ver gráfico 8).

Gráfico 8
CAPEX por eslabón de la cadena de valor, 2025
(En millones de dólares)



Fuente: Elaboración propia sobre la base de fuentes públicas.
Nota: Se omite el segmento de Preparación y gestión de datos y el de plataformas de software y de herramientas por falta de datos exactos, aunque se encontró evidencia de que su nivel de inversión no es significativo.

5. Capitalización de mercado

La estructura de capitalización en la cadena de valor de la IA revela una disparidad profunda en la valoración de los activos físicos frente a la propiedad intelectual, donde inversionistas asignan valores superiores basados en la capacidad de diseñar GPUs y desarrollar modelos fundacionales.

- ***Eslabón 1, Equipamiento*:** en la primera etapa de la cadena de valor de la IA (diseño y manufactura de equipamiento), las empresas con modelos *fabless* son las que registraron una mayor capitalización de mercado con relación a sus ingresos: NVIDIA alcanzó una valoración de US$ 4,340,000 millones, cotizando a 33.4 veces sus ingresos anuales, mientras que AMD se posicionó con una capitalización de US$ 310,000 millones, equivalente a 8.9 veces sus ingresos. En el eslabón de manufactura, TSMC exhibió una valuación de US$ 1.78 billones (14.6 veces sus ingresos) y el ensamblador Foxconn alcanzó los US$ 2.55 billones (11.9 veces sus ingresos). En contraste, los fabricantes de equipos responsables del ensamblaje de servidores presentaron valoraciones relativas menores

dado que su valor está determinado principalmente por capacidad operacional relativamente comodotizada: Dell Technologies cotizó a 1.0 veces sus ingresos (US$ 97,500 millones) y Supermicro a 0.9 veces (US$ 19,500 millones) [140].

- ***Eslabón 2, Infraestructura:*** en el eslabón de infraestructura de nube y centros de datos es complicado analizar la valuación del negocio, ya que tanto AWS, Azure y Google Cloud son empresas que cotizan en bolsa, e incluyen un negocio más extendido. En esos casos, las empresas matrices de los hiperescaladores reflejaron una prima por el crecimiento de sus divisiones de nube. Microsoft alcanzó los US$ 3.04 billones (10.8 veces ingresos) aunque ponderando por ingresos su negocio de infraestructura tuvo una valuación de US$ 1.15 billones, Alphabet registró US$ 3.70 billones (9.2 veces ingresos), aunque ponderando por ingresos su negocio en infraestructura tuvo una valuación de US$ 0.54 billones y Amazon totalizó US$ 2.35 billones, con una valuación ponderada por ingresos para AWS de 0.42 billones. Así también, Oracle expandió su valuación a US$ 429,662 millones (13.4 veces ingresos, y ponderado por ingreso su negocio de infraestructura tuvo una valuación de US$ 368,513 millones) y la "neo-cloud" CoreWeave consolidó una capitalización de US$ 39,279 millones (7.7 veces ingresos) [141].
- ***Eslabón 3, Preparación y gestión de datos***: en la preparación y gestión de datos, el valor se definió mediante rondas de capital y adquisiciones estratégicas. Scale AI alcanzó los US$ 29,000 millones tras la inversión de Meta en 2025, mientras que NVIDIA adquirió Gretel.ai por US$ 320 millones. Las bases de datos vectoriales como Pinecone registraron una valuación implícita de US$ 1,680 millones a inicios de 2026 [142].
- ***Eslabón 4, Modelos fundacionales:*** los laboratorios de modelos fundacionales operaron con valuaciones basadas en expectativas a pesar de sus déficits operativos. OpenAI alcanzó una valoración de US$ 730,000 millones, seguida por Anthropic con US$ 350,000 millones y xAI con US$ 250,000 millones [143].
- ***Eslabón 5, Plataformas de software y herramientas:*** en el eslabón de plataformas de software y herramientas de IA, los múltiplos respondieron a un modelo *asset-light* y con expectativas de crecimiento de ingresos en el futuro. Databricks alcanzó US$ 134,000 millones (24.8 veces ingresos) y Hugging Face US$ 4,500 millones (34.6 veces ingresos). LangChain registró una valuación de US$ 1,250 millones (80 veces ingresos) tras su Serie B, mientras que Suno alcanzó los US$ 2,450 millones [144].
- ***Eslabón 6, Desarrollo e implementación de aplicaciones***: por último, en la implementación de aplicaciones, la capitalización se sustentó en la capacidad de brindar soluciones a segmentos específicos del mercado. Palantir alcanzó una capitalización de US$ 361,932 millones (81 veces ingresos). Startups verticales como Harvey y Abridge registraron valuaciones de hasta US$ 11,000 millones y US$ 5,300 millones respectivamente [145].

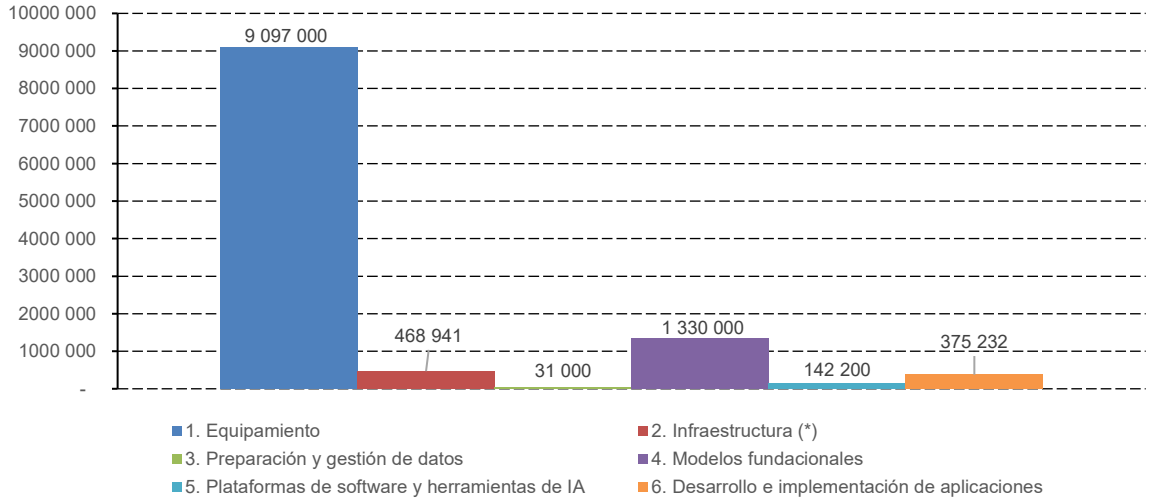
En síntesis, la estructura de capitalización en la cadena de valor de la IA revela una disparidad profunda en la valoración de los activos físicos frente a la propiedad intelectual (bajo la métrica valuación/ingresos anuales). Los eslabones de diseño y manufactura de equipamiento y la infraestructura de nube presentan una dualidad marcada: mientras los diseñadores "fabless" como NVIDIA cotizan a múltiplos de 33.4x ingresos y los hiperescaladores como Microsoft alcanzan las 10.8x,

[140] Ver fuentes detalladas en Anexo 5.
[141] Ver fuentes detalladas en Anexo 5.
[142] Ver fuentes detalladas en Anexo 5.
[143] Ver fuentes detalladas en Anexo 5.
[144] Ver fuentes detalladas en Anexo 5.
[145] Ver fuentes detalladas en Anexo 5.

los ensambladores transaccionales de hardware como Supermicro y Dell se ven penalizados con múltiplos de apenas 0.9x y 1.0x respectivamente. En un estrato intermedio, la gestión de datos y los laboratorios de modelos fundacionales sostienen valoraciones agresivas basadas en la escasez de insumos y el potencial estratégico derivado de la propiedad intelectual. Finalmente, en la parte posterior de la cadena, los eslabones de software y orquestación e implementación de aplicaciones registran los múltiplos más elevados del sistema, ejemplificados por las 80x ingresos de LangChain y las 81x de Palantir, consolidando la tesis de que el mercado asigna una prima extraordinaria a las capas ligeras en activos (asset-light) que operan en la última etapa del valor económico.

Bajo la métrica de valuación en dólares, el eslabón de diseño y manufactura de equipamiento representa la mayor parte de la valuación total del ecosistema, con un monto de US$ 9,097,000 millones. En segundo lugar, se encuentra el segmento de infraestructura, con una valuación de US$ 2,515,529 millones. En un estrato intermedio se ubican los laboratorios de modelos fundacionales con una valuación de US$ 1,330,000 millones, mientras que las capas superiores de la cadena —desarrollo e implementación de aplicaciones, plataformas de software y gestión de datos— presentan valuaciones significativamente menores en términos absolutos (US$ 375,232 millones, US$ 142,200 millones y US$ 31,000 millones), evidenciando que la captura de valor bursátil está hoy anclada primordialmente en el control de los eslabones iniciales (ver gráfico 9).

Gráfico 9
Capitalización de mercado consolidada por etapa de la cadena de valor de la IA en Estados Unidos, marzo de 2026
(En millones de dólares)



Fuente: Elaboración propia sobre la base de fuentes públicas.
Nota: Se omite la valuación de Google en Modelos Fundacionales ya que no reporta ingresos y por ende no puede estimarse la valuación del segmento. Así también en ese segmento también se excluye la valuación de Meta (por su modelo Llama) ya que es un modelo que no genera ingresos de modo directo.

6. Conclusión

El dimensionamiento económico de la cadena de valor en términos de las cinco métricas utilizadas permite generar las siguientes conclusiones en torno a la creación de valor por eslabones:

- **Ingresos:** los eslabones iniciales de la cadena concentran la mayor parte de los ingresos. Reflejando la necesidad de equipamiento y las operaciones de data centers (que incluyen todos los servicios de cloud), se estima que el equipamiento (desarrollo y manufactura de semiconductores y equipamiento) concentra un 61% de los ingresos (ver cuadro 8).

Cuadro 8
Cadena de Valor de IA (Estados Unidos): ingresos totales, 2025
(En millones de dólares)

Eslabón	Ingresos totales	Porcentajes
1. Equipamiento	620 099	61
2. Infraestructura	348 031	34
3. Preparación y gestión de datos	2 041	0
4. Modelos fundacionales	29 428	3
5. Plataformas de software y herramientas	5 844	1
6. Desarrollo e implementación de aplicaciones	10 665	1

Fuente: Elaboración propia.

- **Costos operativos:** debido a la necesidad no solo de desarrollar modelos y aplicaciones sino también de operarlos por inferencia, la estructura de costos operativos difiere de aquella de modelos de software tradicionales donde el costo marginal de operaciones tiende a cero. Es así como los costos operativos como proporción de ingresos tienen un primer efecto mitigante en los ingresos, especialmente en los eslabones de infraestructura y especialmente en modelos fundacionales, donde los costos operativos son extremadamente altos debido a la necesidad de poder de cómputo para el entrenamiento de modelos (ver cuadro 9).

Cuadro 9
Cadena de Valor de IA (Estados Unidos): promedio de costos operativos/ingresos, 2025
(En porcentajes)

Eslabón	OPEX/Ingresos
1. Equipamiento	10,35
2. Infraestructura	66,96
3. Preparación y gestión de datos	Sin información
4. Modelos fundacionales	117,93
5. Plataformas de software y herramientas	87,18
6. Desarrollo e implementación de aplicaciones	57,54

Fuente: Elaboración propia.

- **Rentabilidad:** la rentabilidad en la cadena de valor de la IA revela una asimetría estructural, donde la captura de valor económico depende de la exclusividad tecnológica y el poder de fijación de precios. En este entorno, coexisten rentas elevadas en nodos críticos frente a una comoditización acelerada en otros eslabones. El diseño y manufactura de GPUs registran márgenes de EBITDA extremadamente altos, reflejando rentas cuasi monopólicas de NVIDIA y TSMC. El eslabón de infraestructura mantiene rentas elevadas lideradas por AWS. La preparación y gestión de datos presenta márgenes del 50% en Scale AI, mientras que los laboratorios de modelos fundacionales operan con déficits estructurales masivos debido a los costos de entrenamiento de modelos. Finalmente, si bien las plataformas de software operan en promedio cercanos al margen 0, estas contrastan con el eslabón de implementación de aplicaciones, donde la especialización permite a Palantir consolidar un margen del 50.95% (Ver cuadro 10).

Cuadro 10
Rentabilidad (Estados Unidos): margen EBITDA por eslabón, 2025
(En porcentajes)

Eslabón	Máximo	Mínimo	Promedio
1. Equipamiento	63,85	19,34	41,60
2. Infraestructura	68,89	68,89	68,89
3. Preparación y gestión de datos	9,79	5,97	7,34
4. Modelos fundacionales	52,09	41,55	45,63
5. Plataformas de software y herramientas	50,00	50,00	50,00
6. Desarrollo e implementación de aplicaciones	50,95	50,95	50,95

Fuente: Elaboración propia.

- **Inversión:** La cadena de valor de la IA presenta una estructura de inversión de capital asimétrica, donde la mayor concentración de activos físicos se localiza en la provisión de infraestructura de nube, con desembolsos anuales que superan los US$ 90,000 millones en casos como AWS. En un estrato intermedio se sitúa la fabricación de GPUs y el diseño de semiconductores; con firmas como TSMC, NVIDIA y AMD y los Modelos Fundacionales (ver cuadro 11).

Cuadro 11
Inversión, Estados Unidos, 2025
(En millones de dólares)

Eslabón	CAPEX	Capex/ingresos (*En porcentajes*)
1. Equipamiento	49 349	7,96
2. Infraestructura	267 570	76,88
3. Preparación y gestión de datos	0	0
4. Modelos fundacionales	26 000	88,35
5. Plataformas de software y herramientas	0	0
6. Desarrollo e implementación de aplicaciones	34	0,76

Fuente: Elaboración propia.

Las ratios de capex/ingresos son promedios con lo cual esconden diferencias sustanciales. Por ejemplo, CoreWeave, una neocloud está invirtiendo tres veces su volumen de ingresos, mientras que empresas como Amazon o Alphabet invierten entre 50% y 40%.

- **Capitalización de mercado:** la capitalización del eslabón de equipamiento, infraestructura y del desarrollo de modelos fundacionales explican donde los inversionistas perciben los puntos de la cadena donde se crea más valor (ver cuadro 12).

Cuadro 12
Cadena de valor de la IA en Estados Unidos: valor de mercado por eslabón, 2025
(En millones de dólares y en porcentajes)

Eslabón	Empresas	Valoración	Porcentaje
1. Equipamiento	NVIDIA[a] AMD[a] TSMC[a] Foxconn[b] Supermicro[b] Dell Technologies[b]	9 097 000	67,4
2. Infraestructura de cómputo para IA	AWS[c] Microsoft Azure[c] Google Cloud[c] Oracle[c] CoreWeave[a]	2 515 529	18,6

Eslabón	Empresas	Valoración (en millones US$)	Porcentaje
3. Preparación y gestión de datos (++)	ScaleAI[d] Gretel.ai[e] Pinecone[d]	31 000	0,2
4. Modelos fundacionales (+)	OpenAI[f] Anthropic[f] xAI[d]	1 330 000	9,9
5. Plataformas de software y herramientas (++)	Databricks[d] Hugging Face[d] LangChain[d] Suno[d]	142 200	1,1
6. Desarrollo e implementación de aplicaciones (++)	Palantir[a] Harvey[d] Abridge[d]	375 232	2,8
Total		13 490 961	

Fuente: Elaboración propia.
[a] Valoración bursátil compilada de Yahoo Finance.
[b] Incluye la valoración completa de la empresa y no solamente aquella asignada al ensamblaje de servidores IA.
[c] Valoración bursátil total de Yahoo Finance ponderada por ingresos en segmento de la nube, lo que podría subestimar el valor asignado a los servicios de la nube.
[d] Valoración estipulada a partir de la última ronda de financiamiento.
[e] Precio de adquisición.
[f] Valoración pre-oferta pública (Fuente: Reportes públicos).

III. La cadena de valor de la inteligencia artificial en China y Europa: estructura y dinámicas geopolíticas

El capítulo anterior del documento se enfocó en la cadena de valor de acuerdo con los actores y dimensionamiento económico de Estados Unidos. En el entramado global, Estados Unidos está enfrentado a una competencia geopolítica con la República Popular China, mientras que Europa trata de establecer una posición independiente de ambos polos a partir de una estrategia de soberanía digital. Con base en ello, las siguientes secciones retoman la estructura de cadena de valor de la IA para estudiar los desarrollos de China y Europa. Una vez hecho este análisis, se detalla las implicancias de competencia geopolítica entre estos tres polos mundiales.

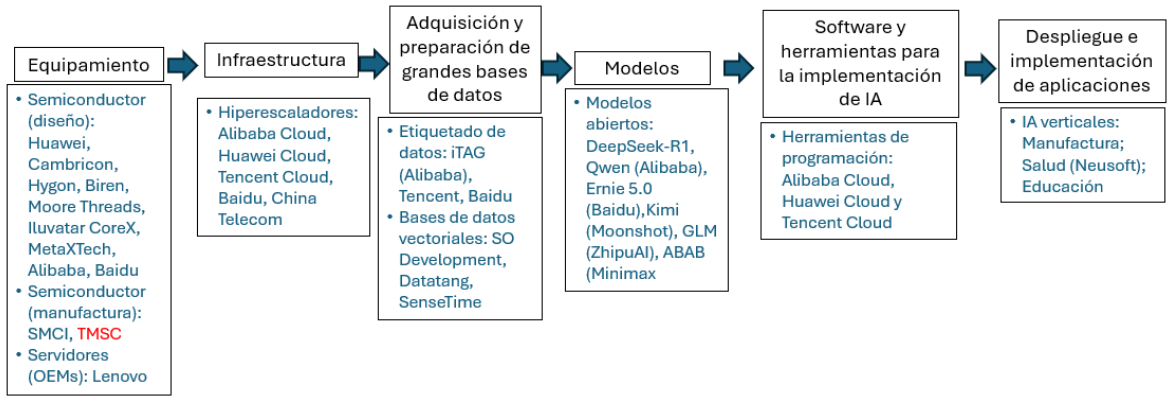
A. La cadena de valor de la inteligencia artificial en la República Popular China

China anunció que la IA representaba un área prioritaria de desarrollo en el 2017 con la formulación del Plan de Desarrollo de la Nueva Inteligencia Artificial, como complemento a la iniciativa de "Made in China 2025". El objetivo de este país en lo referente a la inteligencia artificial no es de posicionarse en algunos eslabones de la cadena de valor para poder competir con Estados Unidos. El propósito es de construir una completa cadena de valor paralela que permita al país desplegar una posición autónoma para transformarse en una alternativa competitiva a la estadounidense. Es por ello por lo que la cadena de valor de China está estructurada de la misma manera que la estadounidense, donde firmas chinas están posicionadas en cada eslabón, desplegando posiciones y movimientos similares a los analizados en el caso anterior con el objetivo de alcanzar la soberanía tecnológica. De todas maneras, al igual que en el caso estadounidense, la cadena de valor china también contiene puntos de riesgo, aunque estos son de diferente tipo.

1. Estructura de la cadena de valor de la IA en China

La cadena de valor de la IA en República Popular China presenta una estructura similar a la de EE. UU., aunque ciertos estadios ya demuestran desde su inicio una integración vertical basada en la presencia de hiperescaladores como Baidu, Alibaba y Tencent en múltiples eslabones (ver diagrama 17).

Diagrama 17
China: estructura de la cadena de valor de la IA



Fuente: Elaboración propia.
Nota: Azul indica desarrollo/capital local.

Contrariamente al caso estadounidense donde la integración vertical en la cadena de la IA ha sido un proceso gradual en respuesta a imperativos de control estratégico y la reducción de costos de transacción, en China la cadena de valor denota una integración desde los orígenes del ecosistema. Por ejemplo, los hiperescaladores (Alibaba Cloud, Tencent Cloud y Baidu) ya se han integrado en el diseño de semiconductores, la adquisición de bases de datos, el desarrollo de modelos y las herramientas de implementación. De todas maneras, en paralelo con los jugadores integrados verticalmente, ciertos eslabones presentan jugadores independientes, como ser en el diseño de semiconductores (Cambricon, Hygon, y otros), adquisición de datos (SO, Datatang), desarrollo de modelos (DeepSeek, Moonshot, ZhipuAI y Minimax), y despliegue de aplicaciones (Neusoft). Los siguientes apartados presentan detalles para cada eslabón. En ciertos casos como ser el de semiconductores se detallan asimismo las tensiones geopolíticas entre China y Estados Unidos debido a su importancia en el desarrollo del sector.

Equipamiento

El proceso de desarrollo de una industria local china de semiconductores comenzó en el 2014 cuando el Consejo del Estado, la autoridad máxima del gobierno, determinó que el país debía alcanzar una autonomía total en la cadena de aprovisionamiento de semiconductores para el 2030. Este objetivo de autonomía industrial estaba fundamentado por las iniciativas restrictivas impuestas por el gobierno estadounidense en lo que respecta a la exportación del equipamiento chino de telecomunicaciones. La irrupción de la IA y el enfrentamiento chino-estadounidense en lo referente al acceso a semiconductores para esta tecnología puso de manifiesto aún más la importancia del objetivo de autonomía industrial.

China reconoce que su posicionamiento en el segmento de semiconductores todavía no está a la altura del desempeño de firmas como NVIDIA (de Estados Unidos) para el diseño de GPUs. Ello puede verificarse en el cuadro 14, el cual compara el poder de procesamiento de chips diseñados en EE. UU. y China.

Cuadro 13
Poder de procesamiento de semiconductores usados en IA
(En Teraflops)

Chip	Poder de procesamiento[a]	Empresa productora	País
B300 (2025)	60 000	NVIDIA	Estados Unidos
B200 (2024)	36 000	NVIDIA	Estados Unidos
B30A (2025)	30 000	NVIDIA	Estados Unidos
H200 (2024)	15 840	NVIDIA	Estados Unidos
H100 (2022)	15 832	NVIDIA	Estados Unidos
Ascend 950PR (2026)	15 600	Huawei	China
Ascend 910C (2024)	10 240	Huawei	China
BW1000	5 600	Hygon	China
Ascend 910B (2022)	5 120	Huawei	China
A100 (2020)	4 992	NVIDIA	Estados Unidos
Siyuan 590	4 493	Cambricon	China
H20 (2023)	2 368	NVIDIA	Estados Unidos

Fuente: Bernstein Research.
[a] Mide el poder de procesamiento para cargas de trabajo de IA. Capacidad teórica máxima.

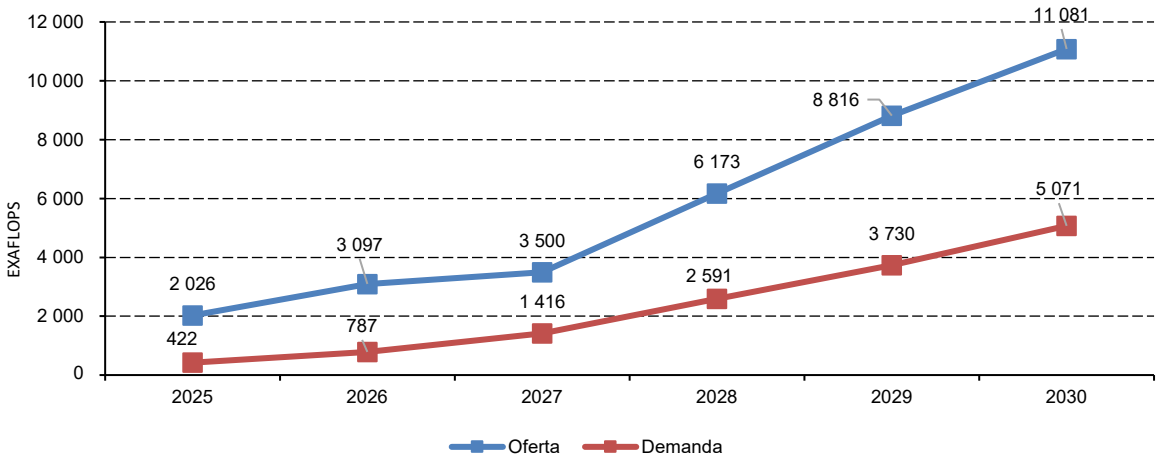
El chip de IA más poderoso producido en China es el Ascend 950C de Huawei, que presenta cerca del poder de procesamiento que la generación de Hopper (marcada con una letra H) de NVIDIA aunque mucho menor que la de Blackwell (marcada con la letra B). Es importante mencionar que DeepSeek está coordinando directamente con Huawei para correr sus modelos en el GPU de la última generación. Simultáneamente, dos fabricantes chinos, Iluvatar y Biren Technology están trabajando en el desarrollo de GPUs con poder de procesamiento similar al H200 de NVIDIA.

A partir de esta situación, la estrategia china está basada en dos ejes paralelos. Por un lado, existe una negociación permanente con Estados Unidos para poder acceder a los productos de NVIDIA principalmente para el entrenamiento de sus modelos, mientras que por el otro se desarrolla una industria de semiconductores de IA local, guiada por objetivos de soberanía tecnológica.

Históricamente, los operadores de plataformas que ofrecen servicios de IA en China han sido un cliente importante de NVIDIA. Por ejemplo, hasta el 2025, ByteDance fue el mayor adquirente de chips de NVIDIA [146]. Sin embargo, el acceso a chips de NVIDIA estaba limitado a la adquisición de modelos inferiores como el H20, y ocasionalmente a los H200. En agosto del 2025, el gobierno chino requirió a empresas locales interrumpir la adquisición de GPUs de NVIDIA, como parte de la estrategia enfocada en el desarrollo de una industria local como alternativa al proveedor estadounidense. Sin embargo, como se indicó arriba, una de las debilidades de los productos locales es que los productos son menos potentes y su capacidad de procesamiento está limitada tan solo al trabajo de inferencias (o sea implementación de servicios de IA donde la plataforma provee respuesta a preguntas e implementa tareas) y no al entrenamiento, donde el modelo es desarrollado a partir del procesamiento de grandes volúmenes de información. De todas maneras, la oferta china de chips enfocados en inferencias ya excede las necesidades del mercado local (ver gráfico 10).

[146] Liu, Q. (21025). "China is slowly but surely breaking free from NVIDIA". *The Information* (November 26).

Gráfico 10
China: oferta y demanda de chips de inferencia de IA, 2025-2030
(En Exaflops)



Fuente: Jefferies, Institute for Progress.
Nota: Un exaflop es una medida de velocidad de cómputo equivalente a 10^{18} operaciones por segundo.

La oferta de chips de inferencia indicada en el grafico 10 está compuesta por la producción local y aquella proveniente de los chips que están siendo autorizados a ser importados. De todas maneras, considerando que la producción local ya satisface la demanda, el gobierno chino requirió a empresas locales suspender la importación de chips H20 de NVIDIA [147].

Por otro lado, reconociendo que la producción local de chips necesarios para el entrenamiento de modelos todavía no puede satisfacer la demanda local, a finales del 2025, el gobierno chino permitió su importación. Sin embargo, en ese momento los obstáculos para la importación de chips avanzados de NVIDIA fueron planteados por el lado estadounidense. En la última década, el gobierno de EE. UU. ha puesto en práctica controles para la exportación de semiconductores avanzados a China (por ejemplo, los chips marcados con la letra B por Blackwell están totalmente prohibidos). Esta restricción también se aplicó a empresas que producen equipamiento para la manufactura de semiconductores de última generación, como lo fue la intervención estadounidense para imponer la prohibición de la exportación de tecnología para este propósito producida por ASML, una empresa holandesa.

En respuesta a la incertidumbre en el aprovisionamiento que apareció en el 2022, las empresas tecnológicas chinas comenzaron a moverse rápidamente para reducir la dependencia de NVIDIA. Originalmente, cuando las dificultades de importación comenzaron a manifestarse, esto tomó al gobierno chino por sorpresa. Los fabricantes locales carecían de experiencia y equipamiento para la producción de chips equiparables a los de NVIDIA. Simultáneamente, en noviembre del 2025, el gobierno de EE. UU. comenzó a estudiar la posibilidad de autorizar la exportación del modelo H200, el cual es dos veces más potente que el H20, el cual permite el entrenamiento de modelos.

En respuesta a esta consideración, el Ministerio de Tecnología de la Información y la Comisión de Desarrollo Nacional y Reforma de China convocó a los mayores clientes de NVIDIA, lo que incluía a ByteDance, Tencent y Aliaba, para informarse de cuántos GPUs ya habían adquirido, cuál era su uso, y cuál era la oportunidad de sustitución de importación. El objetivo manifestado en esas reuniones fue que la adquisición de GPUs fabricados en China no era solamente una política industrial, sino una prioridad macroeconómica de tipo estratégico.

[147] Liu, Q. (2025). "China Demands Companies to Halt Nvidia Chip Orders Over Security Concerns". *The Information* (August 12).

Sin embargo, esas consultas no produjeron resultados. El gobierno comprendió que los GPUs de NVIDIA continuaban siendo superiores a los producidos localmente para el entrenamiento de modelos, y los clientes chinos continuaron intentando adquirirlos, si las condiciones de exportación lo permitían. Es en ese momento que el gobierno cambió de estrategia migrando de la consulta informativa a la imposición regulatoria. La Administración de Ciberseguridad de China ordenó a empresas, como ByteDance y Alibaba, interrumpir toda prueba y compra de chips de NVIDIA. Esto ocurrió a pesar de que muchas empresas ya tenían órdenes de compra en procesamiento. Al mismo tiempo, el gobierno ordenó a empresas a reorientar procesos de compra a alternativas domésticas, como Huawei y Cambricon Technologies. Con ese mandato, muchas firmas ya han comenzado a migrar y a reorientar su infraestructura a partir de los chips locales. El problema en la actualidad es la capacidad limitada de producción doméstica para satisfacer las necesidades locales. Es por ello por lo que el gobierno ha invertido más de US$ 150 mil millones para estimular el desarrollo de otras empresas productoras de chips. A esto se suma la inversión de fondos privados en empresas como Biren Technology, Moore Threads, Iluvatar Corex y MetaXTech, que están desarrollando arquitecturas compatibles con CUDA [148] para facilitar la migración desde NVIDIA (ver cuadro 14).

Cuadro 14
China: empresas privadas para el desarrollo de GPUs

Empresa	Año de fundación	Fondos (En US$)	Valor de mercado	Inversores clave	Privada/pública
Biren Technology	2019	921M	2,07	Qiming Venture partners, GL Ventures	Cotizada en Hong Kong
Moore Threads	2020	525M	3,4 B	Honshan, 5Y Capital, China Mobile	Cotizada en Shanghai
Iluvatar CoreX	2020	177,9M	1B+	Hongsgan	Cotizada en Hong Kong
MetaXTech	2020	164B	2,15B	Hongshan, Matrix Partners China, Lightspedd China Partners	Cotizada en Shanghai
Huawei HSSilicon	1991			Privada	
Alibaba T-Head	2018	- - -	- - -	OPA planeada	
Baidu Kunlunxin	2011	- - -	- - -	OPA planeada	
Cambricon	2016	- - -	- - -	- - -	Cotizada en Shanghai
Hygon	2014	- - -	- - -	- - -	Cotizada en Shanghai
Tencent Enflame	2018	- - -	- - -	OPA planeada	

Fuente: The Information.

El ecosistema chino de producción de chips de IA también también incluye más de diez empresas diseñando GPUs, lo que incluye instituciones públicas que han reclutado personal con años de experiencia en NVIDIA, AMD e Intel, que han retornado al país. Esto incluye, además de Huawei, a las divisiones de chips de grandes empresas: T-head de Alibaba Group, ByteDance, y Kunlunxin de Baidu. En este campo, Huawei se ha posicionado como el líder indiscutible en China con su serie Ascend, proyectando una producción de 1.6 millones de unidades para 2026. Sin embargo, Huawei estima que podrá desarrollar semiconductores de frontera capaces de entrenar modelos fundacionales tan solo en el 2028[149]. Al momento, los productores chinos han capturado 35% del mercado estimado en US$ 16 mil millones [150]. De todas maneras considerando el estímulo a la producción local, se estima que

[148] La arquitectura de CUDA (*Compute Unified Device Architecture*) provee la capa de software donde operan la mayoría de las bibliotecas de *Deep learning*, como PyTorch y TensorFlow.

[149] Tobin, M. (2026). "How China built a chip industry, and why it's still not enough", *The New York Times* (February 14).

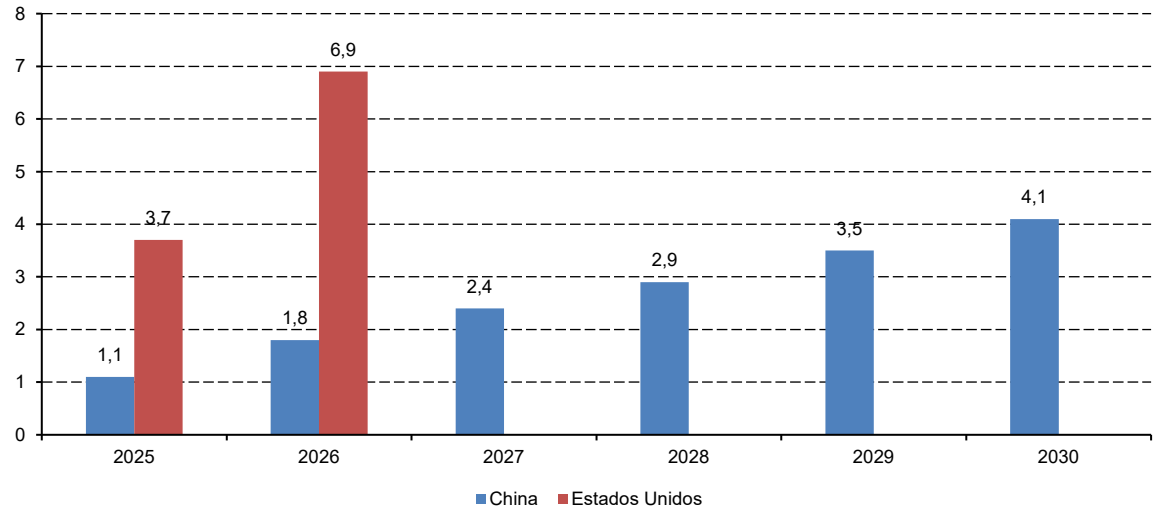
[150] Si, M. (2026). "Chinese AI chips gain huge traction in market," *China Daily* (January, 30). Disponible en: https://www.chinadaily.com.cn /a/202601/30/WS697ccbf7a310d6866eb36b18.html

en el curso de 2026, la cuota del mercado de GPUs en China para NVIDIA disminuirá a 8% de 40% en el 2025, mientras que Huawei alcanzará 50%. [151] Es importante mencionar que la estrategia de todas estas empresas no es de competir directamente con NVIDIA en la manufactura de GPUs para el entrenamiento, sino desarrollar chips para inferencias. Aun así, si el desempeño de los productos chinos no se compara con los de NVIDIA, los fabricantes locales compensan la desventaja agrupando unidades en clústeres de chips.

Simultáneamente, China está desarrollando capacidad de manufactura autónoma, a partir de SMIC, una empresa china y de la inversión de TSMC de Taiwán en una subsidiaria en Shanghái. En el área de avances en manufactura, SMIC ha logrado avances en nodos de 7nm [152] (N+2) y está realizando pruebas en 5nm utilizando litografía DUV mediante técnicas de cuádruple patrón (SAQP), evitando las restricciones de equipos EUV. Hacia finales del 2025, se estima que la capacidad de manufactura de SMIC fue de 45,000 obleas por mes, incrementándose a 60,000 en 2026 y 80,000 en 2027. Una de las fábricas de SMIC está completamente dedicada a la manufactura de chips de Huawei. La expansión de capacidad permitirá a SMIC apoyar también a Cambricon, MetaX y Biren.

Asimismo, el gobierno chino ha encarado políticas orientadas a la coordinación de capacidad de productores locales. A pesar de todas estas iniciativas, la capacidad de manufactura de China todavía no alcanza los niveles estadounidenses (ver gráfico 11).

Gráfico 11
Capacidad comparada de manufactura de chips de AI, 2025-2023
(En nodos de 7nm)



Fuente: Jefferies; Institute for Progress.

Para acelerar la producción de chips de IA, desde el 2025, numerosos gobiernos provinciales chinos también han estructurado fondos y proyectos piloto orientados a subsidiar empresas en el desarrollo de clústeres de semiconductores con el objetivo de asegurar la provisión de chips necesarios para reemplazar a NVIDIA.

[151] Liu, Q. (2026). "China's most important chip design firms". *The Information* (April 19).

[152] El nodo 7nm es un proceso de fabricación de semiconductores que permite una manufactura con 40% reducción en costos energéticos y 10-15% incremento en productividad. Esta tecnología fue introducide en 2018 por TSMC y posteriormente por Samsung y SMIC.

Finalmente, para orientar el desarrollo de la industria en el eslabón, la Academia China de Tecnologías de Información y Comunicación (CAICT) ha comenzado a emitir certificados que atestan que los semiconductores producidos localmente pueden correr rutinas de inferencia en modelos como DeepSeek R1. Esto genera una guía para que empresas con menos recursos tecnológicos puedan seleccionar los semiconductores más apropiados. Este proceso aparece ser más cercano a la certificación técnica de conformidad o adaptación. De acuerdo a ello, la entidad que emite estos certificados (CAICT) opera como institución técnica de referencia en el ecosistema chino de información y comunicaciones. En el detalle institucional, la CAICT explica que desde 2018 desarrolla una tercera parte de evaluación en IA y que ha construido el sistema *Trusted AI Evaluation* [153], organizado en tres capas: testeo de producto y servicio, evaluación de madurez de aplicación y evaluación de riesgo confiable. La CAICT indica que dentro de la primera capa se incluyen pruebas de chips y que el instituto lideró la norma ITU-T F.748.11 (2020), que se describe como el primer estándar internacional de evaluación de chips de IA [154].

Es importante mencionar que cuando se lanzó la evaluación de adaptación de DeepSeek, se señalaron dos objetivos: i) evaluar el efecto de adaptación del modelo en sistemas de hardware y software, incluyendo chips, equipos de cómputo, clústeres y plataformas; y, ii) reflejar la facilidad de uso del stack de software y herramientas, así como el costo de despliegue y desarrollo. Así, las pruebas abarcaron inferencia, fine-tuning y entrenamiento, utilizando la herramienta AISHPerf, con evaluación de costo de adaptación, completitud funcional, efecto de optimización e indicadores de desempeño [155]. Esto sugiere que la certificación no solo mide si el chip pone en funcionamiento el modelo, sino también si la combinación entre hardware, framework y modelo resulta funcional y económicamente utilizable [156].

En mayo de 2025, se informó que la tarjeta de cómputo MTT S4000 de *Moore Threads* pasó la verificación de inferencia para DeepSeek-R1 671B ante la CAICT. El resultado fue considerado conforme a los requisitos de evaluación de adaptación de paso entre chip de IA y el modelo, y la tarjeta obtuvo un certificado de detección, suficiente para uso en inferencia. Esto permite reconstruir una secuencia operativa razonable: presentación del producto, configuración del entorno de prueba, ejecución de escenarios definidos, verificación de estabilidad y conformidad con criterios de paso, y emisión de certificado cuando se cumplen los requisitos [157].

Como puede observarse, la organización del eslabón de equipamiento y su coordinación con el de infraestructura es más flexible y fragmentado que el estadounidense, con múltiples fabricantes de semiconductores combinado con el gobierno coordinando pruebas y estándares. De todas maneras, esta directiva no resolvió el problema de sustitución completa dado que muchas firmas que diseñan chips en China dependen de manufacturas en el extranjero.

El objetivo de China no es solamente sustituir las importaciones de chips para responder a los embargos estadounidenses sino también contribuir a exportaciones a otros países. El objetivo exportador es crear paquetes que incluyan procesadores, modelos y software (el AI stack) para ser exportados a mercados emergentes. Esto también es fuente potencial de enfrentamiento con Estados Unidos, que observa este intento como una maniobra para desplazar la hegemonía de este país [158]. Este enfrentamiento puede acelerar la tensión originada por países que deberán elegir entre dos ecosistemas diferentes.

---

153 Fuente: CAICT. https://service.caict.ac.cn/ywjs/202104/t20210430_3029.html

154 Fuente: CAICT. https://service.caict.ac.cn/jpal/202109/t20210909_3585.html

155 Fuente: China News. https://www.chinanews.com/cj/2025/02-14/10368995.shtml

156 Fuente: Sina Tech. https://finance.sina.com.cn/tech/roll/2025-02-14/doc-inekmuic0855662.shtml

157 Fuente: Sina Tech. https://finance.sina.com.cn/tech/roll/2025-05-07/doc-inevtyis9418933.shtml?froms=ggmp

158 En mayo del 2025, frente al anuncio de Malasia que iba a desplegar 3,000 semiconductores de Huawei como parte de una iniciativa nacional de IA, el Departamento de Comercio de EE. UU. anunció que esto violaría los controles de exportación, a raíz de lo cual Malasia debió cancelar la operación.

Finalmente, a finales del 2025 Estados Unidos anunció que autorizará la venta de una cantidad pre-determinada de semiconductores H200, cerrando el debate de políticas [159]. Si bien esta generación no es la más avanzada de NVIDIA (como los Blackwell), los modelos fundacionales de frontera, como GPT-5 han sido entrenados bajo esta infraestructura. La decisión de Estados Unidos está fundamentada en prevenir a China desarrollar semiconductores de última generación que le permitirían asumir un liderazgo internacional en competencia con Estados Unidos.

La reacción de firmas chinas fue tan inmediata en términos de ordenar chips de NVIDIA, que la empresa debió requerir del fabricante TSMC, el aumento de capacidad adicional de manufactura. Sin embargo, esto determinó una nueva intervención del gobierno chino. A comienzos del 2026, el gobierno chino informó a empresas tecnológicas locales que deben suspender la compra de chips de NVIDIA H200 hasta que decida si las va a autorizar, y en qué condiciones. En particular, el gobierno quiere prevenir que dichas empresas realicen una compra acelerada para almacenar y suplir necesidades futuras [160].

Finalmente, el gobierno chino determinó que autorizará la importación de chips H200 bajo ciertas condiciones, como por ejemplo la adquisición por parte de laboratorios de desarrollo dentro de universidades [161]. Al mismo tiempo, el gobierno está explorando la posibilidad de requerir empresas locales de explorar la posibilidad de comprar una cantidad predeterminada de chips H200 en paralelo con semiconductores producidos localmente [162]. Esta directiva implica que el gobierno chino reconoce que la necesidad de proteger y estimular el desarrollo de la industria local es más importante que el acceso a productos del exterior. De todas maneras, la directiva no es precisa, dejando lugar a interpretaciones que permitan una implementación flexible. La dificultad en la implementación de la directiva reside en que los objetivos nacionales de alcanzar un liderazgo en IA y autoabastecimiento de semiconductores avanzados pueden resultar en un conflicto con el acceso a semiconductores H200 de NVIDIA en el corto plazo para construir modelos de frontera.

Infraestructura de cómputo de la IA

China opera una de las capacidades de cómputo más grandes del mundo (1590 EFLOPS en 2025)[8]. En el 2026, la PRC alberga 483 centros de datos con 3.8 GW dedicados a IA. La proyección al 2031 estima alcanzar 541 centros con una capacidad para IA de 10.3 GW [163].

El desarrollo de la infraestructura de centros de datos ha sido encarado bajo una estrategia de planificación centralizada denominada "East-Data-West-Computing" (*Dongshu Xisu*) formulada en el 2021, que busca desplazar los centros hacia provincias con excedentes energéticos y menores costos operativos, como Guizhou, Gansu y Mongolia Interior. En diciembre del 2021, el gobierno aprobó cuatro hubs nacionales de centros de datos en Mongolia, Guizhou, Gansu y Ningxia. El Information Park de China Telecom en Mongolia es considerado como el más grande del mundo, alcanzando 10 millones de pies cuadrados.

En febrero de 2022, cuatro regiones adicionales fueron agregadas al plan nacional en Beijing-Tianjin-Hebei, el delta del rio Yangtze, Chengdu-Chongqing, y Guangdong-Hong Kong-Macao. En estas zonas, las tarifas eléctricas descienden a 0.28 RMB/kWh, en comparación con los 0.65 RMB/kWh de ciudades como Shanghái, reduciendo el costo total de propiedad (TCO) hasta en un 30% [164]. En estos clústeres los constructores de centros de datos se benefician por concesiones incluyendo bajos costos

[159] Efrati, A. (2025). Trump to let NVIDIA sell advanced AI chips to China. The Information (12/9).
[160] Liu, Q. (2026). "China Tells Tech Companies to halt Nvidia H200 Chip Orders". The information (January 7).
[161] Liu, Q. (2026). "China restricts NVIDIA chip purchases under special circumstances". The Information (January 13).
[162] Los principales adquirentes de semiconductores avanzados son los fabricantes de servidores que, a su vez los proveen a desarrolladores de modelos.
[163] Gergs, L. (2026). Where is AI data center demand growth being driven? ABI Research (February 23).
[164] Mordor Intelligence (2025). China Data Center Server Market Size & Share Analysis - Industry Research Report, Disponible en: https://www.mordorintelligence.com/industry-reports/china-data-center-server-market

de terreno y exenciones impositivas. Esto no implica que operadores o proveedores de servicios no puedan construir centros en áreas diferentes a las designadas; el problema es que entonces no se beneficiarían de las exenciones impositivas y mejores costos [165].

El sector de hiperescaladores en China presenta una concentración similar al de otras economías avanzadas, pero con una mayor presencia de servicios gestionados por el estado y proveedores de servicios de la nube, liderados por Alibaba Cloud, Tencent Cloud, y Baidu Cloud. Alibaba ya ha desplegado siete centros en 2025 y ha anunciado la construcción de otros dos en Chengdu y Wuhan. En términos del mercado chino, Alibaba Cloud lidera con un 36% de cuota, seguido por Huawei Cloud (16%) y Tencent Cloud (9%) (ver cuadro 15).

Cuadro 15
China: cuota de mercado de proveedores de servicios de cloud
(En porcentajes)

Proveedor de Nube	Cuota de Mercado (Q3 2025)	Evolución Estratégica
Alibaba Cloud	36	Crecimiento de tres dígitos en ingresos de IA; despliegue de modelos Qwen
Huawei Cloud	16	Foco en IA industrial; lanzamiento del modelo predictivo Tianji 1.0
Tencent Cloud	9	Optimización de recursos ante escasez de chips; lanzamiento de HY 2.0
Otros (Baidu, China Telecom)	39	Crecimiento impulsado por la digitalización gubernamental y "nubes soberanas"

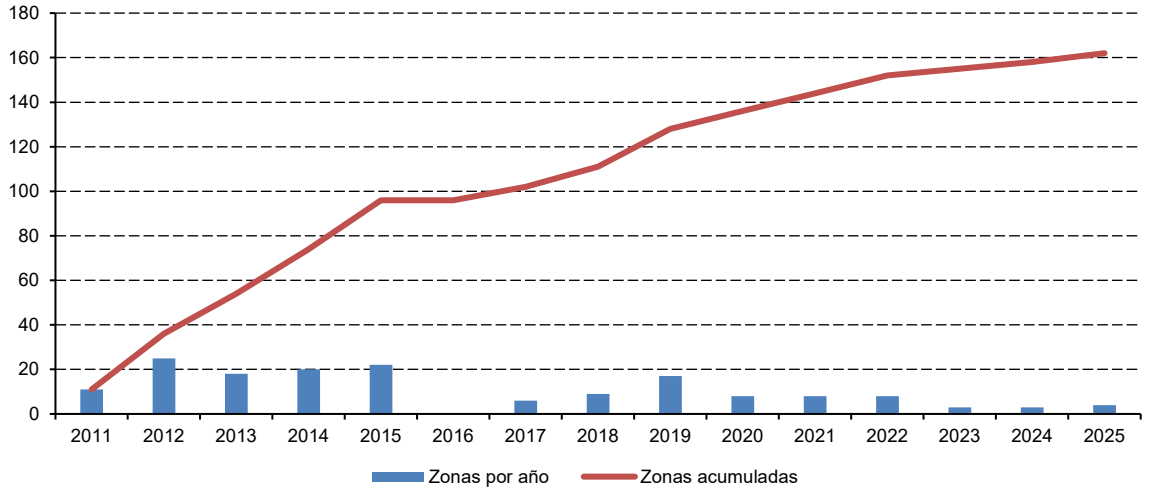
Fuente: Kr-Asia (2026). "China cloud market tops USD 13 billion" Disponible en: https://kr-asia.com/pulses/159929; OMDIA (2026). "Mainland China's cloud infrastructure market accelerates to ... "- (febrero 13). Disponible en: https://omdia.tech.informa.com/pr/2026/feb/mainland-chinas-cloud-infrastructure-market-accelerates-to-24percent-growth-in-q3-2025.

Los hiperescaladores chinos, a los que suman algunos estadounidenses que operan en *joint venture* con capital local, dominan el volumen de computación en la nube, operando 162 zonas en el país (ver gráfico 12).

Adicionalmente, el eslabón de equipamiento cuenta con centros neutrales (es decir independientes de todo operador) como, por ejemplo, VNET, GDS, Chindata, Range Technology y Baosight Software. Este tipo de proveedores son una respuesta a las necesidades de hiperescaladores en la medida de que pueden proveer servicio de manera rápida y costos más reducidos. Su modelo de negocio está basado en poder anticipar la demanda de hiperescaladores sobre la base de la adquisición de terrenos en localidades con alta disponibilidad de energía eléctrica y agua, buena conectividad, y demanda en base a contratos pre-establecidos.

[165] Thompson, D., Lenoir, T. and Lu, P. (2025). *Geopolitics of data centers: an AI showdown that will reshape the world*. S&P Global (December 2). Disponible en: https://www.spglobal.com/en/research-insights/special-reports/look-forward/data-center-frontiers/geopolitics-data-sovereignty-data-centersecurity?utm_source=newsletter&utm_medium=link&utm_campaign=corporates&utm_content=NAMM

Gráfico 12
China: zonas de proveedores de servicios de computación en la nube, 2011-2025
(En cantidad de zonas)



Fuente: Elaboración propia sobre la base de Telegeography.
Nota: Una zona cloud es definida como un centro de datos operando dentro de una región.

Una tendencia crítica en este eslabón es la emergencia de proveedores especializados de servicio en la nube que ofrecen acceso directo a GPUs locales ("GPU-as-a-Service"). Este modelo permite a las empresas mitigar el alto costo de capital inicial necesario para construir infraestructura propia, acelerando el ciclo de adopción desde la experimentación hacia la producción a gran escala[166].

Adquisición y preparación de bases de datos

El tercer eslabón de la cadena, centrado en la captura, etiquetado e integración de datos, ha evolucionado desde el procesamiento manual masivo hacia sistemas de etiquetado automatizado apoyados por IA. China posee una ventaja estructural en la disponibilidad de datos industriales y de comportamiento de usuarios, pero enfrenta límites en el acceso de datos públicos para el entrenamiento de modelos de razonamiento avanzado.

Empresas como Alibaba Cloud, Tencent y Baidu han integrado herramientas de preparación de datos en sus plataformas de IA. Por ejemplo, Alibaba ofrece iTAG, un servicio gestionado que utiliza pre-etiquetado inteligente para tareas multimodales, reduciendo el tiempo de desarrollo de modelos de visión y audio. Tencent ha desarrollado operaciones de etiquetado para mapas, comportamiento de usuarios, y recomendaciones. Asimismo, Tencent ofrece herramientas de etiquetado a partir de su servicio de la nube. Baidu opera su propio servicio de crowdsourcing y plataforma de etiquetado de texto, audio, imágenes y video, de manera integrada con su plataforma de IA.

Más allá de los servicios integrados de grandes operadores, existen empresas especializadas en etiquetado de datos. SO Development ofrece etiquetado multilenguaje, compilación especializada de datos y entrega de datos a constructores de modelos. Asimismo, actores especializados como Datatang y SenseTime proporcionan conjuntos de datos propietarios para sectores verticales como

[166] Tiger Brokers (2026). Mainland China's Cloud Infrastructure Services Market Hits $13.4 Billion in Q3 2025 with 24% Yearly Growth (February 10). Disponible en: https://www.itiger.com/news/1121904572

la conducción autónoma y la imagenología médica, donde la precisión del etiquetado es crítica para la seguridad operativa [167].

Ante el agotamiento de los datos públicos de alta calidad en Internet, el mercado chino está evolucionando hacia el desarrollo de "datos sintéticos". Nuevos startups están desarrollando simuladores para generar escenarios de entrenamiento en robótica y finanzas, permitiendo a los modelos aprender patrones complejos sin comprometer la privacidad de los datos reales, un requisito estricto bajo la Ley de Protección de Información Personal de China [168].

Modelos

China ha registrado una explosión en el desarrollo de modelos fundacionales, con más de 350 modelos certificados por los reguladores a mediados de 2025. En China, el mecanismo aplicado para reconocer a los modelos fundacionales y a los servicios de IA generativa disponibles para el público debe entenderse como un proceso de registro administrativo relacionado con el uso de las herramientas más que una certificación técnica de laboratorio [169]. La base jurídica principal es el Reglamento Provisional para la Gestión de los Servicios de Inteligencia Artificial Generativa [170], emitido conjuntamente por la *Cyberspace Administration of China* (CAC), la Comisión Nacional de Desarrollo y Reforma, el Ministerio de Educación, el Ministerio de Ciencia y Tecnología, el Ministerio de Industria y Tecnologías de la Información, el Ministerio de Seguridad Pública y la Administración Nacional de Radio y Televisión [171]. Este reglamento, vigente desde el 15 de agosto de 2023, establece que los proveedores de servicios de IA generativa deben someterse a reglas de seguridad, evaluación y gobernanza, y prevé expresamente seguridad, proteger los derechos de los ciudadanos, filing algorítmico, denuncias y supervisión. En ese sentido, la autoridad central (CAC) ha aclarado que los servicios de IA generativa que contengan atributos de opinión pública o capacidad de movilización social deben cumplir con el procedimiento de registro, modificación y cancelación de registro de algoritmos [172]. Entre enero y marzo de 2025, la propia CAC reportó 346 servicios generativos registrados a nivel nacional y 159 aplicaciones o funciones registradas a nivel local; mientras que, a cierre de 2025, el acumulado llegó a 748 servicios y 435 aplicaciones registradas [173].

Por otro lado, al registro se suma la norma técnica de la *National Information Security Standardization Technical Committee* (TC260), que estructuró los requisitos básicos de seguridad para servicios de IA generativa alrededor de cuatro bloques: la seguridad del corpus de entrenamiento (incluyendo su origen, contenido y etiquetado), la seguridad del modelo, las medidas de protección aplicadas al servicio y los procedimientos de verificación exigidos para acreditar el cumplimiento de esos requisitos [174].

La organización del proceso de registro de los modelos está basada en el cumplimiento con la Regulación sobre Evaluación de Seguridad de Servicios de Internet con Atributos de Opinión Pública o Capacidad de Movilización Social. Allí se establece que el proveedor debe realizar la evaluación (exante

---

[167] SO Development (2025). Top 10 Chinese Data-Collection companies. (October 10). Disponible en: https://so-development.org/top-10-chinese-data-collection-companies-2025/

[168] Chelidze Group (2026). Artificial intelligence market in China in 2025 (November 12). Disponible en: https://www.chelidze.group/en/post/ai-china-2025

[169] Fuente: Zou, M; Zhang, Lu. (2025). Navigating China's regulatory approach to generative artificial intelligence and large language models. https://www.cambridge.org/core/journals/cambridge-forum-on-ai-law-and-governance/article/navigating-chinas-regulatory-approach-to-generative-artificial-intelligence-and-large-language-models/969B2055997BF42DE693B7A1A1B4E8BA

[170] Fuente: Central Cyberspace Administration. *Reglamento provisional para la administración de servicios de inteligencia artificial generativa*. https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

[171] Central Cyberspace Administration. https://www.cac.gov.cn/2023-07/13/c_1690898326795531.htm

[172] Central Cyberspace Administration. https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm (artículo 17)

[173] Central Cyberspace Administration. https://www.cac.gov.cn/2026-01/09/c_1769688009588554.htm

[174] Comité Técnico Nacional de Normalización de la Ciberseguridad. https://www.tc260.org.cn/upload/2024-03-01/1709282398070082466.pdf

y expost), y que dicha evaluación debe revisar la legalidad del servicio, la efectividad de las medidas de seguridad y la eficacia del control de riesgos. La norma enumera aspectos específicos como responsable de seguridad, verificación de identidad de usuarios, conservación de registros y logs, prevención y manejo de información ilegal o dañina, protección de datos personales, mecanismos de denuncias, y capacidad de apoyo técnico a las autoridades [175].

En términos generales, la dinámica competitiva en este eslabón está determinada por una apuesta agresiva por el código abierto para erosionar la dominancia de los modelos propietarios desarrollados principalmente en EE. UU. De acuerdo con Nikkei, los modelos fundacionales chinos han capturado en noviembre del 2025 15% de cuota a nivel global, habiendo aumentado de 1% en un año [176].

Los "Seis Tigres" de la IA en China (Moonshot AI, Zhipu AI, MiniMax, Baichuan AI, StepFun y DeepSeek) han desafiado a los incumbentes (Baidu, Alibaba, Tencent) mediante innovaciones en eficiencia algorítmica. Un caso paradigmático es DeepSeek, cuya arquitectura y entrenamiento de bajo costo ($5.6 millones) ha demostrado que es posible alcanzar rendimientos de frontera sin las inversiones de firmas como OpenAI (ver cuadro 16).

Cuadro 16
China: principales modelos fundacionales

Modelo/empresa	Especialización	Desempeño
DeepSeek-R1	Razonamiento y Codificación	Rendimiento de nivel GPT-4 con 1/20 del costo de entrenamiento Rendimiento similar a GPT-4 con un coste de entrenamiento inferior a 6 millones de dólares
Qwen (Alibaba)	Ecosistema Abierto	700 millones de descargas en Hugging Face; adopción en Japón y SE Asia, superando incluso a Llama de Meta en integración comunitaria
Ernie 5.0 (Baidu)	Búsqueda y Agentes	2.4 billones de parámetros; integración nativa en el ecosistema de Baidu Mantiene el dominio en la comprensión profunda de la lengua y cultura china, integrado masivamente en su ecosistema de búsqueda y aplicaciones
Kimi (Moonshot)	Contexto Largo	Capacidad de procesar hasta 2 millones de caracteres en un solo prompt
GLM (Zhipu AI)	Ecosistema abierto y colaborativo	Modelo enfocado principalmente en programación
ABAB Series (Minimax)	Capacidad de creación de aplicaciones	M2 cuesta $1.20 por millón de tokens, comparado con $15 para Claude Sonnet

Fuente: The Efficiency Trap: China Won The Cost Curve, The US Owns The ..., https://www.techconstant.com/the-efficiency-trap-china-won-the-cost-curve-the-us-owns-the-cash-register/; Chinese AI Models Reportedly Hit ~15% Global Share in Nov. 2025, Fueled by DeepSeek Open-Source Push – TrendForce; Baidu integrates AI into search as rivals pursue superapp ambitions, https://amp.kr-asia.com/baidu-integrates-ai-into-search-as-rivals-pursue-superapp-ambitions; The 2025 China AI Index: Ranking the Most Powerful and Influential Models, https://www.athenasia.com/post/the-2025-china-ai-index-ranking-the-most-powerful-and-influential-models.

El lanzamiento de modelos abiertos de alta calidad ha reducido las barreras de entrada para desarrolladores chinos, permitiéndoles realizar un *fine-tuning* para aplicaciones industriales específicas sin incurrir en costos de licencia prohibitivos. Esta estrategia busca establecer a los modelos chinos como la infraestructura básica de la IA, exportando no solo tecnología sino también estándares de gobernanza. Los modelos de código abierto reducen el costo total de propiedad, permiten la adaptación lingüística y minimizan la fricción regulatoria permitiendo preservar datos locales dentro de las fronteras del país. Esto

[175] Central Cyberspace Administration. *Reglamento sobre la evaluación de la seguridad de los servicios de información en Internet con capacidad de movilización social o de opinión pública*. https://www.cac.gov.cn/2018-11/15/c_1123716072.htm.

[176] NEWS (2016). *Chinese AI Models Reportedly Hit ~15% Global Share in November 2025, Fueled by DeepSeek Open-Source Push (January 26).* Disponible en: https://www.trendforce.com/news/2026/01/26/news-chinese-ai-models-reportedly-hit-15-global-share-in-nov-2025-fueled-by-deepseek-open-source-push/

ha generado a China una ventaja comparativa en el desarrollo de cargas de trabajo en finanzas, logística y servicios públicos, especialmente en países como Singapur y Alemania, donde los modelos chinos están posicionados en cuarta posición en términos de compartición de tokens [177].

Software y herramientas

El quinto eslabón de la cadena de valor incluye las herramientas necesarias para la implementación de aplicaciones, donde se está produciendo una transición desde interfaces conversacionales hacia la "IA Agéntica" [178]. La transición hacia la "IA Agéntica" es más acelerada en China debido a la integración nativa de agentes en super-apps como WeChat (Tencent) y el ecosistema de productividad de Alibaba. Las plataformas de desarrollo de agentes están pasando de la demostración de capacidades a una fase de escalabilidad operativa. Los agentes de IA en China no solo conversan, sino que ejecutan flujos de trabajo autónomos, interactuando con APIs empresariales, sistemas ERP y dispositivos de hardware [179].

Hacia finales de 2025, el mercado de agentes en China está madurando hacia una fase de ingeniería operativa escalable. Alibaba Cloud, Huawei Cloud y Tencent Cloud han actualizado sus plataformas para incluir protocolos de contexto de modelos, orquestación de flujos y bases de conocimientos RAG (*Retrieval-Augmented Generation*) comercializables [180]. El valor se está desplazando desde el modelo base hacia la "orquestación", donde la capacidad de un agente para resolver problemas complejos de múltiples pasos define la ventaja competitiva.

Desarrollo e implementación de aplicaciones

El último eslabón involucra la implementación de IA para optimizar procesos de negocio. En China, este mercado está dominado por aplicaciones de IA Vertical, diseñadas para industrias específicas:

- **Manufactura:** China ha priorizado la integración de LLMs en equipos industriales, vehículos eléctricos y maquinaria. Empresas como TCL han desplegado agentes de red locales que utilizan razonamiento de modelos DeepSeek para detectar peligros ocultos con una eficiencia del 92% [181].
- **Salud:** Neusoft lidera soluciones de inteligencia de datos clínicos adoptadas por los 20 mejores hospitales de China, facilitando la conversión de hallazgos de investigación en aplicaciones prácticas para cuidados intensivos y enfermedades cardiovasculares [182].
- **Educación:** se registra un uso masivo de herramientas de tutoría personalizada, con una tasa de aprobación del 83% entre la población china respecto a los beneficios de la IA en la enseñanza [183].

---

[177] Wu, J. (2025). *China market news* (December 9). Disponible en: https://nai500.com/blog/2025/12/9-chinese-ai-stocks-riding-the-open-source-wave/

[178] Dudarenok, A. (2026). High Tech Companies Leading AI, Robotics, and EVs in.. (February 9)., Disponible en: https://chozan.co/chinese-high-tech/

[179] The State Council (2025). *China to unveil plans to integrate AI with manufacturing*, (November 4). Disponible en: https://english.www.gov.cn/news/202511/04/content_WS6909f081c6d00ca5f9a07504.html

[180] Tiger Brokers (2026). *Mainland China's Cloud Infrastructure Services Market Hits $13.4 Billion in Q3 2025 with 24% Yearly Growth* – (February 10). Disponible en: https://www.itiger.com/news/1121904572

[181] The State Council (2025). *China to unveil plans to integrate AI with manufacturing*, (November 4). Disponible en: https://english.www.gov.cn/news/202511/04/content_WS6909f081c6d00ca5f9a07504.html

[182] Neusoft (2025). *Neusoft Leads China's Healthcare Data Intelligence Solutions Industry with Dual Market Share Leadership,* (October 28). Disponible en: https://www.neusoft.com/neusoft-leads-chinas-healthcare-data-intelligence-solutions-industry-with-dual-market-share-leadership/

[183] The 2025 AI Index Report | Stanford HAI, Disponible en: https://hai.stanford.edu/ai-index/2025-ai-index-report

El papel de los servicios de telecomunicaciones en la cadena de valor de IA de China

Los operadores de telecomunicaciones chinos han adoptado una estrategia de diversificación para competir con los proveedores de nube tradicionales, posicionándose como la plataforma base para la entrega de servicios de IA soberana [184].

Cuadro 17
China: estrategia de IA de operadores de telecomunicaciones

Operador	Estrategia de IA	Inversión Clave
China Mobile	Estrategia "Pirámide de IA"	Triplicar potencia de cómputo para 2028; modelo Jiutian
China Telecom	Iniciativa "AI Flow"	Integración de IA en infraestructura de red; modelo TeleChat
China Unicom	CDSA (Cómputo y Aplicaciones Inteligentes)	Incremento del 28% en presupuesto de cómputo para 2025

Fuentes: Enki (2026). China Mobile AI Energy Strategy 2026: Inside the Blueprint, Disponible en: https://enkiai.com/china-mobile-ai-energy-strategy-2026-inside-the-blueprint. Matrix BCG (20926). What is Competitive Landscape of China Unicom Company? Disponible en: https://matrixbcg.com/blogs/competitors/chinaunicom.

China Mobile ya ha demostrado la viabilidad financiera de esta inversión. Sus ingresos por transformación digital alcanzaron los 156,900 millones de yuanes en la primera mitad de 2025, proporcionando el capital necesario para sostener el despliegue de redes 5G-Advanced nativas de IA.

2. Movimientos de integración vertical en la cadena de valor de la inteligencia artificial en China

Como se mencionó al inicio del capítulo, la organización industrial de la IA en China se caracteriza por un alto nivel de concentración en la capa de infraestructura y una fragmentación acentuada en la capa de aplicaciones, combinada con una integración vertical intensa.

En el eslabón de equipamiento, la posición de dominio de Huawei a través de su ecosistema Ascend es incuestionable. Cambricon y Moore Threads emergen como los especialistas independientes más valorados, con Cambricon alcanzando una valoración de 630,000 millones de yuanes. Estas empresas se han convertido en "habilitadores" de otros sectores, como la electrónica de consumo y los vehículos inteligentes, permitiendo que las marcas chinas compitan globalmente con hardware optimizado localmente.

En el eslabón de infraestructura, Alibaba Cloud mantiene el liderazgo con una cuota del 36%, apalancando su escala global y su capacidad de inversión que supera los 100,000 millones de yuanes anuales en IA e infraestructura de nube. Huawei Cloud y Tencent Cloud completan el oligopolio estable, mientras que ByteDance, a través de su motor Volcano, ha capturado una cuota del 14.8% en servicios de nube de IA, impulsada por la necesidad de procesar los inmensos volúmenes de video de sus plataformas Douyin y TikTok [185].

En resumen, la necesidad de maximizar la eficiencia energética y el rendimiento computacional, sumada a la escasez de chips avanzados, ha desatado un proceso de integración vertical agresivo, superior al observado en otros mercados [186]. En esta tendencia, el papel del gobierno chino de incentivo a la integración vertical es indiscutible.

---

[184] Enki (2026). *China Mobile's AI Data Center Energy Efficiency Strategy 2026.* Disponible en: https://enkiai.com/china-mobile-ai-energy-strategy-2026-inside-the-blueprint

[185] Technasia (2025) *Alibaba outpaces ByteDance, Tencent in China's AI cloud:* report (September 10). Disponible en: https://www.techinasia.com/news/alibaba-outpaces-bytedance-tencent-in-chinas-ai-cloud-report

[186] Zi, Ch. (2026). *Baidu integrates AI into search as rivals pursue superapp ambitions,* (February 11). Disponible en: https://amp.kr-asia.com/baidu-integrates-ai-into-search-as-rivals-pursue-superapp-ambitions

La integración vertical de hiperescaladores hacia el eslabón de equipamiento

Ante los altos márgenes de los proveedores de hardware y la incertidumbre del suministro, los hiperescaladores chinos están diseñando sus propios chips personalizados (ASICs).

- **Alibaba y T-Head:** Alibaba ha consolidado a T-Head como su vehículo principal de desarrollo de chips, lanzando procesadores de inferencia Hanguang 800 y aceleradores PPU Zhenwu que ya alimentan múltiples clústeres de 10,000 chips dentro de Alibaba Cloud [187]. En 2026, Alibaba planea escindir T-Head como una entidad independiente previo a una oferta pública, buscando apalancar su valoración [188].
- **Baidu y Kunlunxin:** Baidu ha integrado sus chips Kunlunxin en toda su infraestructura, alcanzando una sinergia que, de acuerdo con informes de la empresa, mejora el control de costos y la resiliencia de la cadena de suministro [189]. Los pedidos de Kunlunxin superaron las 200,000 unidades en 2025, convirtiendo a la división en un proveedor externo para otros sectores, incluyendo a los operadores de telecomunicaciones.

Esta integración vertical es una respuesta racional a los costos de transacción: al controlar el diseño del chip, los hiperescaladores eliminan la necesidad de adaptar sus modelos a arquitecturas genéricas de terceros, optimizando el consumo energético.

La integración de modelos fundacionales al suministro de datos

Los laboratorios de IA en China están integrándose hacia atrás para asegurar el acceso a datos de calidad y hacia delante para controlar la experiencia del usuario final.

- ByteDance y Tencent operan internamente sus propias granjas de etiquetado para entrenar sus modelos de recomendación y generación de video [190].
- Baidu ha integrado su modelo Ernie Bot directamente en su aplicación de búsqueda principal (700 millones de usuarios mensuales), eliminando la fricción de una aplicación independiente y capturando datos de interacción en tiempo real para el aprendizaje continuo del modelo [191].

3. Los puntos de presión en la cadena de valor de inteligencia artificial china

A pesar del impulso masivo de la IA, la cadena de valor en China enfrenta ciertos puntos de presión que cuestionan su rentabilidad a largo plazo y su estabilidad financiera.

El Dilema de la Monetización y la Brecha de Software

Existe una contradicción fundamental entre el avance técnico de los modelos chinos y su capacidad para generar ingresos. Mientras que en Estados Unidos el mercado de software se basa en modelos de suscripción SaaS consolidados, en China prevalece una cultura de "gratuidad" o de pago por hardware [192].

---

[187] Si, M. (2026) "Chinese AI chips gain huge traction in market", *Chinadaily.com.cn*, (January 30). Disponible en: https://www.chinadaily.com.cn/a/202601/30/WS697ccbf7a310d6866eb36b18.htmlacceso

[188] Young, D. (2025) "Alibaba eyes sky-high chip valuations with T-Head spinoff plan ", Bamboo Works (September 10). Disponible en: https://amp.kr-asia.com/baidu-integrates-ai-into-search-as-rivals-pursue-superapp-ambitions

[189] Zi, Ch. (2026). *Baidu integrates AI into search as rivals pursue superapp ambitions,* (February 11). Disponible en: https://amp.kr-asia.com/baidu-integrates-ai-into-search-as-rivals-pursue-superapp-ambitions

[190] SO Development (2025). *Top 10 Chinese Data-Collection Companies* (October 10) Disponible en: https://so-development.org/top-10-chinese-data-collection-companies-2025/

[191] Zi, Ch. (2026). *Baidu integrates AI into search as rivals pursue superapp ambitions,* (February 11). Disponible en: https://amp.kr-asia.com/baidu-integrates-ai-into-search-as-rivals-pursue-superapp-ambitions

[192] Bloomberg Intelligence (2025). China AI software Outlook. Disponible en: https://assets.bbhub.io/professional/sites/41/China-AI-Deep-Dive.pdf

- **La Brecha de Ingresos:** los 10 principales chatbots de China generaron menos de US$ 1 millón en ingresos en los primeros nueve meses de 2024, comparado con los US$ 160.5 millones de ChatGPT. Esta desproporción sugiere que las empresas de IA chinas podrían tardar al menos tres años en alcanzar la rentabilidad en un mercado fragmentado y saturado de herramientas gratuitas.
- **La Guerra de Precios de DeepSeek:** el lanzamiento de modelos con precios de API extremadamente bajos (un 92% inferiores a los precios de mercado previos) ha destruido la capacidad de fijación de precios de los startups, forzando una erosión de márgenes [193].

Sobreinversión en Infraestructura y capacidad ociosa

La inversión masiva en centros de datos, impulsada por subsidios estatales y competencia entre provincias, plantea el riesgo de sobrecapacidad. En Estados Unidos, la preocupación es similar, pero en China el riesgo se ve amplificado por el hecho de que muchas infraestructuras se construyen con hardware nacional que tiene un ciclo de vida corto y requiere actualizaciones constantes.

La construcción de centros de datos de IA consume una densidad de energía por metro cuadrado muy superior a los centros tradicionales, y aunque China ha aumentado su capacidad de generación, la infraestructura de transmisión aún presenta cuellos de botella para absorber las solicitudes de carga masivas de los nuevos clústeres de GPUs en el oeste del país [194].

B. La cadena de valor de la inteligencia artificial en Europa

La cadena de valor de la IA en Europa está transitando un momento crítico en el que el continente busca emerger de una posición de dependencia tecnológica de Estados Unidos hacia una autonomía estratégica construida alrededor de una infraestructura de centros de datos locales, el desarrollo de modelos soberanos, y el liderazgo en aplicaciones industriales verticales. En este marco, la inteligencia artificial no es concebida meramente como una aplicación secundaria de IT, sino como una tecnología de propósito general capaz de comprimir los ciclos de innovación y actuar como multiplicador de la productividad laboral en sectores donde Europa ya posee una ventaja comparativa, como la manufactura avanzada, la salud pública y la transición energética.

De todas maneras, este proceso no está exento de riesgos y puntos de presión. El contexto económico actual, subrayado por el Informe Draghi, señala una brecha de inversión masiva de aproximadamente 800,000 millones de euros anuales necesarios para que la Unión Europea mantenga su competitividad frente a los Estados Unidos y China [195].

Por otro lado, a diferencia de Estados Unidos y China, la cadena de valor europea de la IA está intrínsecamente ligada al marco regulatorio de la Ley de IA de la UE (EU AI Act), que actúa no solo como un mecanismo de protección ciudadana, sino como un vector de organización industrial que establece las barreras de entrada y las ventajas competitivas de las firmas locales [196].

---

[193] TechConstant (2026). The Efficiency Trap: China Won the Cost Curve, The US Owns The Cash Register. Disponible en: https://www.techconstant.com/the-efficiency-trap-china-won-the-cost-curve-the-us-owns-the-cash-register/

[194] Mordor Intelligence (2025). *China Data Center Server Market Size & Share Analysis - Industry Research Repor*t, Disponible en: https://www.mordorintelligence.com/industry-reports/china-data-center-server-market

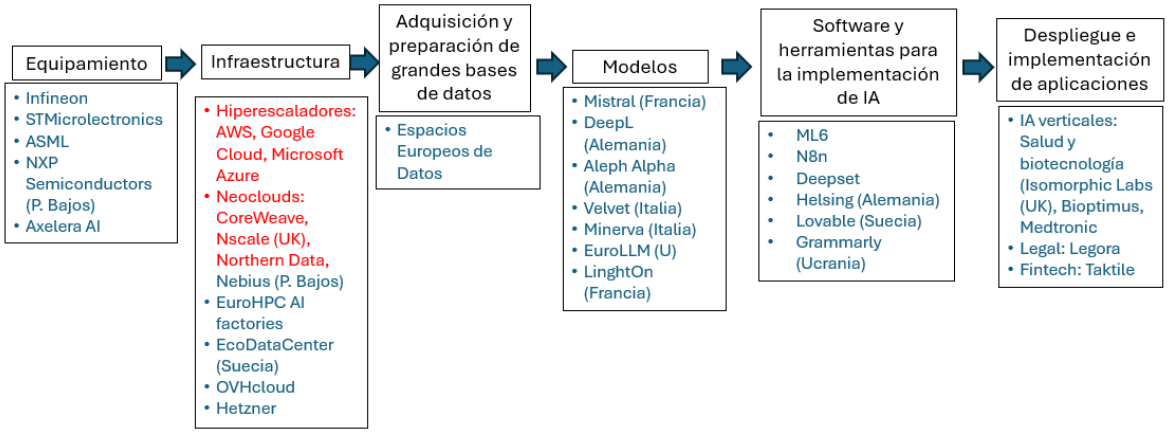
[195] El 'Informe Draghi' incide en la innovación tecnológica, la descarbonización y la seguridad para impulsar la competitividad de la UE – *Adigital* (2024), Disponible en: https://www.adigital.org/actualidad/el-informe-draghi-incide-en-la-innovacion-tecnologica-la-descarbonizacion-y-la-seguridad-para-impulsar-la-competitividad-de-la-ue/

[196] Comisión Europea (2024). *Ley de IA*. Disponible en: https://digital-strategy.ec.europa.eu/es/policies/regulatory-framework-ai

Estructura de la cadena de valor de la IA en Europa

La cadena de valor de la IA en Europa ya demuestra un avance en la mayoría de los eslabones, aunque las áreas principales de desarrollo se concentran en infraestructura (centros de datos), modelos fundacionales soberanos y herramientas para la implementación de IA y aplicaciones en mercados verticales (ver diagrama 18).

Diagrama 18
Europa: cadena de valor de IA



Fuente: Elaboración propia.
Nota: Azul indica desarrollo/capital local.

Equipamiento

En el diseño de semiconductores para centros de datos, Europa es mayoritariamente un importador de GPUs de NVIDIA. Por otro lado, si bien el componente clave dentro de este eslabón son los GPUs y TPUs (*Tensor Processing Units*), en el caso de chips especializados en eficiencia energética como los semiconductores de potencia de carburo de silicio (SiC) y nitruro de galio (GaN), firmas europeas como Infineon Technologies y STMicroelectronics mantienen posiciones dominantes en el sector automotriz e industrial [197].

Asimismo, potenciada por ASML (Países Bajos), Europa posee un liderazgo mundial en la fabricación de maquinaria de litografía ultravioleta extrema (EUV), una tecnología esencial para la producción de chips de 3nm y 5nm a escalas nanométricas [198]. De todas maneras, el continente enfrenta el desafío de trasladar ese dominio de maquinaria hacia una capacidad interna de diseño y fabricación de GPUs de propósito general. En este contexto, el *European Chips Act* prevé una inversión pública de 43,000 millones de euros con el objetivo de duplicar la cuota europea de producción mundial de semiconductores hasta alcanzar el 20 % en 2030 [199].

---

[197] Lotus Trading Academy (2026). *Las mejores acciones de semiconductores y chips*. Disponible en: https://expertosentrading.com/las-mejores-acciones-de-semiconductores-y-chips/

[198] The Associated Press (2026). "ASML made record $11.5 billion profit in 2025 thanks to AI-driven demand, plans to cut 1,700 jobs". *Newsday* (January 28). Disponible en: https://www.newsday.com/business/asml-semiconductor-profit-netherlands-jobs-artificial-intelligence-y27701

[199] García Retuerta, O; Garcia-Retuerta, D. (2026) *La cadena de valor de la inteligencia artificial: estrategias de autonomía para España*. Documento de Opinión IEEE Enero/12. Disponible en: https://www.defensa.gob.es/documents/2073105/3095923/cadena_de_valor_ia_2026_dieee003.pdf/9201d73c-38c0-efa7-023b-4bf8433c0f56?t=1765972424619

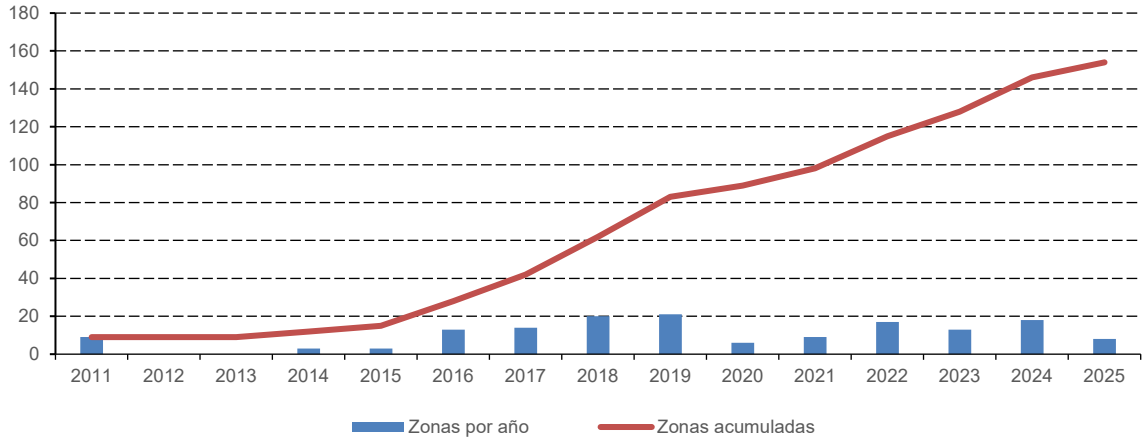
Finalmente, empresas como Infineon Technologies AG, NXP Semiconductors y Axelera AI están liderando el desarrollo de unidades de procesamiento de IA para inferencia eficiente en dispositivos de consumo y vehículos autónomos, capturando valor en nichos donde la eficiencia energética es más crítica que la potencia de procesamiento [200].

Infraestructura de cómputo de la IA

En Europa, este eslabón está experimentando un cambio estructural hacia la descentralización impulsada por la disponibilidad energética [201]. Históricamente, el continente controlaba 25% de la capacidad global de cómputo, aunque debido al avance estadounidense y chino, este porcentaje se redujo a 15% en el 2024. En el 2026, Europa contenía 1,879 centros de datos con una capacidad dedicada a la IA de 4.2 GW. Al 2031, el número de centros de datos está proyectado a alcanzar 2,303 y la capacidad dedicada a IA será de 9.7 GW [202]. La mayor parte de la capacidad de centros de cómputo europeos está concentrada en Frankfurt, Londres, Ámsterdam, Paris y Dublín, aunque existen nuevos desarrollos en Madrid y Helsinki. La infraestructura de IA está provista por tres categorías de proveedores: i) hiperescaladores, ii) neoclouds especializados, y iii) factorías de IA.

En la primera categoría, los proveedores globales (AWS, Google, Microsoft) dominan el volumen de computación en la nube, operando 154 zonas en el continente. En la actualidad, estos operadores enfrentan presiones para ofrecer zonas de disponibilidad soberanas que cumplan con las leyes locales de datos. Por ejemplo, AWS y Oracle ya se han asociado con el sector público para desplegar zonas en la nube en Frankfurt (ver gráfico 13).

Gráfico 13
Europa: zonas de proveedores de servicios de computación en la nube, 2011-2025
(En cantidad de zonas)



Fuente: Telegeography; análisis de los autores.
Nota: Una zona cloud es definida como un centro de datos operando dentro de una región.

[200] TBRC Business Research (2026). Artificial Intelligence (AI) In Chip Design Market Competition Analysis 2025: How Players Are Shaping Growth - *EIN Presswire* (January 7). Disponible en: https://www.einpresswire.com/article/880783563/artificial-intelligence-ai-in-chip-design-market-competition-analysis-2025-how-players-are-shaping-growth

[201] Jacamon, V., Dallard, J. Spenser, T. (2025). *Overcoming energy constraints is key to delivering on Europe's data center goals.* International Energy Agency (November 16). Disponible en: https://www.iea.org/commentaries/overcoming-energy-constraints-is-key-to-delivering-on-europe-s-data-centre-goals

[202] Gergs, L. (2026). *Where is AI data center demand growth being driven.* ABI Research (February 23).

Los Neo-clouds especializados incluyen a empresas como CoreWeave (originalmente de capital estadounidense, pero con inversiones crecientes en Europa), Nscale, Nebius, y Northern Data, que ofrecen acceso directo a clústeres masivos de GPUs con configuraciones optimizadas para IA, a menudo con costos un 50-70% inferiores a los de los hiperescaladores [203].

Finalmente, los supercomputadores públicos, desarrollados bajo la iniciativa de EuroHPC constituye una red de supercomputadores financiados por la UE, como LUMI (Finlandia) y Leonardo (Italia), diseñados para proporcionar capacidad de entrenamiento de modelos a startups y centros de investigación, reduciendo la dependencia de infraestructuras comerciales externas [204]. El Plan de Acción para el Continente de la IA de la Comisión Europea ha establecido el despliegue de una red de 19 "Factorías de IA" distribuidas en 17 estados miembros (ver cuadro 18).

Cuadro 18
Europa: lista de factorías de IA

País	Factoría	Enfoque principal
Alemania	JUPITER / JAIF	Exaescala, IA para investigación climática, ingeniería, manufactura, movilidad, tecnologías automotrices y ámbito académico
	HammerHAI	Ingeniería, manufactura, movilidad, tecnologías automotrices y ámbito académico
España	BSC AIF/ 1HealthAI	Salud pública y modelos lingüísticos, Clima y Agricultura, Finanzas, Asuntos Jurídicos, Energía, Comunicaciones y Medios de Comunicación
Finlandia	LUMI AIF	Entrenamiento de LLMs multilingües, Industrias manufactureras, Salud y ciencias de la vida, Tecnologías y redes de comunicación, Clima y meteorología, Tecnología de materiales, Lenguaje
Italia	Leonardo / IT4LIA	Agroalimentación, Ciencias de la Tierra, Ciberseguridad, Fabricación
País	Factoría	Enfoque principal
Suecia	MIMER	Ciencias de la vida, ciencia de materiales, sistemas autónomos y gaming
Portugal	1HealthAI	Salud, clima/agricultura, finanzas, energía
Italia	IT4LIA	IA industrial y soberana para manufactura, agroalimentos, ciberseguridad y ciencias de la Tierra
Francia	A12F	Plataforma multisectorial para investigación, innovación y servicios públicos
Grecia	Pharos	Salud, cultura e idioma, sostenibilidad
Bulgaria	Brain++	LLMs para lengua búlgara, observación espacial, robótica, agricultura, alimentos y salud
Francia	A12F	Energía, Aeroespacial, Salud, Defensa y seguridad, Ciencias de la Tierra, Continuo digital, Ciencias de los materiales, Robótica, Agricultura, Finanzas y Derecho, Educación (tecnología educativa), movilidad y humanidades
Eslovenia	SLAIF	Transición verde, salud y biotecnología, sociedad digital, ciencia
Lituania	LitAI	Infraestructura soberana de IA orientada a ciberseguridad, energía verde, industria inteligente y salud digital
Polonia	Gaia AI	LLMs, atención médica, espacio
P. Bajos	NLAIF	IA para datos sensibles en salud, movilidad, finanzas y agricultura
Rumania	RO AI	Adopción de IA por pymes y sector público

Fuente: Listado oficial de EuroHPC JU AI Factories 2025.

Estas instalaciones están equipadas con superordenadores optimizados para IA y actúan como ventanillas únicas para que PYMEs y startups accedan a potencia de cálculo y soporte técnico [205].

[203] Djurickovic, T. (2026). "The 25 tech companies that dominated European funding in 2025", TechEli (January 15). Disponible en: https://tech.eu/2026/01/15/the-25-tech-companies-that-dominated-european-funding-in-2025/

[204] European Commission (2025). *AI Factories*. Disponible en: https://digital-strategy.ec.europa.eu/en/policies/ai-factories

[205] Big Data Value Association (2025). *EuroHPC JU selects six new AI factories*. Disponible en: https://bdva.eu/news/eurohpc-ju-selects-six-new-ai-factories/

Es importante remarcar que el modelo de negocio y la propuesta de valor de cada uno de estos tres proveedores de infraestructura es diferente. Los hiperescaladores ofrecen una capacidad integrada del "AI stack", que va desde el poder de cómputo, las plataformas de IA y los servicios necesarios para su instalación en empresas. Desde ese punto de vista, el objetivo de estos es acelerar la adopción de IA en empresas, capturando ingresos recurrentes por el uso de plataformas de IA. En general, los adoptantes incluyen startups digitales nativos, grandes empresas y Pymes. Los neo-clouds ofrecen GPUs como servicio, lo que implica que el modelo de negocio está basado en la renta de infraestructura, y su rentabilidad considera el precio de alquiler de GPU por hora. En general, los clientes de neo-clouds son desarrolladores de modelos y aplicaciones de IA y no empresas usuarias. Las factorías de IA tienen como propósito la creación de capacidad soberana bajo el control estratégico de cada país. El modelo de negocio incluye la oferta de capacidad de cómputo para el entrenamiento de modelos, aunque dado su objetivo la oferta puede ser subsidiada.

Adquisición y preparación de bases de datos

En Europa, el eslabón de captura y etiquetado de datos está fuertemente condicionada por el marco del Reglamento General de Protección de Datos, lo que ha incentivado el desarrollo de mercados de datos industriales y el uso de datos sintéticos [206].

Debido al inminente agotamiento de los datos públicos de alta calidad, los desarrolladores europeos están recurriendo a los "Espacios Europeos de Datos" (*European Data Spaces*), repositorios únicos de datos que permiten el intercambio seguro de conjuntos de datos en sectores como la salud, la movilidad y la energía para entrenar modelos especializados [207]. Los "Espacios Europeos de Datos" están enmarcados en el proceso de integración de bases de datos o *Data Lakes*. En esa línea, conviene precisar que la política europea no promueve un único *lago de datos* centralizado, sino que impulsa una red de espacios de datos comunes, interoperables y federados, que se encuentra organizada por sectores o dominios. La base política de este modelo nace en la Estrategia Europea de Datos de 2020 [208], que planteó la creación de un mercado único de datos en Europa mediante espacios comunes en salud, agricultura, manufactura, energía, movilidad, finanzas, administración pública, capacidades, ciencia abierta y pacto verde. El enfoque consiste en crear infraestructuras [209] y marcos de gobernanza comunes [210] que permitan que distintos actores compartan, reutilicen y combinen datos bajo reglas homogéneas de acceso, seguridad e interoperabilidad.

La construcción de estos espacios se apoya en una arquitectura política y regulatoria que comprende varios aspectos:

- En el plano jurídico, la Data Governance Act [211] creó reglas para la reutilización de ciertos datos protegidos del sector público, para los servicios de intermediación de datos y para el llamado *data altruism* [212], además de establecer la European Data Innovation Board.
- La Data Act añadió reglas para ampliar el acceso y uso de datos, sobre todo datos industriales y de dispositivos conectados.

---

[206] Lagarde, C. (2025). *The transformative power of AI: Europe's moment to act.* European Central Bank (November 24). Disponible en: https://www.ecb.europa.eu/press/key/date/2025/html/ecb.sp251124_1~c239fb4a7f.en.html

[207] Comisión Europea (2025). *Enfoque europeo de la inteligencia artificial* Disponible en: https://digital-strategy.ec.europa.eu /es/policies/european-approach-artificial-intelligence

[208] Fuente: Comisión Europea. Estrategia Europea de Datos. https://digital-strategy.ec.europa.eu/en/policies/strategy-data

[209] Fuente: Comisión Europea. Centro de soporte de Data Spaces. https://digital-strategy.ec.europa.eu/en/news/data-spaces-support-centre

[210] Fuente: EUR-Lex. Gobernanza de datos europea. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=legissum%3A4604122

[211] Fuente: EUR-Lex. https://eur-lex.europa.eu/EN/legal-content/glossary/data-governance.html

[212] *Data altruism* es la compartición voluntaria de datos personales o no personales por individuos u organizaciones motivados por el bien común, como ser servicios de investigación científica, Cuidado de salud, o protección del medio ambiente. El concepto, promovido por el *Data Governance Act*, busca construir la compartición de datos para promover innovación y beneficios sociales.

- En el plano técnico y organizacional, la Comisión Europea financia el Data Spaces Support Centre (DSSC) para definir requisitos comunes, buenas prácticas, glosarios, kits de arranque y una arquitectura de referencia, y financia también Simpl [213] (un middleware abierto) para dar interoperabilidad entre infraestructuras cloud-to-edge y entre distintos espacios de datos.

En ese sentido, los espacios europeos funcionan como ecosistemas federados con reglas comunes, catálogos, servicios de acceso, control de permisos y capas de interoperabilidad. En la práctica, cada espacio se construye de manera sectorial y con una combinación distinta de norma, gobernanza y casos de uso.

Por su parte, la relación con el sector privado es fundamental, porque la política europea no concibe los espacios de datos como una infraestructura exclusivamente pública. La propia Comisión señala que los espacios están abiertos a la participación de empresas, administraciones y personas. Así, los titulares de datos pueden ponerlos a disposición gratuita o a cambio de una compensación. Esto implica que empresas, operadores de infraestructura, desarrolladores de software, proveedores de cloud, startups, universidades y entidades públicas pueden participar como proveedores de datos, usuarios, intermediarios o desarrolladores de servicios. En el caso del *Language Data Space* [214], esta relación con el sector privado es aún más explícita. La Comisión definió ese espacio como una plataforma y marketplace para el intercambio de datos, modelos, herramientas y servicios lingüísticos [215], y organizó talleres específicos con sectores como noticias, radiodifusión, publicidad, edición, tecnología lingüística y telecomunicaciones [216]. Además, la Comisión aclaró que la plataforma del *Language Data Space* apoya el intercambio, pero no aloja los datos por sí misma, lo que refuerza el carácter federado y comercializable del sistema.

En lo que respecta específicamente a editoriales, periódicos, revistas y otros titulares privados de contenido, la política europea no crea un derecho automático del Estado o de los desarrolladores de IA a absorber esos contenidos dentro de los espacios de datos. La relación está mediada por el régimen de propiedad intelectual [217] y, en particular, por la Directiva europea sobre derechos de autor en el mercado único digital [218]. Esa directiva permite la minería de texto y datos sobre materiales accesibles legalmente, pero en el caso general del artículo 4 deja claro que esa excepción opera solo si los titulares de los derechos no han reservado expresamente su uso. Además, la propia directiva prevé que dicha reserva pueda hacerse mediante medios legibles por máquina cuando el contenido está disponible en línea. Esto significa que, para actores privados como editoriales y revistas, el enfoque europeo tiende más hacia licencias, marketplaces, acuerdos de compartición y reglas que hacia una apropiación obligatoria del contenido.

La iniciativa incluye el despliegue de 14 espacios en agricultura, aspectos culturales, energía, finanzas, sostenibilidad, salud, manufactura, lenguaje, medios de comunicación, movilidad, administración pública, investigación e innovación, habilidades, y turismo.

Modelos

Los modelos fundacionales son clasificados bajo la Ley Europea de IA como modelos de IA de uso general desde agosto de 2025. Esto les impone obligaciones específicas de transparencia sobre los datos de entrenamiento y la mitigación de riesgos sistémicos [219].

---

213 Fuente: Comisión Europea: Simpl. https://digital-strategy.ec.europa.eu/en/policies/simpl
214 Fuente: Comisión Europea. https://digital-strategy.ec.europa.eu/en/news/public-access-european-language-data
215 Fuente: Comisión Europea. https://digital-strategy.ec.europa.eu/en/funding/language-data-space-call-tenders
216 Fuente: Comisión Europea. https://digital-strategy.ec.europa.eu/en/library/language-data-space-workshop-report
217 Fuente: EUR-Lex. https://eur-lex.europa.eu/legal-content/EN-RO/TXT/?from=EN&qid=1699380467196&uri=CELEX%3A32019L0790
218 Fuente: EUR-Lex. https://eur-lex.europa.eu/eli/dir/2019/790/oj
219 Comisión Europea (2025). Ley de IA. Disponible en: https://digital-strategy.ec.europa.eu/es/policies/regulatory-framework-ai

En este eslabón, Europa ha logrado posicionar a Mistral AI como un líder global, ofreciendo modelos que maximizan el rendimiento por parámetro y permiten a empresas ejecutar IA en sus propias infraestructuras privadas [220]. La propuesta de valor de Mistral AI se basa en tres pilares:

- **Eficiencia de parámetros:** El modelo Mistral 7B demuestra que modelos pequeños bien diseñados pueden equiparar el desempeño de modelos mucho más grandes, reduciendo así el gasto de infraestructura y el consumo energético.
- **Soberanía y Despliegue Privado:** Al ofrecer modelos con pesos abiertos, Mistral permite a las empresas implementar aplicaciones de IA dentro de sus propios entornos, garantizando el cumplimiento del Reglamento General de Protección de Datos, evitando el "lock-in" de proveedores de nube no europeos [221].
- **Multilingüismo Nativo:** A diferencia de los modelos estadounidenses optimizados para el inglés, los modelos de Mistral permiten operar en las 24 lenguas oficiales de la UE, facilitando la adopción de IA en la administración pública y sectores culturales locales.

Mistral no es el único modelo europeo. Entre otras ofertas, se cuentan Aleph Alpha (Alemania), Velvet AI (Italia), Minerva (Italia), LightOn (Francia), y EuroLLM-9B (Paneuropeo). La comparación entre ellos muestra que el énfasis europeo se centra en modelos abiertos, la sostenibilidad en términos del consumo energético, y el apoyo pluri-lingüístico (ver cuadro 19).

Cuadro 19
Europa: modelos fundacionales de IA

Empresa	Modelo	Número de parámetros	Tokens de entrenamiento	Lenguajes	Fortalezas
Mistral AI (Francia)	l Small 3.1	24B	128k tokens	EN, Multimodal (imágenes)	Eficiencia, multimodal
	Mixtral 8x7B	56B	32k tokens	Multilingüe	Líder en categoría
	Magistral Small	24B	40k tokens	EN, Multilingüe	Razonamiento
	Devstral Small	24B	128k tokens	Coding, EN	Agentic, programación
Aleph Alpha (Alemania)	Luminous	Varios	---	Cinco lenguajes de UE	Representación semántica
	Pharia-1-LLM-7b-Control	7 mil millones	---	Alemán, Frances, español	Código Abierto, corpus multilenguaje
Velvet (Italia)	Velvet 14B	14B	128k tokens	IT/DE/FR/ES/PT/EN	Sostenibilidad, multilenguaje
	Velvet 2B	2B	32k tokens	IT/EN	Eficiente
Minerva (Italia)	Minerva 7B	7.4 mil millones	2.5 trillones	Italiano/Ingles	Código abierto
EuroLLM-9B (EU)	EuroLLM-9B	9 mil millones	4 trillones	24 lenguas de UE + 11 adicionales	Código abierto
	EuroLLM-1.7B	1.7 mil millones	---	Multilenguaje	Código Abierto, multilenguaje
LightOn (Francia)	Pagnol, RITA, Mamboutai	---	---	---	Código abierto
	Reason-ModernColBERT	---	---	---	Específico a vertical
	BioClinical ModernBERT	---	---	---	Específico a biomedicina

Fuente: Razzaq, A. (2025). "Europe's Top AI Models of 2025: Multilingual, Open, and Enterprise- Ready".

[220] *Mistral AI Models 2025: European AI Excellence Guide for Developers & Researchers*, Disponible en: https://local-ai-zone.github.io/brands/mistral-ai-european-excellence-guide-2025.html

[221] Razzaq, A. (2025). "Europe's Top AI Models of 2025: Multilingual, Open, and Enterprise- Ready*", MarkTechPost*, (August 25), Disponible en: https://www.marktechpost.com/2025/08/15/europes-top-ai-models-of-2025-multilingual-open-and-enterprise-ready/

Software y herramientas

El desarrollo del eslabón de las plataformas de software y herramientas de IA en Europa confirma la tesis de George Stigler, citada en el capítulo I del estudio, en el sentido de que la división del trabajo está limitada por la extensión del mercado. En la fase actual de desarrollo de la IA en Europa puede verificarse un proceso de desintegración vertical donde la capa de chatbots y agentics se desacopla de la infraestructura de computación para la IA. Al mismo tiempo, esta especialización genera asimetrías de mercado significativas. Mientras que las interfaces desarrolladas por jugadores globales muestran una elevada penetración en el tejido productivo europeo, el ecosistema europeo ha elegido responder mediante el despliegue de opciones de código abierto y soluciones locales guiadas por el principio de la soberanía lingüística [222]. De acuerdo al razonamiento europeo, la dependencia de interfaces de usuario propietarias controladas por laboratorios de IA extranjeros expone a las corporaciones locales a una desventaja competitiva importante. Esto es particularmente crítico cuando los modelos de lenguaje o interfaces conversacionales actúan como el canal primario de interacción con el cliente final. Para mitigar esta vulnerabilidad, Europa fomenta consorcios diseñados bajo principios de soberanía tecnológica y multilingüismo. Es así como el proyecto OpenEuroLLM, un consorcio de veinte instituciones de investigación, empresas y centros de la EuroHPC (European High Performance Computing Joint Undertaking) enfocado en el desarrollo de modelos fundacionales abiertos y transparentes tiene implicancias para el desarrollo local del eslabón de software y herramientas.

Esta reconfiguración del mercado de software europeo se evidencia en la proliferación de soluciones nativas diseñadas para la ejecución autónoma de procesos. La empresa europea ERP.net, por ejemplo, ha introducido Operator.net, una plataforma que despliega agentes de IA directamente insertos en el núcleo del ERP y CRM corporativo para automatizar flujos de facturación, logística y soporte técnico a partir de comandos verbales [223]. Asimismo, plataformas regionales de automatización como eZintegrations y su módulo Goldfinch AI operan como capas neutrales de integración agéntica que exponen flujos empresariales complejos como puntos finales del Model Context Protocol, facilitando que modelos externos interactúen de forma segura con la infraestructura operativa interna [224]. Esta tendencia se complementa con la integración de agentes LLM-agnósticos en plataformas de código abierto como Odoo 19, que automatizan desde la asignación de leads comerciales hasta la emisión de recordatorios de cobro [225].

El desarrollo de software y herramientas en Europa no depende exclusivamente de protocolos corporativos privados si se aspira a garantizar la soberanía digital. Por este motivo, la Comisión Europea ha financiado y promovido Simpl, un middleware inteligente, seguro y de código abierto diseñado para viabilizar la interoperabilidad técnica y el acceso fluido a los Espacios Comunes Europeos de Datos, detallados arriba [226]. Simpl actúa como el motor de orquestación que traduce las directrices teóricas de soberanía y confianza de la Unión Europea en una arquitectura operativa real para dar soporte a iniciativas estratégicas y sectores específicos locales. Entre estas instancias destacan el Espacio de Datos de Contratación Pública (PPDS), el Espacio de Datos de Salud de la Unión Europea (EHDS2), el Espacio de Datos Lingüísticos (LDS), la Nube Europea de la Ciencia Abierta (EOSC), el proyecto de modelado climático

---

[222] AMS Silo (AI) 2025). *Europe's Leading AI Companies and Research Institutions Combine Their Forces to Develop Next-generation Open LLMs* (February 3). Disponible en: https://www.amd.com/en/blogs/2025/open-llms-for-transparent-ai-in-europe.html

[223] Nkosi, M. (2026). "ERP.net Launches an Integrated AI Agentic Platform for Enterprise Resource Planning", *IT News Africa.com* (May 13). Disponible en: https://www.itnewsafrica.com/2026/05/erp-net-launches-an-integrated-ai-agentic-platform-for-enterprise-resource-planning/

[224] Wilson, J. (2026). "Agentic AI Platform Comparison: Top 5 Enterprise Tools 2026" *Zintegrations* (March 24) Disponible en: https://ezintegrations.ai/agentic-ai-platform-comparison/

[225] Simsek, H. (2026). "Top 10 Agentic AI ERP Systems & 6 Solutions" *AIMultiple* (March 5), Disponible en: https://aimultiple.com/agentic-ai-erp

[226] Ferrer, Ana (2026). *Simpl Programme - CoE-DSC*, Disponible en: https://coe-dsc.nl/wp-content/uploads/2026/04/Ana-Ferer-180326_Simpl_DataSharingFestival.pdf

Destination Earth (DestinE) y el Espacio de Datos de Comunidades Inteligentes (SCDS). Esta estructura de middleware federado reduce de manera sistemática los costos de transacción para las empresas europeas, permitiéndoles participar en ecosistemas de datos compartidos bajo reglas contractuales transparentes, automatizadas e inmunes a la injerencia de legislaciones de terceros países [227].

En segundo lugar, para contrarrestar el poder competitivo de los proveedores de infraestructura física, el eslabón en Europa registra un proceso de alianzas corporativas e integraciones de carácter horizontal y vertical. Este fenómeno de consolidación sectorial se manifiesta en las estrategias de los principales actores europeos de software de IA. Por ejemplo, Mistral AI, el laboratorio de IA, ha consolidado su posicionamiento mediante un modelo híbrido de código abierto modificado (*Open Core*) [228]. Sobre esta base, Mistral ha establecido alianzas estratégicas con firmas de consultoría e integración de sistemas como Accenture [229] y Reply [230] para desplegar modelos personalizados en las infraestructuras de nube de corporaciones.

Simultáneamente, la intensidad de capital requerida para entrenar de forma competitiva modelos fundacionales de frontera ha forzado una recomposición de los activos dentro de la región. Aleph Alpha se apartó del desarrollo exclusivo de LLMs para especializarse en PhariaAI, una suite de software de orquestación y gobernanza empresarial que actúa como un sistema operativo para entornos altamente regulados [231]. Este viraje facilitó su adquisición del laboratorio canadiense Cohere, creando un bloque transatlántico apoyado por la Alianza de Tecnología Soberana Canadá-Alemania. Asimismo, STACKIT ha consolidado su posición en el mercado de seguridad mediante una alianza estratégica con la firma estadounidense CrowdStrike para ofrecer la plataforma de ciberseguridad nativa Falcon provista directamente desde centros de datos europeos [232], combinada con una asociación de defensa digital con la Oficina Federal de Seguridad de la Información de Alemania (BSI) [233].

El despliegue operativo del marco regulatorio de la Ley de Inteligencia Artificial de la Unión Europea (EU AI Act) condiciona la posición competitiva de las empresas de software de aplicación europeas en comparación con sus rivales de Estados Unidos y China [234]. Al categorizar los sistemas de IA en una escala de riesgo de cuatro niveles —inaceptable, alto, limitado y mínimo—, la ley regula con especial énfasis a aquellas soluciones integradas en infraestructuras críticas, gestión de recursos humanos, evaluación crediticia y filtrado de datos sensibles, catalogándolas como de "alto riesgo" [235]. Este esquema introduce una barrera no arancelaria significativa debido al incremento en los costos regulatorios necesarios para la introducción de un producto extranjero en el mercado europeo.

---

[227] European Commission (2026). *Simpl: Cloud-to-edge federations empowering EU data spaces*, Disponible en: https://digital-strategy.ec.europa.eu/en/policies/simpl

[228] Steinschaden, J. (2026). "Mistral CEO Warns Europe Could Become US "Vassal State" Within 2 Years", *Trending Topics*, Disponible en: https://www.trendingtopics.eu/mistral-ceo-warns-europe-could-become-us-vassal-state-within-2-years/

[229] Consultancy.eu (2026). *Accenture and Mistral AI join forces for sovereign AI in Europe*. March 2. Disponible en: https://www.consultancy.eu/news/13286/accenture-and-mistral-ai-join-forces-for-sovereign-ai-in-europe

[230] Business Wire (2026). *Reply Announces a partnership With Mistral AI to Develop Sovereign and Enterprise-grade Artificial Intelligence Solutions* | Business Wire (March 18). Disponible en: https://via.ritzau.dk/pressemeddelelse/14840053/reply-announces-a-partnership-with-mistral-ai-to-develop-sovereign-and-enterprise-grade-artificial-intelligence-solutions?publisherId=90456&lang=en

[231] Patience, N. (2026). "Cohere Acquires Aleph Alpha: Deal Born of Sovereignty, Necessity." Futurum (April 27). Disponible en: https://futurumgroup.com/insights/cohere-acquires-aleph-alpha-a-deal-born-of-sovereignty-necessity/

[232] Crowdstrike (2026). *CrowdStrike and Schwarz Digits Partner to Deliver AI-Native and Sovereign Cybersecurity Platform on STACKIT*, (March 5). Disponible en: https://ir.crowdstrike.com/news-releases/news-release-details/crowdstrike-and-schwarz-digits-partner-deliver-ai-native-and/

[233] Schwarzdigits (2026). *Digital sovereignty: BSI Schwarz Digits strategic partnership* (February 13). Disponible en: https://schwarz-digits.de/en/presse/archive/2026/digital-sovereignty-bsi-and-schwarz-digits-enter-into-strategic-partnership

[234] Elad, B. (2026). "EU AI Act Compliance Cost Statistics 2026: Key Trends Now". *SQ Magazine* (April 10). Disponible en: https://sqmagazine.co.uk/eu-ai-act-compliance-cost-statistics/

[235] Workstreet (2026). "EU AI Act Compliance: What US SaaS Companies Need to Know" (May 13). Disponible en: https://www.workstreet.com/blog/eu-ai-act-compliance

Para las pequeñas y medianas empresas tecnológicas europeas, la certificación técnica con las obligaciones de los sistemas de alto riesgo requiere esfuerzos organizativos y financieros considerables. El cumplimiento del Artículo 10 (gobernanza y representatividad de datos de entrenamiento para mitigar sesgos), el Artículo 12 (registro inalterable de eventos y *logging* de operaciones) y las directrices de supervisión humana y ciberseguridad demanda reestructuraciones internas, desviando inversión de capital hacia áreas regulatorias, incrementando el alto costo de la conformidad legal y técnica. Estas asimetrías de costos regulatorios se traducen en barreras geográficas y de escala para las empresas europeas. Mientras que las grandes corporaciones multinacionales disfrutan de economías de escala que les permiten diluir estos costos fijos de cumplimiento normativo entre múltiples sistemas de IA, los startups europeos afrontan un costo de oportunidad severo al restar recursos de ingeniería a sus procesos de innovación básica. Asimismo, la implementación de sistemas de gestión de calidad duplica la frecuencia de las auditorías internas de desarrollo en comparación con los esquemas previos a la vigencia del reglamento.

Es por ello por lo que, como contrapartida al objetivo de soberanía, el diseño regulatorio europeo genera un entorno contra-productivo para los desarrolladores y los proveedores de financiamiento europeos. Esta situación fuerza a las empresas europeas a enfocar sus esfuerzos en nichos altamente específicos de menor riesgo normativo, en lugar de escalar soluciones transversales de alta monetización. Así, mientras que los startups de software en Estados Unidos aceleran el despliegue de copilotos y herramientas automatizadas con dinámicas ágiles, las plataformas y desarrolladores europeos enfrentan un ecosistema fragmentado donde el cumplimiento normativo debe asumirse como un factor de diseño de ingeniería desde la fase preliminar de cualquier proyecto de software.

En conclusión, el análisis del eslabón de software y herramientas de la cadena de valor de la IA en Europa revela que este se encuentra en una encrucijada estratégica caracterizada por la tensión entre la búsqueda de autonomía tecnológica y la rigidez de su marco regulatorio. Desde la perspectiva de la organización industrial, el modelo organizativo de baja intensidad de capital ofrece una flexibilidad financiera indispensable para los startups locales, permitiéndoles competir mediante el uso intensivo de talento altamente cualificado e innovación en propiedad intelectual. Por otro lado, la sostenibilidad de este enfoque se ve amenazada por las altas barreras de entrada y los costos de transacción impuestos por la Ley de IA, los cuales limitan la escalabilidad de los nuevos entrantes frente a las economías de escala de los incumbentes estadounidenses. Iniciativas institucionales como Simpl y consorcios como OpenEuroLLM representan esfuerzos sistemáticos para crear un entorno de confianza capaz de reducir los riesgos de dependencia tecnológica. Asimismo, las alianzas estratégicas que integran el desarrollo de modelos de código abierto modificado con arquitecturas de nube soberana certificada —como la asociación entre Mistral AI, Cohere-PhariaAI y la plataforma STACKIT del Schwarz Group— demuestran que la soberanía digital de la región depende de la capacidad de articular de forma coherente la capa de software con los entornos de ejecución locales.

Desarrollo e implementación de aplicaciones

Este eslabón involucra la implementación de soluciones de IA para optimizar procesos de negocio. Europa se diferencia por su enfoque en la "IA Vertical", es decir, aplicaciones entrenadas en la ontología y regulaciones de industrias específicas como la salud, la jurídica y la manufactura [236].

El valor de la IA en Europa está siendo capturado en la capa de aplicaciones verticales especializadas. Por ejemplo, en el caso de IA Industrial y Manufacturera, Siemens ha liderado con su "Industrial Copilot", una herramienta de IA diseñada para generar automáticamente código de ingeniería, permitiendo que las máquinas se autoprogramen y optimicen en función de los datos de

[236] McKinsey & Co. (2025). *Accelerating Europe's AI adoption: The role of sovereign AI*, (December 19). Disponible en: acceso: https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/accelerating-europes-ai-adoption-the-role-of-sovereign-ai

producción en tiempo real. De la misma manera, en el sector automotriz, Bosch está integrando IA para la conducción automatizada en series de producción masiva, apalancando su ecosistema "ctrlX" para el control inteligente de maquinaria.

Europa posee también una ventaja competitiva en el entrenamiento de modelos para biotecnología gracias a sus sistemas de salud universales que generan vastos conjuntos de datos anonimizados [237]. Isomorphic Labs (basado en el Reino Unido), una escisión de DeepMind, aplica la IA para transformar el descubrimiento de fármacos [238]. De manera similar, Bioptimus (Francia) es un startup que desarrolla modelos fundacionales específicos para biología [239]. Finalmente, Medtronic (Irlanda) ha integrado el módulo "GI Genius", el primer sistema de endoscopia asistido por IA para la detección de pólipos en tiempo real, utilizando hardware de Edge AI de NVIDIA [240].

Europa también ha crecido significativamente en el segmento jurídico y financiero. Legora (Suecia) se enfoca en la automatización de flujos de trabajo de propiedad intelectual, mientras que, en el sector financiero, Taktile (Alemania) proporciona plataformas de decisión automatizadas para la suscripción de crédito en bancos, integrando el cumplimiento de condiciones regulatorias directamente dentro del algoritmo.

1. Movimientos de integración vertical en la cadena de valor de la inteligencia artificial en Europa

De la misma manera que en las cadenas estadounidense y china, las empresas activas en Europa despliegan estrategias de integración vertical tendiendo a asumir control de eslabones clave, o mitigar el riesgo político.

Para mitigar el riesgo de ser excluidos de los sectores regulados en Europa, los hiperescaladores estadounidenses están ejecutando una estrategia de "soberanía por asociación".

- **Microsoft y Mistral:** Microsoft invirtió 15 millones de euros en Mistral AI para distribuir sus modelos a través de la plataforma Azure, reconociendo que los clientes europeos demandan opciones locales para cargas de trabajo que requieran adaptación al entorno local [241].
- **Nubes soberanas locales:** la alianza entre Aleph Alpha y STACKIT (el proveedor de nube del Schwarz Group, dueños de Lidl) es un caso paradigmático de integración vertical hacia la infraestructura para ofrecer "PhariaAI-as-a-Service", una solución de IA totalmente soberana alojada en Alemania [242].

En el caso de integración de modelos a infraestructura, Mistral AI ha comenzado a integrarse verticalmente hacia el eslabón de infraestructura mediante asociaciones estratégicas con NVIDIA para construir centros de datos dedicados en Francia. Este movimiento permite a Mistral transformarse de

---

[237] Lagarde, C. (2025). *The transformative power of AI: Europe's moment to act.* European Central Bank (November 24). Disponible en: https://www.ecb.europa.eu/press/key/date/2025/html/ecb.sp251124_1~c239fb4a7f.en.html

[238] Escarzaga, A. (2025). "Record breakers: Europe's top 10 AI funding rounds of 2025". *EU-Startups* November 6). Disponible en: https://www.eu-startups.com/2025/11/europe-top-10-ai-funding-rounds-of-2025/

[239] Sifted AI 100: Europe's rising AI startups in 2025, Disponible en: https://sifted.eu/rankings/ai-100-2025

[240] Intuition Labs (2026). *Top 20 Medtech Companies Leveraging AI in 2025* (January 13). Disponible en: https://intuitionlabs.ai/articles/top-20-medtech-companies-using-ai-2025

[241] Mistral AI nears $14B valuation in latest funding round, positioning it as Europe's AI frontrunner, Disponible en: https://techfundingnews.com/mistral-ai-raises-2-billion-valuation-europe-ai-frontrunner/

[242] McKinsey & Co. (2025). *Accelerating Europe's AI adoption: The role of sovereign AI*, (December 19). Disponible en: acceso: https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/accelerating-europes-ai-adoption-the-role-of-sovereign-ai

desarrollador de modelos fundacionales a proveedor de servicios de nube integrados, capturando el margen económico tanto de la creación del modelo como de su ejecución (inferencia) [243].

2. Los puntos de presión en la cadena de valor de inteligencia artificial en Europa

Así como en el caso estadounidense y chino, la cadena de valor de IA en Europa también presenta puntos de presión y riesgo.

Sobreinversión en Infraestructura de cómputo de la IA

La construcción masiva de centros de datos en Europa, proyectando una inversión acumulada de 176,000 millones de euros hasta 2031, genera el interrogante de si la demanda real de servicios de IA podrá rentabilizar este capital [244].

La "circularidad" de los acuerdos, donde fabricantes de GPUs invierten en nubes que a su vez compran sus propios chips, puede inflar artificialmente las valoraciones y ocultar riesgos de sobrecapacidad. Un indicador crítico para monitorear debe ser la tasa de ocupación de las nuevas "Neo-clouds" y si la voluntad de pago de las empresas por servicios premium de IA compensa los elevados costos marginales de inferencia.

El volumen de deuda y la estabilidad financiera

A diferencia de los hiperescaladores estadounidenses que financian su expansión con flujos de caja operativos masivos, muchos participantes de la cadena europea están recurriendo a instrumentos de deuda [245].

- **Endeudamiento de capital de riesgo:** el volumen de financiamiento de riesgo en empresas de IA europeas alcanzó un récord de 5.600 millones de euros en 2025, representando el 12% de todo el capital invertido en startups [246]. Nscale, que opera en el segmento de data centers levantó 1.53 mil millones para adquirir GPUs de última generación, mientras que Mistral recibió 1.7 mil millones de euros. NXP Semiconductors recaudó 1,000 millones de euros, mientras que EcoDataCenter, un operador de centros de datos sueco lo hizo por 1.05 mil millones de euros [247].
- **Riesgo de Refinanciación:** si las tasas de interés se mantienen elevadas o si el retorno de la inversión se demora debido a cuellos de botella en la adopción organizacional, el sector podría enfrentar una oleada de reestructuraciones y consolidaciones forzosas en 2026 [248].

---

[243] PYMNTS.com (2025). Mistral Targets $14B Valuation in New Round (September 3). Disponible en: https://www.pymnts.com/artificial-intelligence-2/2025/mistral-ai-nears-close-of-funding-round-lifting-valuation-to-14-billion/

[244] Aussemsn M. (2026). "European data center market reaches tipping point due to AI and energy constraints", IT Daily (February 10). Disponible en: https://itdaily.com/news/datacenter/european-data-center-market-reaches-tipping-point-due-to-ai-and-energy-constraints/

[245] Mazanti Pulse (2026). *VC market update: Status of 2025 and the outlook for 2026* (January 28). Disponible en: https://www.mazantipulse.com/insights/vc-market-update-status-of-2025-and-the-outlook-for-2026/

[246] AI CertsNews (2026). European VC 2025 Investment Trends Reach New Heights. Disponible en: https://www.aicerts.ai/news/european-vc-2025-investment-trends-reach-new-heights/

[247] Djurickovic, T. (2026). "The 25 tech companies that dominated European funding in 2025", TechEli (January 15). Disponible en: https://tech.eu/2026/01/15/the-25-tech-companies-that-dominated-european-funding-in-2025/

[248] Asisteapp (2025). Startups maduras y rentables: IA, regulación y capital (noviembre 19). Disponible en: https://asisteapp.co/startups-maduras-y-rentables-ia-regulacion-y-capital/

Fragmentación del mercado único y brecha de innovación

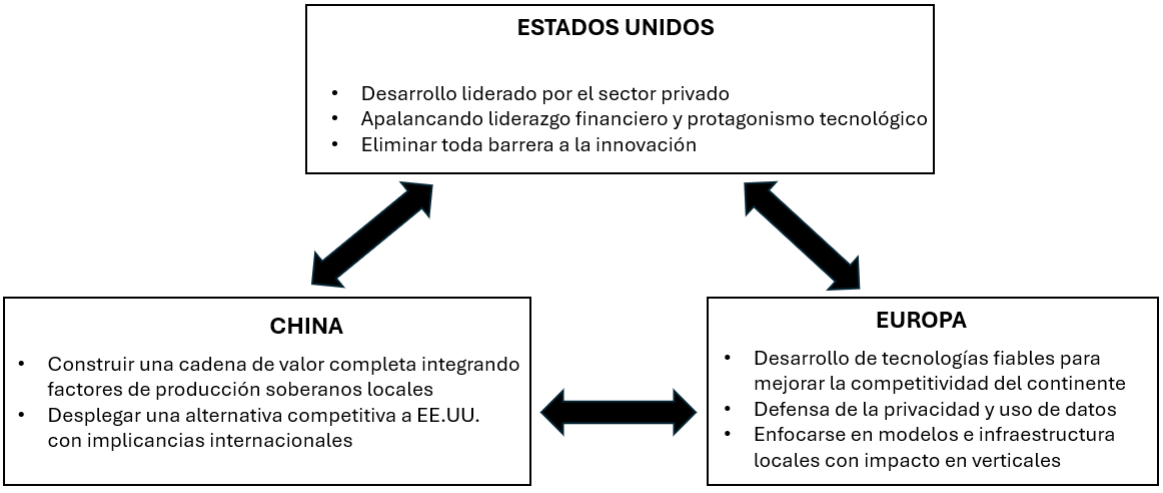
A pesar del esfuerzo armonizador de la Ley de IA, Europa sigue operando como 27 mercados distintos en términos de permisos energéticos, acceso al talento y despliegue de fibra óptica [249]. El Informe Draghi advierte que esta fragmentación impide que las empresas europeas escalen con la rapidez necesaria, resultando en la probabilidad de adquisiciones tempranas por parte de gigantes estadounidenses.

C. Dimensiones geopolíticas de la cadena de valor de la inteligencia artificial

Reconociendo la importancia económica de la IA como tecnología de propósito general, los países despliegan estrategias para posicionarse en su cadena de valor para cumplir dos objetivos: i) encarar el desarrollo soberano (es decir independiente) de IA para optimizar su acceso y garantizar control sobre eslabones críticos, y/o ii) acelerar la adopción de la tecnología por parte de organizaciones o individuos. Ambas opciones reconocen el valor económico intrínseco de la IA como acelerador de productividad y competitividad económica, lo que significa que la diferencia entre ambos objetivos es hasta qué punto un país puede construir capacidades autónomas para desarrollar modelos de IA, y adaptarlos a las necesidades del aparato productivo. Esta opción tiene, obviamente, implicancias en términos de desarrollo de infraestructura y acceso a capital humano.

El esfuerzo para desarrollar y desplegar IA representa una de las fronteras de competencia entre naciones, con implicancias geoestratégicas. Países y regiones compiten para controlar el acceso a ventajas comparativas para asegurar la preeminencia económica, política y militar derivadas de la IA en un contexto internacional fragmentado y polarizado. En general, el conflicto geopolítico se despliega entre tres polos geográficos (ver diagrama 19).

Diagrama 19
Objetivos geoestratégicos de tres polos mundiales de desarrollo de la inteligencia artificial



Fuente: Elaboración propia.

[249] Rua, C. (2025). "El futuro de la innovación en Europa: la respuesta al informe Draghi" *PierNext,* (febrero 21), Disponible en: https://piernext.portdebarcelona.cat/gobernanza/el-futuro-de-la-innovacion-en-europa-la-respuesta-al-informe-draghi/

El desarrollo de la industria de la IA en Estados Unidos está guiado por la idea que el sector privado debe estar direccionando el proceso innovador, posicionando al país como líder mundial. Apalancando las posiciones de liderazgo financiero y protagonismo tecnológico, Estados Unidos busca eliminar toda barrera a la innovación, especialmente en términos regulatorios, asegurando el acceso a insumos intermedios. Al mismo tiempo, reconociendo el avance chino, Estados Unidos despliega iniciativas tendientes a obstaculizar el desarrollo tecnológico del país asiático. Las barreras al desarrollo chino de la IA están dirigidas a controlar al país asiático en su cadena de aprovisionamiento tecnológico, principalmente en GPUs. Asimismo, Estados Unidos interviene en la adquisición de tecnologías clave (por ejemplo, prohibir a China acceder a tecnología de punta para la fabricación de semiconductores producida por ASML, una empresa holandesa), o en los programas de asistencia internacional para el desarrollo de IA (frente al anuncio de Malasia que iba a desplegar 3,000 semiconductores de Huawei como parte de una iniciativa nacional de IA, el Departamento de Comercio de EE. UU. anunció que esto violaría los controles de exportación, a raíz de lo cual Malasia debió cancelar la operación).

Por otro lado, el objetivo de China no es posicionarse en ciertos eslabones de la cadena de valor de la IA para poder competir con Estados Unidos, sino desplegar una cadena de valor paralela completa con factores de producción soberanos que permitan al país construir una alternativa competitiva a la estadounidense. En última instancia, el posicionamiento de China en IA busca replicar los avances de este país en industrias como el equipamiento de telecomunicaciones, maquinaria eléctrica, vehículos, y farmacéutica, todas ellas de capital intensivo con altas economías de escala. Internacionalmente, el mercado objetivo para la estrategia de la PRC es el denominado "sur global" (lo que incluye a América Latina y el Caribe, Asia Subsahariana, y Asia Pacífico, excluyendo a Australia, Nueva Zelanda y Singapur) [250], lo que representa 34% de la economía mundial. En este entorno, China se posiciona como líder del Sur Global, diferenciándose del modelo occidental, a partir del despliegue de apoyo financiero y tecnológico a gobiernos. La construcción de una posición internacional de liderazgo en IA es crítica debido al carácter de la tecnología como insumo intermedio a múltiples industrias estratégicas para China como defensa, aeroespacial, y automotriz. En este contexto, el Estado chino juega un papel importante en términos de provisión de financiamiento y de direccionamiento industrial.

Europa, concebida como una alianza de los denominados "*middle powers*", busca usar la IA como palanca que le permita asumir una posición de liderazgo mundial, apalancada por sus sistemas avanzados de educación e innovación. Asimismo, a diferencia de los mercados de Estados Unidos y China, la cadena de valor europea de la IA está intrínsecamente ligada a un marco regulatorio pionero, basado en la Ley de IA de la UE (EU AI Act) y el Plan de Acción de IA, centrados en el desarrollo de tecnologías fiables para mejorar la competitividad del continente, al tiempo que preserva los valores democráticos y la privacidad. El objetivo último es implementar los beneficios de la IA en diversos sectores, como la salud, la educación, la industria y la sostenibilidad ambiental, redefiniendo las barreras de entrada y las ventajas competitivas de las firmas locales [251]. El marco de la Unión Europea está complementado con políticas industriales nacionales, como "AI para la Humanidad" (Francia), y la Estrategia Federal de la Inteligencia Artificial (Alemania).

La competencia geopolítica entre naciones y bloques es potenciada por el acceso y/o control que puede existir en ciertos eslabones clave de la cadena de valor de la IA. En primer lugar, el acceso a semiconductores de última generación determina que Estados Unidos (NVIDIA en el diseño de chips), Taiwán (TSMC en la fabricación de chips) y China (Huawei y SMIC en la fabricación de chips) (con una posición embrionaria de Europa) son las naciones mejor posicionadas. En segundo lugar, el acceso al poder de cómputo para el entrenamiento e inferencia de modelos enfrenta a Estados Unidos (con AWS,

[250] Ver lista completa de países en Balbontin, R. (2026). *The global trade battleground: US-China competition in the global South*" Information Technology & Innovation Foundation (April).

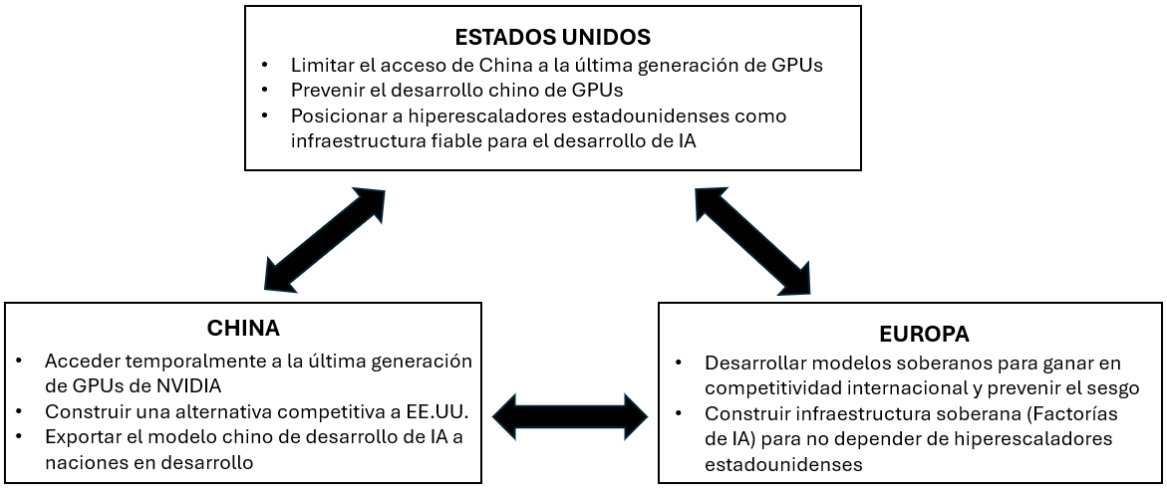
[251] Ver Comisión Europea (2025). *Enfoque europeo de la inteligencia artificial*. Disponible en: https://digital-strategy.ec.europa.eu/es/policies/european-approach-artificial-intelligence

Google Cloud, Microsoft Azure), China (Alibaba, Baidu, y ByteDance) y Europa (con el proyecto de factorías públicas). En tercer lugar, el acceso a los recursos humanos necesarios tanto para el desarrollo de modelos, herramientas como en su implementación en organizaciones determina que los Estados Unidos, China y Europa estén compitiendo en condiciones de paridad competitiva.

Finalmente, el acceso a datos para entrenar modelos resulta en una ventaja para Estados Unidos (Meta, Amazon y Google), y China (Baidu y ByteDance), mientras que Europa encara un desarrollo compartido mediante "Espacios Europeos de Datos".

En estas cuatro dimensiones de competencia geopolítica (chip, computo, talento y datos), Estados Unidos, China y Europa despliegan políticas públicas orientadas a maximizar su desarrollo autónomo y control, políticas que conllevan implicancias geopolíticas. De todas maneras, el tenor de estas políticas varía entre el énfasis europeo en preservar derechos a privacidad y gobernanza y un eje intervencionista estatal de China (ver diagrama 20).

Diagrama 20
Estrategias de posicionamiento geopolíti



Fuente: Elaboración propia.

En este marco, corresponder examinar el papel que puede cumplir América Latina, una región que incluye países de desarrollo medio y naciones en desarrollo. En términos generales, es dable estipular que los países de desarrollo medio de la región (como Brasil, Chile y Uruguay) pueden apalancar sistemas educativos de calidad y ecosistemas de innovación para jugar un papel en el juego competitivo internacional, aunque esto requerirá aumentar significativamente el nivel de inversión. Por otro lado, las naciones en desarrollo de América Latina deberán examinar su alineamiento con alguno de los tres líderes de la competencia geopolítica para obtener acceso a los factores de producción de la tecnología.

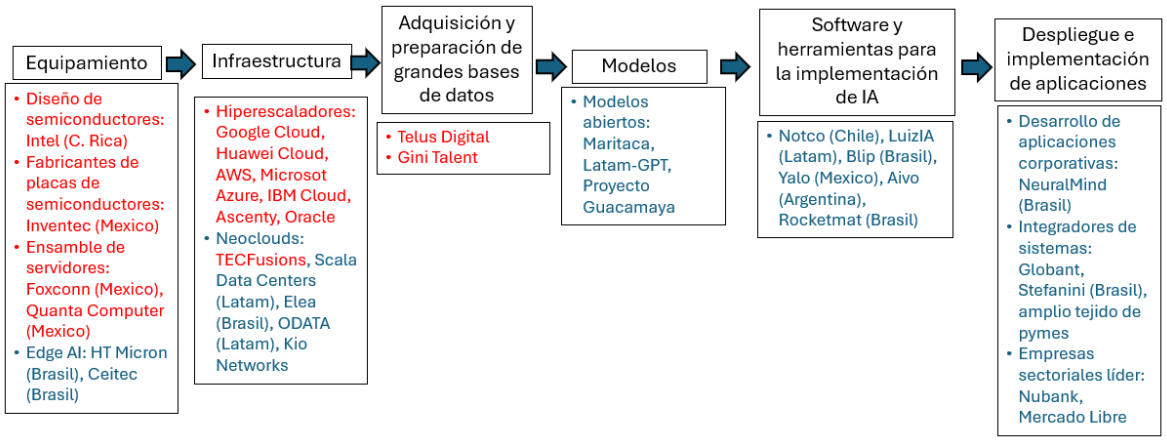
IV. La cadena de valor de la inteligencia artificial en América Latina

Como se anticipó en la introducción de este estudio, reconociendo las limitaciones estructurales en inversión y desarrollo tecnológico, América Latina y el Caribe, históricamente relegada al consumo de tecnología, intenta (i) asumir ciertas posiciones en el entrenamiento de modelos fundacionales en español y portugués, (ii) impulsar el desarrollo de centros de datos (apalancando capital privado principalmente extranjero), y (iii) desarrollar recursos humanos para la adaptación de IA a necesidades del aparato productivo. Este capítulo se enfoca en analizar la situación de la región en la cadena de valor de la IA. La sección 1 presenta el desarrollo y posicionamiento de jugadores en la región, mientras que la siguiente sección detalla el dimensionamiento económico de la región por eslabón de la cadena de valor.

A. Estructura de la cadena de valor

La cadena de valor de la IA en América Latina y el Caribe demuestra un avance en los eslabones de desarrollo local de software, herramientas e implementación de sistemas, con una presencia principalmente extranjera en infraestructura, y un movimiento embrionario en modelos fundacionales soberanos (ver diagrama 21).

Diagrama 21
América Latina y el Caribe: cadena de valor de IA



Fuente: Elaboración propia sobre la base de información corporativa y fuentes secundarias especializadas.
Nota: Azul indica desarrollo/capital local.

1. Equipamiento

Contrariamente a lo esperado, la región muestra una presencia incipiente en el eslabón de equipamiento para IA, aunque de manera segmentada y concentrada en ciertos países. Esta presencia responde, por una parte, a la reconfiguración geopolítica de las cadenas globales (es decir, la reducción de la dependencia de regiones afectadas por riesgos militares, como la originada en el Estrecho de Taiwán) y de estrategias de "*nearshoring*" hacia el mercado estadounidense (en el caso de México). Adicionalmente, la presencia en el eslabón de equipamiento está dirigida a apalancar capacidades locales específicas, como la ingeniería vinculada a semiconductores (en el caso de Costa Rica), y algunos nichos de componentes y electrónica para satisfacer la demanda regional y mundial (en el caso de Brasil).

Con respecto al primer factor, en parte impulsado por la necesidad de mitigar el riesgo de su localización en Taiwán y China, Foxconn ha confirmado una inversión de US$ 168 millones para expandir su capacidad productiva en Jalisco para ensamblar los GPUs de última generación de NVIDIA. [252] Esta inversión está dirigida no solo a reducir la presencia de Foxconn en China, sino también para responder a la demanda global de servidores de IA. La inversión de Foxconn en el segmento de IA se agrega a la presencia de la empresa en instalaciones en Juárez, Chihuahua, Monterrey y Guadalajara enfocadas en componentes electrónicos avanzados, y productos para la industria automotriz, donde ya emplea 35,000 trabajadores [253].

Asimismo, Inventec, un fabricante de placas para servidores de IA de Taiwán ha iniciado actividades de fabricación en México [254]. En el área de ensamble de racks de servidores de GPUs se cuenta también en México con la presencia de Quanta Computer [255]. El despliegue de ensambladores

[252] MBN (2025). "Foxconn invests US$ 168 million to expand serve production". *Mexico Business*. Disponible: https://mexicobusiness.news/cloudanddata/news/foxconn-invests-us168-million-expand-ai-server-production
[253] Mexico Daily (2025). "Foxconn invests US$ 168 million to expand AI server production." *Mexico News Daily* (August 27). Disponible en: https://mexiconewsdaily.com/business/foxconn-investing-in-mexico/
[254] Trendforce News (2023). *Quanta, Wiwynn, and major manufacturers scale up to meet rising demand for AI servers*. (October 24). Disponible en: https://www.trendforce.com/news/2023/10/24/insights-quanta-wiwynn-and-major-manufacturers-scale-up-to-meet-rising-demand-for-ai-servers/
[255] Qanta Says AI server plans on track. *Taipei Times* (August 13). Disponible en: https://www.taipeitimes.com /News/biz/archives/2025 /08/13/2003841930

de equipamiento para la IA está no solamente determinado por consideraciones geopolíticas, sino también logísticas. La localización de México reduce el costo logístico de transporte de servidores de IA a los centros de datos que están siendo desplegados en Estados Unidos, especialmente en Texas.

Adicionalmente al despliegue de manufactura de equipamiento de IA en México, Costa Rica mantiene una posición en la ingeniería de alto nivel a partir de la presencia ya histórica de Intel de su Centro de Excelencia en Ingeniería y servicios globales de diseño [256]. Es importante mencionar que Intel mantendrá operaciones de ingeniería y servicios corporativos aun después de relocalizar sus actividades de ensamble y prueba de alto volumen a Vietnam y Malasia. Ello sugiere que el país conserva capacidades relevantes en segmentos intensivos en conocimiento, aunque ya no en la manufactura de mayor escala [257].

Finalmente, Brasil demuestra la presencia de actores enfocados en satisfacer nichos específicos de componentes y semiconductores para satisfacer la demanda global y latinoamericana. Por ejemplo, HT Micron se ha posicionado en soluciones System-in-Package que integran conectividad y procesamiento para Edge AI (sensores inteligentes, dispositivos de IoT industrial y módulos de conectividad) que alimentan a modelos de los grandes centros de datos [258]. Adicionalmente, CEITEC, una empresa estatal está enfocada en la producción de semiconductores ligados a la gestión inteligente de energía y dispositivos para vehículos eléctricos [259]. En conjunto, ello sugiere una inserción enfocada en segmentos especializados más que en la producción masiva de hardware crítico para IA.

2. Infraestructura de cómputo de la IA

El despliegue de centros de datos en América Latina y el Caribe puede ser analizado a partir de dos series estadísticas: la presencia de centros de datos sin diferenciación, con lo cual se incluye infraestructura propiedad de empresas y centros de co-localización (reportados por *Data Center Map*) y el despliegue de zonas de computación en la nube por parte de hiper escaladores (recolectados por Telegeography). La segunda serie es más ajustada a los requerimientos de IA. En ambos planos, la infraestructura regional se concentra principalmente en Brasil, México y Chile, mientras que América Latina y el Caribe mantiene una escala menor que la observada en América del Norte, Asia y Europa.

Considerando veinte países de América Latina y el Caribe para los que se dispone de información, la región ha evolucionado de 95 centros de datos en el 2015 a 475 diez años más tarde (ver cuadro 20).

Cuadro 20
América Latina y el Caribe: centros de datos, 2015-2025

	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
Argentina	10	12	11	11	14	14	14	15	15	29	34
Bolivia (Estado Plurinacional de)	2	2	2	2	3	3	3	2	2	4	5
Brasil	33	35	39	41	45	63	64	73	82	146	185
Chile	8	9	9	9	11	12	14	15	15	50	64
Colombia	6	6	6	6	6	6	7	7	7	32	40
Costa Rica	7	7	7	7	7	7	7	7	7	8	10

[256] Essential News (2022). "In the last two years INTEL invested close to US$ 1 billion in Costa Rica and officially inaugurated an Assembly and Test plant." August 9. Disponible en: https://www.cinde.org/en/essential-news/in-the-last-two-years-intel-invested-close-to-usd-1billion-in-costa-rica-and-officially-inaugurated-an-assembly-and-test-plant

[257] Nhat, P. (2025). "Intel expands Vietnam operations as it relocates chip assembly from Costa Rica". *Hanoi Times* (October 25). Disponible en: https://hanoitimes.vn/intel-expands-vietnam-operations-as-it-relocates-chip-assembly-from-costa-rica.887172.html

[258] HT Micron – *STMicroelectronics*. Disponible en: https://www.st.com/content/st_com/en/partner/partner-program/partnerpage/HT-Micron.html

[259] Castro, E. (2024). "Ministry invests R$220 million in CEITEC for semiconductor production". *Canal Solar* (December 11). Disponible en: https://canalsolar.com.br/en/ministry-invests-in-ceitec-semiconductor-production/

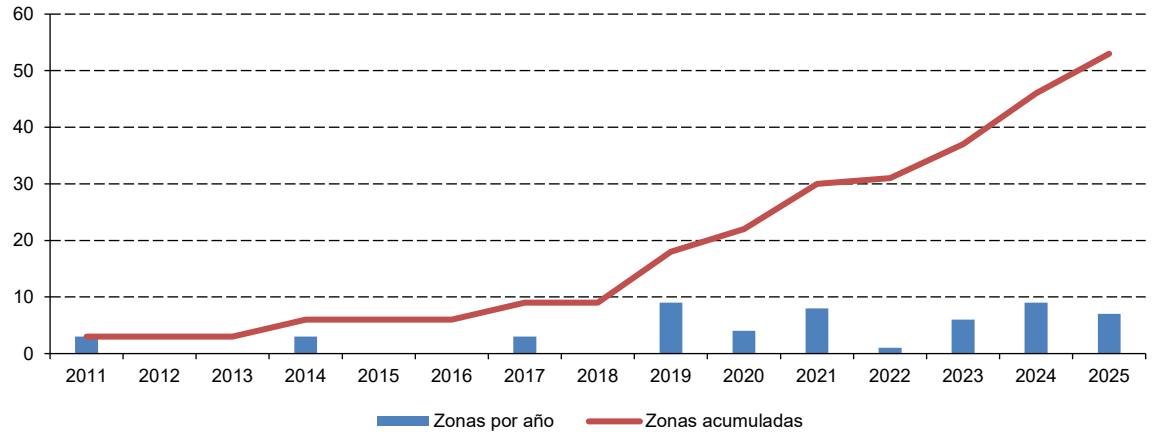
	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
Ecuador	1	1	1	1	1	1	1	1	1	7	9
El Salvador	0	0	0	0	0	0	1	1	1	3	3
Guatemala	1	1	1	1	1	1	2	2	2	5	5
Honduras	0	0	0	0	0	0	1	1	1	2	4
Jamaica	0	0	0	0	0	0	0	0	0	0	1
México	11	11	11	11	10	10	13	14	15	51	59
Nicaragua	1	1	1	2	2	2	3	2	2	3	3
Panamá	5	6	6	6	6	6	7	7	7	14	14
Paraguay	1	1	1	1	1	1	1	0	1	3	6
Perú	1	1	1	1	1	1	1	4	4	10	15
República Dominicana	0	0	0	0	0	0	0	0	0	1	1
Trinidad y Tabago	0	1	1	1	1	1	1	1	1	2	2
Uruguay	6	7	7	7	7	7	7	7	7	7	9
Venezuela (República Bolivariana de)	2	2	2	3	5	5	5	5	5	5	6
Total	95	103	106	110	121	140	152	164	175	382	475

Fuente: Data Center Map.

En el 2026, el número de centros de datos alcanza 475, mientras que la capacidad instalada dedicada a cargas de IA es de 551 MW (megavatios). Para el 2031, se estima que el número de centros de datos alcanzará 627, y que la capacidad instalada dedicada a IA 1.6 GW (gigavatios), equivalente a 1.600 MW [260].

Por otra parte, el despliegue de centros de datos de proveedores de servicios de computación en la nube en la región alcanza a la fecha 53 (este valor ya está incluido en la serie precedente). El país más equipado con este tipo de infraestructura es Brasil con 21 zonas, seguido de México con 17, y luego Chile con 13. Como puede observarse en el gráfico 14, el despliegue de zonas de hiperescaladores proveedores de la nube ha registrado un salto a partir del 2019, reflejando no solo la migración de empresas a servicios de la nube sino también el avance en IA.

Gráfico 14
América Latina y el Caribe: despliegue de Zonas de proveedores de servicios en la nube, 2011-2025



Fuente: Elaboración propia sobre la base de Telegeography (2025).

[260] Gergs, L. (2026). Where is AI data center demand growth being driven. ABI Research (February 23).

Más allá de los centros de hiperescaladores, el despliegue y adopción de IA en la región está generando la entrada de operadores especializados en infraestructura digital y data centers preparados para cargas de IA (este es un movimiento similar al registrado en los mercados estadounidense y europeo con los "neo-clouds"). En junio de 2025, TECfusions, una empresa estadounidense, anunció el despliegue de un centro de 100 MW en Palo Alto (Chile), aunque la capacidad comenzará en 10 MW, incrementándose a medida que la demanda crezca. De manera similar, Scala Data Centers, una empresa brasilera, ya tiene centros operando en México (1), Chile (1), Brasil (4), y cuatro centros en construcción en Chile (1), Brasil (2), y Colombia (1). Scala no está construyendo simples almacenes de servidores, sino que está diseñando instalaciones de alta densidad preparadas para refrigeración líquida y conectadas a fuentes de energía renovable masiva. Scala invertirá US$ 50 mil millones en la construcción de un campus denominado AI City en Rio Grande do Sul. Scala también está construyendo tres centros en Chile con una capacidad proyectada de 53 MW. [261] De acuerdo con el sitio de Scala, la empresa se posiciona como un socio estratégico indispensable para que empresas como Microsoft u OpenAI puedan desplegar sus modelos en la región en centros con baja latencia. Otro operador, Elea Data Centers, está construyendo un centro enfocado en IA en Rio de Janeiro. De manera agregada, se proyecta que el despliegue de centros de datos en América Latina agregará 1831 MW de capacidad entre 2025 y 2030, de los cuales 858 MW estarán localizados en Brasil y 431 MW en México [262].

Es interesante remarcar la inversión rezagada en el despliegue de centros de datos en Argentina. Diversos factores como la inestabilidad económica e institucional, la falta de políticas sostenidas para atraer inversiones tecnológicas, y la ausencia de incentivos fiscales específicos han hecho que muchos actores internacionales del sector prioricen países vecinos. En general, los hiperescaladores han optado por instalar sus data centers en entornos más previsibles y con mejores condiciones regulatorias. Esto ha limitado la capacidad local de ofrecer servicios de baja latencia y ha aumentado la dependencia de nodos externos para el procesamiento de información crítica [263]. Al momento, Argentina contiene un pequeño centro de Huawei, desplegado en el 2019 y una zona Edge de AWS. En este marco, OpenAI y Sur Energy anunciaron en el 2025 una inversión de US$ 25 mil millones para construir un centro de 500 MW (en su última etapa) en Patagonia que sería parte del proyecto Stargate. Al momento, el proyecto fue solo refrendado con un Memorándum de Intención [264]. Sin embargo, diferentes prioridades entre los socios de Stargate —Oracle, OpenAI, y Softbank— recientemente reportadas, han cambiado la estrategia de construcción y financiamiento de data centers, con lo cual el plan argentino estaría en duda de si se concretará [265].

En conclusión, de manera agregada, la región presenta una posición más limitada respecto a otras regiones del mundo en términos de despliegue de centros de datos (ver gráfico 15).

---

[261] Scala Data Centers (2025). Scala Data Centers closes USD 328 million international deal, securing funds for three data centers and a power substation in Chile (July). Disponible en: https://scaladatacenters.com/scala_na_midia/scala-data-centers-closes-usd-328m-international-deal-securing-funds-for-three-data-centers-and-a-power-substation-in-chile/

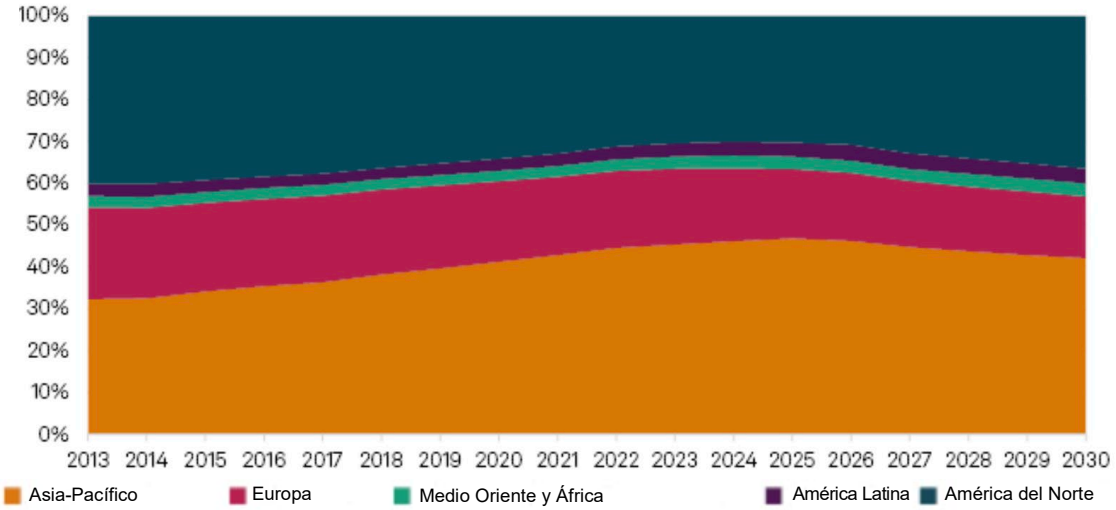
[262] Business Wire (2025). Latin America Data Center Colocation market industry outlook 2025-2030. (October 21). Disponible en: https://www.businesswire.com/news/home/20251021184222/en/Latin-America-Data-Center-Colocation-Market-Industry-Outlook-2025-2030-Key-Players-Such-As-Ascenty-Equinix-ODATA-Scala-Data-Centers-and-KIO-Networks-are-Leading-Large-scale-Developments---ResearchAndMarkets.com

[263] Cushman & Wakefield (2025). *El mercado de data centers estaba dormido. ¿Cómo se está despertando?* (Junio, 6). Disponible en: file:///Users/raulkatz/Desktop/Proposal%20AI%20Argentina/El%20mercado%20argentino%20de%20data%20centers%20estaba%20dormido.%20%C2%BFCo%CC%81mo%20se%20esta%CC%81%20despertando%3F%20%7C%20AR%20%7C%20Cushman%20&%20Wak.html

[264] Swinhoe, D. (2015). "OpenAI plans 500 MW data center in Argentina". *Data Center Dynamics* (October 13). Disponible en: https://www.datacenterdynamics.com/en/news/openai-plans-500mw-data-center-in-argentina/

[265] Gardizy, A. (2026). "Inside OpenAI's scramble to get computing power after Stargate stalled". *The Information* (February 22). Disponible en: https://www.theinformation.com/articles/inside-openais-scramble-get-computing-power-stargate-stalled?utm_campaign=article_email&utm_content=article-16643&utm_medium=email&utm_source=sg&rc=exdbp6

Gráfico 15
Superficie actual y proyectada de centros de datos
(En pies cuadrados)



Fuente: S&P Global Market Intelligence; 451 Research.

Considerando las proyecciones de despliegue de capacidad de superficie de centros de datos al 2030, América Latina muestra una cierta estabilidad alrededor de 4%, mientras que América del Norte y Asia alcanzan el 40%, y Europa 15% [266].

3. Adquisición y preparación de bases de datos

En la región opera un ecosistema de empresas extranjeras que organiza tareas de recolección, anotación, validación y preparación de datos para IA, en algunos casos recurriendo a trabajadores con formación técnica o conocimientos especializados en áreas como salud, finanzas o agricultura. El suministro de datos de la región es realizado para alimentar las cadenas globales de desarrollo de IA, muchas de ellas lideradas por empresas de países desarrollados, como ya se ha evidenciado en este estudio. En este sentido, el etiquetado se lleva a cabo a través de plataformas digitales internacionales mayoritariamente provenientes de países desarrollados que tercerizan estas tareas realizadas de forma remota, por trabajadores de la región. Estos trabajadores deben sincronizarse con horarios comerciales extranjeros, trabajar en un segundo idioma y depender muchas veces de traductores automatizados [267].

Entre estas empresas se encuentran:

- Telus Digital, empresa canadiense emplea a miles de colaboradores locales y provee servicios de anotación multilingüe en múltiples formatos, incluyendo texto, audio, video e imágenes, y posee oficinas físicas en Brasil, Costa Rica y México [268]. La empresa posee oficinas en El Salvador y Guatemala, y operaciones en Argentina, Brasil y Costa Rica [269]. Solo en Guatemala la empresa posee 14,000 colaboradores [270]. En el 2020 Telus adquirió otra empresa de suministro de datos Lionbridge IA [271].

[266] Struta, I. (2025). *Record-breaking data center investments, M&A in 2025 amid AI demand*. S&P Global (December 17).
[267] https://www.memoria.fahce.unlp.edu.ar/art_revistas/pr.19668/pr.19668.pdf
[268] https://www.lionbridge.com/es/get-in-touch/#locations
[269] https://www.telusdigital.com/about/locations
[270] https://distritoguate.com/telus-digital-celebra-30-anos-impulsando-el-progreso-y-el-desarrollo-economico-en-guatemala/
[271] https://www.lionbridge.com/press-release/lionbridge-completes-sale-of-ai-division-to-telus/

- Gini Talent, con base en Turquía, que presta servicios especializados de etiquetado de datos a medida para empresas [272]. Este tipo de etiquetado requiere no solo fluidez lingüística, sino conocimiento profundo. Por ejemplo, la anotación de imágenes médicas o la clasificación de documentos legales en portugués y español exige un nivel de especialización que los datos disponibles no pueden ofrecer con la precisión necesaria [273].

Este eslabón no está exento de restricciones regulatorias. Las exigencias en materia de protección de datos personales, tratamiento de datos sensibles, seguridad y transferencias internacionales de datos aplican a cualquier empresa —local o extranjera— que opere en estos mercados. Sin embargo, sus efectos pueden ser asimétricos, ya que los costos de cumplimiento tienden a ser más fácilmente absorbidos por actores de mayor escala y con más capacidades jurídicas, técnicas y organizacionales.

Junto con esta inserción en cadenas internacionales de preparación de datos, en la región también emergen esfuerzos propios de curaduría y construcción de corpus. El caso más significativo es LatamGPT, iniciativa regional liderada por el Centro de Inteligencia Artificial de Chile (CENIA), que contempla la recopilación, depuración y estructuración de datos provenientes de diversos países de América Latina y el Caribe para entrenar un modelo adaptado a los contextos lingüísticos y culturales de la región. Este esfuerzo muestra que el eslabón 3 no se limita a la provisión de trabajo tercerizado para actores externos, sino que también puede constituir una base para la generación de activos estratégicos regionales en materia de IA.

4. Modelos fundacionales

Siguiendo el modelo europeo, América Latina está encarando el desarrollo de modelos soberanos, es decir basados en lenguaje castellano o portugués. La necesidad de desarrollar modelos soberanos está fundamentada en el hecho de que los modelos desarrollados en Estados Unidos o China contienen un sesgo cognitivo y lingüístico distinto al de otras naciones. Este sesgo ya ha sido fundamentado en numerosos estudios. Por ejemplo, en un estudio comparado entre ChatGPT y DeepSeek en relación con la evaluación financiera de empresas chinas revela que ChatGPT es más optimista respecto a su futuro desarrollo que DeepSeek, lo que está basado en las diferentes fuentes informativas usadas por cada plataforma. [274] Otro sesgo documentado es el grado de evaluación positiva de regímenes políticos del país donde ha sido desarrollado el modelo [275]. Finalmente, un sesgo adicional ha sido detectado en la visión positiva de valores culturales del país donde se ha desarrollado el modelo [276].

Desde el punto de vista puramente tecnológico, todo modelo podría ser adaptado al contexto especifico de países para evitar sesgos cognitivos. Sin embargo, esto requiere acceder a las instrucciones (*prompts*) del código. El obstáculo es que muchos modelos no permiten acceder a dichas instrucciones dado que definen una codificación rígida en el sistema ("*hard code*"). Todo ello lleva a que muchas naciones, más allá de las europeas intentan desarrollar modelos soberanos (véase el caso de Indonesia).

---

[272] Top Data Annotation Companies in LATAM Fueling AI Innovation in 2025 - Gini Talent, fecha de acceso: abril 8, 2026, https://ginitalent.com/top-data-annotation-companies-in-latam-fueling-ai-innovation-in-2025/

[273] Modelo de referencia para políticas de IA en el sector público de américa latina y el caribe - IADB Publications, fecha de acceso: abril 8, 2026, https://publications.iadb.org/publications/spanish/document/Modelo-de-referencia-para-politicas-de-IA-en-el-sector-publico-de-America-Latina-y-el-Caribe.pdf

[274] Cao, S., Wang, C.Y., and Yi, X. (2025). *When LLMs Go abroad: Foreign bias in AI Financial predictions*. Harvard Business School Working Paper 26-013.

[275] Ho-Chun, H., Weener, T., Chen, Y-C., Noh, S., Zha, M., and Lo, H. (2025). Do language models favor their home countries? Asymmetric propagation of positive misinformation and foreign influence audits. *Harvard Kennedy School (HKS) Misinformation Review*, *6*(5).

[276] Tao, Y., Viberg, O., Baker, R. and Kizilcec, R. (2024). "Cultural bias and cultural alignment of large language models". *PNAS Nexus*, 3, pp. 346.

El desarrollo de modelos soberanos no implica que países en desarrollo lo hagan de manera totalmente independiente. De hecho, NVIDIA, Google Cloud y AWS son algunas de las empresas que colaboran con países en desarrollo en el entrenamiento de modelos soberanos.

En América Latina, se han identificado en particular tres modelos desarrollados localmente, el primero impulsado por el sector privado y el segundo a partir de una colaboración público-privada. En Brasil, Maritaca AI es una empresa local privada basada en Campinas, y vinculada indirectamente a la Universidad de Campinas, enfocada en el desarrollo de grandes modelos de lenguaje con un foco en calidad, y eficiencia. La empresa ya ha lanzado dos modelos Sabia-2 y Sabia-3 en portugués, posicionándolos en competencia con GPT-4. Estos modelos se basan en una premisa técnica distinta: en lugar de entrenar modelos con billones de parámetros de internet (donde el portugués es una fracción menor), Maritaca utiliza un enfoque de *Curriculum Learning* y especialización de dominio sobre corpus brasileños de alta calidad. De acuerdo con informes de desempeño, Sabiá-4 (y la versión ligera eficiente Sabiazinho-4) ha demostrado un rendimiento igual o superior a GPT-4 en tareas específicas de lengua portuguesa y conocimiento cultural brasileño, superando a los modelos globales en exámenes estandarizados como el ENEM (ingreso universitario) y el examen de la OAB (Orden de Abogados de Brasil) [277]. Asimismo, la eficiencia de su arquitectura les permite ofrecer costos por token de 3 a 4 veces menores que los modelos de frontera globales. Esta ventaja es crítica para empresas locales que facturan en moneda local pero que, al usar OpenAI, deben pagar APIs en dólares, exponiéndose a la volatilidad cambiaria.

En el frente académico y público-privado, CENIA (Chile) lidera el proyecto LatamGPT, un ente coordinando 33 instituciones latinoamericanas, con participación de CAF y Amazon Web Services, lanzado en junio de 2025 [278]. El apoyo de AWS fue realizado en el marco del acuerdo del hiperescalador con el Data Observatory y permitió el entrenamiento de la primera versión del modelo en el curso de cuatro meses y medio [279]. A diferencia de Maritaca AI, que es una iniciativa comercial, CENIA coordina un esfuerzo panregional para crear modelos de código abierto sin fines comerciales, impulsado por el Ministerio de Ciencia, Tecnología, Conocimiento e Innovación de Chile, al que ya se han unido instituciones de Brasil, República Dominicana, Perú, Costa Rica, y Panamá. El objetivo de LatamGPT no es el de competir con los grandes laboratorios de Estados Unidos sino el de generar un modelo basado en un corpus latinoamericano y, en ese camino, generar capacidades locales.

El modelo está basado en la arquitectura de Llama 3.1 – 70 B de Meta constituido por un pre-entrenamiento continuo (CPT), y ya ha sido entrenado y está compuesto por 70 mil millones de parámetros derivados de 8+ terabytes de datos latinoamericanos (300 mil millones de *tokens* [280]), con características similares a ChatGPT 3.5. Si bien la primera versión fue entrenada en 9 días, utilizando la infraestructura de AWS [281], en el futuro el modelo estará alojado para inferencias en la supercomputadora de la Universidad de Tarapacá (Chile) [282], y tendrá aplicaciones en educación, política pública y la preservación de lenguas no mayoritarias (Mapuche, Rapanui, y Guaraní) [283]. En lo

---

[277] Abonizio, H., Sales Almeida, T., Laitz, T., Malaquias Junior, R., Kerche Bonas, G. Nogeira, R. and Pires, R. (2024). *Sabia-3 Technical Report*. Disponible en: https://arxiv.org/abs/2410.12049

[278] Latam GPT: Learn about the Latin American AI model developed in Chile (February 13, 2025). Disponible en: https://www.gob.cl/en/news/LatamGPT-learn-about-the-latin-american-ai-model-developed-in-chile/

[279] Entrevista con Alexandra Garcia, de LatamGPT.

[280] Los *tokens* son unidades mínimas de texto que el modelo procesa. No siempre equivalen a palabras completas. Un *token* puede ser una palabra entera, una parte de palabra, un signo de puntuación, un número, o incluso un espacio en ciertos sistemas de tokenización.

[281] Fuente: https://www.LatamGPT.org/en/faq. El supercomputador de la Univesidad de Tarapaca constara de 12 nodos con 96 GPUs H200 de NVIDIA.

[282] La Universidad opera una computadora equipada con 96 GPUs H200 de NVIDIA.

[283] The Tech Buzz (2025). *Latin America builds rival to ChatGPT wih 50 billion parameter model*. (September 1). Disponible en: https://www.techbuzz.ai/articles/latin-america-builds-rival-to-chatgpt-with-50b-parameter-model

que corresponde a la multimodalidad operativa [284], LatamGPT es esencialmente textual. La dimensión de verificación o encuadre regulatorio externo explícito [285] de LatamGPT declara procesos de curación ética, anonimización y enfoque de derechos, lo cual es importante, aunque esta sigue siendo sobre todo interna al proyecto.

En el ámbito académico, el Proyecto Guacamaya: IA es un modelo de aplicación especifica en el campo ambiental para la Amazonia. Este modelo ha sido desarrollado como resultado de una colaboración entre el Instituto Alexander Von Humboldt, *Microsoft*, el Centro de Investigación y Formación en Inteligencia Artificial (CINFONIA) de la Universidad de Los Andes en Colombia y el Instituto Amazónico de Investigaciones Científicas (SINCHI) [286]. El proyecto ha desarrollado un modelo fundacional multimodal que integra datos de cámaras trampa, bio-acústica e imágenes satelitales para el monitoreo de la deforestación y la fauna. "El sistema SPARROW (*Solar-Powered Acoustic and Remote Recording Observation Watch*), desarrollado por Microsoft, combina sensores inteligentes, energía solar e inteligencia artificial para capturar, transmitir y analizar datos directamente desde la selva" [287]. Los modelos pre entrenados están disponibles en *Pytorch Wildlife*, una plataforma de código abierto de *Microsoft* para la biodiversidad, y el espíritu del proyecto es que toda la comunidad científica y otros países puedan colaborar [288]. El proyecto se está comenzando a expandir: en el año 2025 Perú decidió adherir [289], instancia necesaria porque la integridad del Amazonas depende de una coordinación regional sólida apoyada por la tecnología.

A pesar de estos intentos de desarrollo local latinoamericano, corresponde mencionar que la adopción de IA, medida en términos de tráfico a sitios clave, está liderada por los modelos estadounidenses (ver cuadro 21).

Cuadro 21
Países de América Latina: visitantes a principales sitios de herramientas de IA, noviembre 2025
(En cantidad de visitas)

País	ChatGPT	Perplexity	Copilot	Deepseek	Gemini	Claude	Total
Brasil	289 890 514	9 124 266	4 439 806	12 169 248	35 833 768	7 330 034	358 787 636
Chile	39 346 564	1 348 101	1 305 349	1 469 075	4 081 823	857 457	48 408 369
Colombia	78 404 394	2 698 712	1 399 207	3 576 000	7 056 423	2 403 906	95 538 642
Costa Rica	6 764 818	185522	141 599	339 380	600 884	241 423	8 273 626
El Salvador	4 830 779	108 698	142 298	304 671	685 951	162 323	6 234 720
Guatemala	6 703 902	170 325	217 359	488 330	808 189	237 759	8 625 864
Uruguay	5 463 219	4 729 571	122 992	60 691	164 929	273 336	10 814 738

Fuente: Semrush.

Esta dominancia no es patrimonio del uso casual de individuos. Las APIs de estas plataformas son la infraestructura sobre la que se construyen aplicaciones corporativas en la región, lo que resulta en una ventaja de distribución difícil de replicar para cualquier plataforma local.

---

[284] Capacidad de un sistema de IA para trabajar con más de un tipo de información.
[285] Se refiere cuando el modelo es evaluado por su equipo y está validado por una entidad externa, normalmente, pública, regulatoria o independiente con la finalidad de tener reglas formales sobre transparencia del modelo, auditorías de riesgos, criterios de uso en sector público, revisión de sesgos o seguridad, cumplimiento con normas de protección de datos.
[286] https://news.microsoft.com/es-xl/proyecto-guacamaya-inteligencia-artificial-para-preservar-la-amazonia/
[287] https://sinchi.org.co/la-inteligencia-artificial-llega-a-la-selva-amazonica-para-monitorear-y-conservar-su-biodiversidad
[288] https://indicelatam.cl/el-vuelo-de-guacamaya-el-proyecto-en-que-microsoft-y-la-investigacion-confluyeron-en-la-ia/
[289] https://www.gob.pe/institucion/minam/noticias/1128778-minam-y-microsoft-lanzan-proyecto-para-monitorear-ecosistemas-usando-inteligencia-artificial

5. Software y herramientas

En América Latina y el Caribe, el eslabón de software y herramientas para la implementación de IA es más incipiente que el de desarrollo de aplicaciones, pero constituye un espacio relevante para la acumulación de capacidades tecnológicas. Este segmento comprende plataformas, componentes de software, herramientas de automatización, sistemas conversacionales, soluciones de procesamiento de lenguaje natural, plataformas de datos, RPA, APIs, herramientas de integración y entornos de orquestación que permiten construir, adaptar, desplegar o escalar soluciones de IA en organizaciones públicas y privadas.

En la región, la frontera entre el eslabón de software y herramientas y el de desarrollo de aplicaciones es particularmente porosa. A diferencia de ecosistemas más maduros, como los de Estados Unidos, China o Europa, donde existe una mayor especialización entre plataformas, herramientas, integradores y aplicaciones finales, muchas empresas regionales combinan plataforma propia, consultoría, integración e implementación de soluciones verticales. Por ello, la clasificación debe basarse en la función dominante de la empresa o producto dentro de la cadena: herramientas reutilizables que habilitan a terceros a desarrollar, integrar o desplegar IA corresponden principalmente al eslabón de software y herramientas, mientras que soluciones finales orientadas a casos de uso sectoriales corresponden principalmente al eslabón e desarrollo de aplicaciones.

La región no cuenta aún con una masa significativa de plataformas horizontales comparables a las de los grandes proveedores globales de nube, modelos fundacionales o software empresarial. Sin embargo, sí existen capacidades emergentes en herramientas especializadas, especialmente en plataformas conversacionales, automatización empresarial, gestión de datos, procesamiento documental, integración de APIs e incorporación de IA en flujos de trabajo organizacionales.

Uno de los subsegmentos más visibles es el de plataformas conversacionales y agentes de IA. Empresas como Blip en Brasil, Botmaker en Argentina, Zenvia en Brasil, Aivo en Argentina, Yalo en México y Vozy en Colombia ofrecen herramientas para crear, desplegar y operar chatbots, voicebots, agentes de IA y soluciones de comunicación automatizada en canales como WhatsApp, correo electrónico, voz, aplicaciones y redes sociales. Estas plataformas cumplen una función habilitante en la medida en que permiten a empresas de distintos sectores automatizar atención al cliente, ventas, cobranza, soporte, marketing y comercio conversacional.

Otro grupo de empresas se orienta a plataformas de datos, analítica y orquestación de IA. Semantix, de origen brasileño, constituye un caso relevante por su posicionamiento en big data, analítica e inteligencia artificial empresarial, mientras que NeoSpace representa una experiencia más reciente en plataformas de datos y modelos orientados a predicciones en tiempo real, especialmente en servicios financieros. Aunque estas empresas operan en segmentos todavía acotados frente a las grandes plataformas globales, muestran capacidades regionales para desarrollar capas intermedias de software que conectan datos, modelos y decisiones empresariales.

También existen herramientas de automatización e integración empresarial que cumplen un papel importante en el despliegue de IA. Rocketbot, de origen chileno, ofrece una plataforma de automatización robótica de procesos que permite construir robots de software e integrar flujos de trabajo. Stefanini, de origen brasileño, ha desarrollado soluciones como SophieX, un asistente de IA generativa para entornos corporativos con capacidades multilingües. Estas herramientas se ubican en una zona intermedia entre software habilitante y aplicación empresarial, pero resultan relevantes para el eslabón 5 cuando funcionan como plataformas reutilizables para automatizar procesos, conectar sistemas y apoyar la implementación de IA en organizaciones.

Finalmente, destacan capacidades regionales en procesamiento de lenguaje natural, búsqueda semántica y gestión documental. NeuralMind, en Brasil, desarrolla agentes de IA orientados a buscar información, automatizar flujos de trabajo y apoyar decisiones sobre bases de conocimiento empresariales y gubernamentales. Este tipo de soluciones es especialmente relevante para la región, dada la importancia de adaptar herramientas de IA a idiomas, marcos regulatorios, documentos administrativos, sistemas judiciales, servicios financieros y necesidades institucionales locales.

Un caso particular de frontera entre los eslabones de software y herramientas y desarrollo de aplicaciones es NotCo, empresa de origen chileno que inicialmente utilizó IA para desarrollar sus propios productos alimentarios de origen vegetal, pero que progresivamente ha convertido su plataforma tecnológica en una herramienta B2B para la innovación de productos en la industria de alimentos. Su plataforma Giuseppe AI permite analizar ingredientes, características moleculares, propiedades funcionales, costos y perfiles nutricionales para apoyar procesos de reformulación y desarrollo de nuevos productos. La alianza con The Magnum Ice Cream Company refuerza esta transición: Magnum anunció que piloteará la plataforma de NotCo para desarrollar nuevos formatos de producto y perfiles nutricionales, así como para acelerar sus procesos de innovación. En este sentido, NotCo ilustra la porosidad entre los eslabones 5 y 6 en América Latina y el Caribe: cuando la empresa desarrolla productos propios basados en IA se ubica en el eslabón de aplicaciones, pero cuando ofrece Giuseppe AI como plataforma reutilizable para terceros se aproxima al eslabón de software y herramientas.

En conjunto, estos casos sugieren que la oportunidad de América Latina y el Caribe en el eslabón de software y herramientas no reside necesariamente en competir con las grandes plataformas globales de IA generalista, sino en desarrollar capas intermedias, herramientas especializadas y software empresarial que faciliten la adopción de IA en contextos locales. No obstante, el segmento enfrenta restricciones importantes: dependencia de modelos fundacionales y nubes globales, escasez de capacidades avanzadas en MLOps y LLMOps, limitaciones de financiamiento, baja escala regional, dificultades de internacionalización y necesidad de estándares de interoperabilidad, seguridad y gobernanza. Por ello, el fortalecimiento de este eslabón requiere políticas orientadas a software avanzado, datos, talento, infraestructura de prueba, compras públicas innovadoras y articulación con sectores demandantes sofisticados.

6. Desarrollo e implementación de aplicaciones

En América Latina y el Caribe, el eslabón de desarrollo, despliegue e implementación de aplicaciones de IA presenta una base empresarial más amplia que los segmentos aguas arriba de la cadena. Esta posición se sustenta en capacidades acumuladas en desarrollo de software, servicios digitales, consultoría tecnológica, integración de sistemas y adaptación de soluciones a necesidades sectoriales. No obstante, dentro de este eslabón conviene distinguir dos subsegmentos con dinámicas económicas y tecnológicas diferentes: la integración e implementación de tecnologías de IA en organizaciones usuarias, y el desarrollo de aplicaciones y soluciones propias de IA.

El primer subsegmento corresponde a la integración, consultoría e implementación empresarial de soluciones de IA. En este espacio participan empresas regionales de servicios tecnológicos, firmas globales de consultoría, integradores de sistemas y proveedores especializados que acompañan a grandes organizaciones en la identificación de casos de uso, adaptación de modelos y plataformas de terceros, integración con sistemas legados, rediseño de procesos, gestión del cambio organizacional y despliegue de aplicaciones en entornos productivos.

La distribución de herramientas y empresas de IA en diferentes ámbitos de aplicaciones permite observar la estructura del eslabón en América Latina. Aunque algunas herramientas de IA operan en múltiples dominios, cada herramienta o empresa se categoriza según su área de aplicación principal o el ámbito en el que está más activa. (ver cuadro 22).

Cuadro 22
Países de America Latina: distribución de herramientas y empresas de IA
(En porcentajes)

Dominio principal	Brasil	Chile	Colombia	México	Uruguay	Total
Consultoría y desarrollo de herramientas	26,7	26,5	44,0	39,1	39,2	31,5
Finanzas y Seguros	7,1	5,5	5,7	8,2	3,3	6,7
Atención al cliente y CRM	6,9	3,9	7,1	5,1	4,6	6,1
Ciencia de datos y Análisis	6,0	7,5	6,4	4,2	4,6	5,9
Sanidad y Ciencias de la Vida	5,9	5,5	4,1	3,4	6,7	5,2
Marketing y publicidad	5,4	2,0	3,4	4,3	4,6	4,5
Seguridad y vigilancia	3,8	5,8	3,4	3,6	1,3	3,9
Productividad y colaboración	3,7	3,9	2,5	2,8	7,1	3,6
Recursos Humanos y Desarrollo Profesional	2,6	3,6	2,1	4,1	2,5	2,9
Fabricación e Industria	2,8	5,0	1,6	2,5	2,1	2,9
Educación y Aprendizaje	3,5	3,4	1,6	1,5	1,7	2,8
Legal y Cumplimiento	3,3	2,7	0,9	2,3	2,9	2,7
Logística y cadena de suministro	2,4	2,0	2,3	2,5	2,5	2,4
Agricultura, Acuicultura y Alimentación	2,5	2,4	1,6	1,7	4,2	2,3
Comercio minorista y comercio electrónico	2,7	2,5	1,4	1,3	3,8	2,3
Otros dominios	14,7	17,7	11,9	13,4	9,2	14,2

Fuente: Weltevreden, J.W.J., & Kakanis, D. (2026). Informe sobre Ecosistemas de IA de América Latina 2026: Un análisis comparativo de los paisajes de IA en Brasil, Chile, Colombia, México y Uruguay. Ámsterdam: RankmyAI y Universidad de Ciencias Aplicadas de Ámsterdam.

El panorama de la IA en América Latina está dominado por empresas de Consultoría y Desarrollo de Herramientas. En los cinco países, este es el dominio más grande, lo que demuestra que gran parte del ecosistema regional de IA sigue centrada en ayudar a las organizaciones a implementar soluciones de IA, desarrollar aplicaciones personalizadas y apoyar proyectos de transformación digital, en lugar de ofrecer productos de IA independientes.

Más allá de la Consultoría y Desarrollo de Herramientas, los sectores más fuertes de la región son aplicaciones en Finanzas y Seguros, Atención al Cliente y CRM, Ciencia de Datos y Analítica, Sanidad y Ciencias de la Vida, y Marketing y Publicidad. Estos dominios reflejan áreas donde la IA ya cuenta con muchas aplicaciones empresariales prácticas, incluyendo automatización de flujos de trabajo, detección de fraude, atención al cliente, predicción, personalización y análisis de datos.

A nivel de país, Brasil y Chile presentan los ecosistemas de IA más diversificados de la región. Brasil cuenta con una sólida representación en Finanzas y Seguros, Atención al Cliente y CRM, Ciencia de Datos y Analítica, Sanidad y Ciencias de la Vida, Marketing y Publicidad, Educación y Aprendizaje, Legal y Cumplimiento, Retail y Comercio Electrónico, Ventas, Manufactura e Industria, y Agricultura, Acuicultura y Alimentación. Chile también demuestra actividad en numerosos ámbitos, pero destaca más en Ciencia y Análisis de Datos, Seguridad y Vigilancia, Manufactura e Industria, Energía y Servicios Públicos, y Monitorización Ambiental.

Colombia presenta una concentración mucho más fuerte en Consultoría y Desarrollo de Herramientas que los demás países analizados. Muchas empresas colombianas de IA parecen centrarse en desarrollar soluciones de IA personalizadas y apoyar proyectos de adopción de IA para clientes. Fuera de la Consultoría y Desarrollo de Herramientas, los principales ámbitos son Atención al Cliente y CRM, Ciencia de Datos y Analítica, Finanzas y Seguros, y Salud y Ciencias de la Vida.

México exhibe un ecosistema de IA más equilibrado tras Consultoría y Desarrollo de Herramientas, con una representación más fuerte en Finanzas y Seguros, Atención al Cliente y CRM, Marketing y Publicidad, Recursos Humanos y Desarrollo Profesional, y Desarrollo y Programación. Uruguay también muestra un ecosistema relativamente diversificado a pesar del tamaño más reducido de su economía, con una mayor representación en Productividad y Colaboración, Salud y Ciencias de la Vida, Agricultura, Acuicultura y Alimentación, Retail y Comercio Electrónico, y Legal y Cumplimiento.

Varios dominios siguen siendo relativamente pequeños en todos los países analizados. Esto incluye áreas como Infraestructura de IA, Plataformas y Ecosistemas, Sostenibilidad e IA para el Bien, Servicios Públicos y Gobernanza, Transporte y Movilidad, Viajes, Turismo y Hostelería, Bienestar Personal y Salud Mental, Accesibilidad y Discapacidad, Investigación Académica y Militar y Defensa. Los datos también muestran que Media & Creación de Contenido, Escritura y Lenguaje, y Diseño & Herramientas Creativas desempeñan actualmente solo un papel limitado en la mayoría de los ecosistemas de IA latinoamericanos. En general, los datos sugieren que la región está actualmente más centrada en aplicaciones prácticas de IA empresariales, operativas y relacionadas con la industria que en herramientas de IA generativa orientadas al consumidor.

Entre los ejemplos más representativos de este eslabón, Globant, empresa de origen argentino y alcance global, constituye un caso relevante de firma regional que ha construido capacidades internacionales en servicios digitales e IA. En 2024, la empresa reportó ingresos totales por US$ 2.415,7 millones, y al tercer trimestre de 2025 registraba 29.020 trabajadores, de los cuales 27.123 correspondían a profesionales de tecnología, diseño e innovación. Globant ha orientado parte de su oferta hacia la transformación empresarial basada en IA mediante estudios, soluciones y capacidades sectoriales, incluyendo servicios financieros, industrias intensivas en procesos y aplicaciones empresariales apoyadas en alianzas con proveedores tecnológicos globales como SAP, AWS, Salesforce, Oracle, Microsoft, Google y OpenAI. Ya en 2023 Globant generó US$ 100 millones en ingresos por integración de sistemas de IA [290]. De manera similar, firmas globales como Accenture, Deloitte, EY y KPMG han fortalecido sus prácticas de IA, datos, nube y transformación digital en los principales mercados de la región. Su papel se concentra en la traducción de tecnologías de frontera en soluciones organizacionales concretas para sectores como banca, energía, telecomunicaciones, empresas de consumo minorista, líneas aéreas, manufactura, salud, servicios públicos y gobierno.

El segundo subsegmento corresponde al desarrollo de aplicaciones, plataformas o soluciones propias basadas en IA. A diferencia de la integración empresarial, donde el valor se genera principalmente mediante la adaptación de tecnologías de terceros a procesos organizacionales existentes, este subsegmento se caracteriza por la creación de productos digitales o soluciones sectoriales en los que la IA forma parte del núcleo de la propuesta de valor. En América Latina y el Caribe, este espacio incluye tanto startups especializadas como empresas digitales de mayor escala. NotCo, empresa de origen chileno, constituye un caso ilustrativo al utilizar IA para el desarrollo de alimentos de origen vegetal mediante una plataforma que combina datos sobre composición molecular, propiedades funcionales e ingredientes.

Mercado Libre y Nubank, por su parte, incorporan IA en plataformas digitales masivas para mejorar búsqueda, recomendaciones, prevención de fraude, evaluación crediticia, logística, atención al cliente y personalización de servicios. En Fintech, Nubank (Brasil, Latam) es el caso paradigmático de aplicación de IA para alimentar una estrategia segmentación crediticia en un mercado no tradicional para la banca física. Las capacidades de Nubank están basadas en el control de una base de datos sobre el comportamiento financiero de clientes latinoamericanos [291].

[290] Globant (2025). *Investors*. Disponible en: https://investors.globant.com/The-AI-opportunity-we-dare.

[291] Data intelligence company Hyperplane - Nubank, Disponible: https://cdn.nubank.com.br/newsletter-migration/campaigns/NU_HEADLINES_EN/NuHeadlines_25_EN_preview_migrated.htmlEN/NuHeadlines_25_EN_preview_migrated.html

También destacan empresas especializadas como NeuralMind en Brasil, Aivo en Argentina, Yalo en México y Rocketmat en Brasil, que desarrollan soluciones de procesamiento de lenguaje natural, agentes conversacionales, comercio conversacional o gestión de talento. Estos casos muestran que la región cuenta con capacidades para generar aplicaciones de IA en sectores diversos, aunque con una fuerte dependencia de modelos, plataformas cloud e infraestructura provistos por actores globales.

Junto con estos actores de mayor escala, existe un tejido amplio de pequeñas y medianas empresas dedicadas al desarrollo de software, automatización, analítica avanzada, agentes conversacionales, integración de APIs, soluciones verticales y consultoría especializada. Según información preliminar del estudio, en Uruguay se identifican 243 empresas que participan en la cadena de IA, de las cuales 55% tienen diez o menos empleados y 39% se posicionan en el segmento de consultoría y desarrollo de herramientas. Este patrón de atomización también se observa en economías de mayor tamaño: el mismo relevamiento identifica 2.641 empresas en Brasil y 742 en México en actividades vinculadas al desarrollo de software, herramientas, consultoría e implementación de IA (RankmyAI, 2026). Estas cifras sugieren que el eslabón de aplicaciones e implementación tiene una mayor densidad empresarial relativa que otros segmentos de la cadena, aunque con predominio de empresas de menor escala.

Esta estructura empresarial contribuye a la capilaridad de la adopción de IA en el tejido productivo regional, especialmente entre empresas medianas, gobiernos subnacionales y sectores que requieren soluciones adaptadas a contextos locales. Sin embargo, también plantea desafíos de escala, financiamiento, acceso a talento avanzado, certificaciones, infraestructura de prueba y capacidad para participar en proyectos complejos o de alcance regional. En muchos casos, las empresas locales actúan como adaptadoras, integradoras o desarrolladoras de soluciones sobre modelos, plataformas cloud e infraestructuras provistas por actores globales, lo que limita su capacidad de captura de valor en los componentes más intensivos en capital, datos y propiedad intelectual.

En consecuencia, la posición de América Latina y el Caribe en este eslabón combina oportunidades y limitaciones. La región cuenta con capacidades relevantes para adaptar, integrar y desplegar soluciones de IA, así como para desarrollar aplicaciones orientadas a problemas específicos de sus economías y sociedades. Sin embargo, gran parte de estas capacidades depende de tecnologías, modelos, plataformas cloud e infraestructuras provistas por actores globales. Por ello, el desafío de política pública no consiste únicamente en promover más aplicaciones de IA, sino en fortalecer las capacidades locales de desarrollo, integración, validación, escalamiento y gobernanza de estas soluciones, articulando empresas tecnológicas, sectores usuarios, universidades, centros de investigación, infraestructura de datos, compras públicas innovadoras y cooperación regional.

7. Los puntos de presión en la cadena de valor de inteligencia artificial en América Latina

Así como en los casos estadounidense, chino y europeo, la cadena de valor de la IA en América Latina también presenta puntos de presión y riesgo, aunque estos son diferentes de los países desarrollados.

Subinversión en modelo soberano

El presupuesto de desarrollo inicial de LatamGPT, el modelo soberano para América Latina, es de US$ 550,000 en financiamiento directo, a lo que se suma un aporte de CAF de US$ 250,000, por un total de US$ 800,000. A ello, se suma la contribución de AWS para el procesamiento subsidiado de entrenamiento de la primera versión del modelo. Estas sumas son significativamente inferiores a lo requerido para desarrollar un modelo con capacidad equivalente a algunos de los modelos soberanos. otros modelos en esta categoría se benefician de financiamientos sustancialmente más elevados

(AMALIA [292] US$ 5.5 millones; ALIA [293] US$ 45 millones; SEA-LION [294] US$ 65 millones). Si bien es cierto que la versión actual de LatamGPT está más cerca de una infraestructura para desarrolladores que de un servicio desplegado al público y que sus características son menos desarrolladas (sin multimodalidad, falta de familia amplia de modelos, ausencia de pilotos públicos), el hecho es que la inversión debe ser aumentada sustancialmente para alcanzar un lanzamiento público. En este momento no existen fondos comprometidos para el futuro.

La inversión latinoamericana es aún inferior a las contribuciones de otras naciones en el desarrollo de modelos soberanos. Por ejemplo, el gobierno de India, apoyado por *IndiaAI Mission* y empresas privadas, está invirtiendo US$ 1.25 mil millones en un modelo de dimensión equivalente al 7B desarrollado por Mistral [295]. De manera similar, Singapur invirtió US$ 65 millones en el desarrollo de Sea-Lion, esfuerzo enfocado en el desarrollo de LLMs para la región asiática, basado en lenguas locales y construido alrededor del modelo Qwen de Alibaba.

Como fue desarrollado previamente en este documento, la cadena de valor de IA está guiada por economías de escala e integración vertical donde el volumen de inversión es clave para generar resultados viables. Esto implica que, para América Latina y el Caribe, los objetivos de desarrollo de un LLM soberano deben estar apoyados por un esfuerzo inversor más importante.

Recuadro 1
Aspectos particulares que condicionan el desarrollo de LATAMGPT

Si bien LatamGPT no busca competir con los modelos fundacionales desarrollados por los grandes laboratorios de IA, esto no implica que no deba estar financiado para su desarrollo.

Si bien en su primera versión, el equipo recibió la colaboración de AWS, ésta ya ha terminado dado que estaba basada en un crédito que el hiperescalador había acordado con el *Data Observatory* firmado con el gobierno chileno (Ver https://www.dataobservatory.net/). El objetivo para el entrenamiento de la versión 2.0 es utilizar el supercomputador de la Universidad de Tarapacá. Sin embargo, a mediados del 2026, el equipamiento para la construcción del centro no ha sido instalado, con lo cual CENIA está experimentando ya una falta de capacidad de cómputo para continuar el entrenamiento del modelo. Adicionalmente, CENIA no recibe fondeo adicional por parte de otros gobiernos de la región, dado que la colaboración de estos se limita a ayudar en la coordinación de recursos humanos. En la actualidad, CENIA está explorando fuentes adicionales de financiamiento provenientes de otros operadores privados para continuar el trabajo de entrenamiento del modelo.

Adicionalmente, el equipo de desarrollo de LatamGPT ha enfrentado dificultades en el acceso de datos para el entrenamiento del modelo. Si bien para la primera versión se han recopilado 2.8 trillones parámetros, esto ha sido superando complejidades por la existencia de leyes restrictivas a nivel regional respecto del uso de datos. Por ello, el tiempo requerido para la obtención de licencias y digitalizar los datos para ser usados de manera legal y transparente en el entrenamiento afecta la inversión requerida para el entrenamiento del modelo. Adicionalmente, más allá de la obtención de permisos para el uso de datos, el equipo de LatamGPT debe trabajar para poner en los mismos, la "marca de agua", que limita el "web scraping" ilegal y defender los derechos de autor. Este proceso limita la capacidad que tienen otros desarrollos de modelos soberanos como los europeos.

Fuente: Elaboración propia.

Adicionalmente, LatamGPT representa el resultado del trabajo de más de 100 profesionales, alojados en 60 instituciones, distribuidos en 15 países [296]. La experiencia de proyectos nacionales como

[292] Modelo LLM en portugués, diseñado para la administración pública, la ciencia y aplicaciones de código abierto.
[293] Primer modelo europeo de IA en castellano y lenguas cooficiales.
[294] Familia de LLMs abiertos para el Sudeste Asiático, basado en modelos fundacionales abiertos Gemma, Llama y Qwen.
[295] Ghosh, S. (2024) "India's US$ 1.25 billion push to power AI", NatureIndia (March 17). Disponible en: https://www.nature.com/articles/d44151-024-00035-5
[296] Fuente: CENIA.

los de la Unión Europea [297] demuestran que, si bien disponen de infraestructura de nivel mundial, están limitados en su desempeño por fallos de coordinación, acceso limitado a capital de riesgo, capacidad limitada en desarrollo de producto y comercialización, entre otros factores [298].

Fragmentación de la cadena de valor

De manera similar a la cadena europea, el desarrollo de la IA en América Latina implica la necesidad de enfrentar la complejidad de fragmentación institucional lo que reduce el efecto que puede tener los esfuerzos de inversión y desarrollo encarados por cada país.

Adicionalmente, más allá de Brasil, México y Chile, que poseen una capacidad autosuficiente a lo largo de la cadena de valor, el resto de los países deben buscar la especialización vertical, afinando modelos y construyendo aplicaciones para sectores verticales donde la región tiene datos y ventajas comparativas, como ser en agricultura, biodiversidad, minería y Fintech.

En este contexto, la fuga de talento especializado continúa siendo un obstáculo crítico. Mientras Globant y las consultoras capacitan a miles, los salarios en dólares del trabajo remoto para empresas estadounidenses drenan el talento senior de los startups locales. Por lo tanto, iniciativas educativas y la creación de centros de ingeniería de alto nivel deben contribuir a retener una masa crítica de investigadores y desarrolladores que son la base del ecosistema futuro.

Limites en la capacidad de disponibilidad de recursos humanos

America Latina y el Caribe todavía están experimentando limitaciones en la disponibilidad de capital humano. Esta conclusión requiere un análisis más profundo para resaltar diferentes aspectos contradictorios.

Por el lado positivo, la región registra un aumento importante en instituciones de formación de profesionales de IA. En la versión del 2025 del Índice Latinoamericano de Inteligencia Artificial se documenta que 17 de los 19 países de la región tienen programas de magíster en IA, con 179 programas en total, mientras que se en el último año se duplicó el número de países con programas de doctorado en IA, de 4 a 8, con la incorporación de Argentina, Colombia, Paraguay y Perú y 29 programas en total [299].

Sin embargo, considerando el posicionamiento internacional de las instituciones de formación, la región se encuentra distanciada significativamente de los lideres mundiales. Independientemente de la confiabilidad de rankings de instituciones de formación, la región demuestra unas pocas instituciones en relación con los grandes polos de desarrollo de la IA (ver cuadro 23).

Cuadro 23
Posicionamiento en el ranking de instituciones de formación de IA, 2025

País	Número de instituciones	Autores	Ranking de publicaciones	Índice de IA[a]
Estados Unidos	187	3 337	38 349,6	3 151,59
China	33	1 097	21 737,4	1 146,88
Reino Unido	43	970	7 742,1	738,86
Alemania	71	592	7 045,3	586,15
Canadá	30	473	4 362,3	341,47

[297] European Commission (2025). *A pionering AI Project awarded for opening Large Language Models to European languages I Shaping Europe's digital future*. Disponible en: strategy.ec.europa.eu/en/news/pioneering-ai-project-awarded-opening-large-larnguage-models-european-languages.

[298] Ver Tan, J., Jackson, B., Berjon, R. and Coyle, D. (2025). *Airbus for AI: a global strategy for value creation*. Bennett School for Public Policy, University of Cambridge. Disponible en: https://publicai.co/airbus-for-ai.pdf

[299] Soto, Á., Durán, R., Moreno, A., Adasme, S., Rovira, S., Jordán, V. y Poveda, L. (Coords.) (2026). Índice Latinoamericano de Inteligencia Artificial (ILIA) 2025. Hallazgos principales, IA aplicada y talento humano. *Documentos de Proyectos* (LC/TS.2026/1). Comisión Económica para América Latina y el Caribe y Centro Nacional de Inteligencia Artificial, pag. 36.

País	Número de instituciones	Autores	Ranking de publicaciones	Índice de IA[a]
Australia	13	337	4 119,6	269,80
Singapur	5	161	3 187,7	181,02
Corea	9	197	3 146,2	157,10
India	25	244	1 480,8	142,20
Israel	8	181	1 903,6	133,50
Brasil	14	147	343,3	27,56
Chile	1	8	20,7	1,44
Argentina	2	12	14,1	1,25
Colombia	1	3	5,5	0,35

Fuente: AICPB. AI Rankings.
[a] Incluye las siguientes áreas: Computer Vision, Natural Language, Machine Learning, Cognitive Reasoning, Robotics, and Multi-Agent Systems. In addition, we also consider two areas: Simulation for training AI agents in graphics environments like VR and AR; and AI In General.

Por otra parte, a pesar del crecimiento en los programas de formación, las tasas de crecimiento de talento en IA indican que la brecha entre Latinoamérica y los lideres globales se está ampliando.[300] Esta tendencia se acentúa con la migración de talento: de acuerdo con el Índice Latinoamericano de Inteligencia Artificial, la mayor parte de los países de la región experimentan una pérdida neta de profesionales en IA[301]. Esta situación contrasta con la de los polos de innovación globales, que muestran un flujo neto positivo, lo que representa un obstáculo estructural para el desarrollo de la IA en la región.

B. Dimensionamiento económico de la cadena de valor de la inteligencia artificial en América Latina

La organización industrial y el dimensionamiento económico de la cadena de valor de la IA en América Latina y el Caribe presentan divergencias sustanciales frente a los mercados de frontera como Estados Unidos y China. Mientras que los ecosistemas avanzados exhiben una alta intensidad de capital y una fuerte concentración de valor en el desarrollo de GPUs y modelos fundacionales, la inserción regional sigue una dinámica asimétrica en términos de eslabones generadores de mayor valor.

En términos del análisis, a diferencia del estudio dedicado a Estados Unidos en el capítulo II, el dimensionamiento económico de la cadena de valor latinoamericana se ve limitado por una menor disponibilidad de datos e información pública detallada, lo que no permite realizar el dimensionamiento por métrica, como fue hecho en el capítulo anterior. En consecuencia, para evaluar los mecanismos de captura de valor en la cadena latinoamericana, este capítulo replica la taxonomía de los seis eslabones fundamentales: 1) Equipamiento, 2) Infraestructura, 3) Preparación y gestión de datos, 4) Modelos fundacionales, 5) Plataformas de software y herramientas de IA, y 6) Desarrollo e implementación de aplicaciones. Mediante este análisis, se evidenciará cómo la región está posicionada en un desarrollo embrionario en los eslabones iniciales de fabricación y un papel más activo en los estadios superiores de la cadena de valor, enfocados en la consultoría, la integración tecnológica y la resolución de problemas corporativos verticales.

1. Equipamiento

En el eslabón base de la cadena de valor, correspondiente al diseño, manufactura y ensamblaje de servidores para IA, América Latina y el Caribe presenta una profunda asimetría estructural respecto a

[300] Soto, Á., Durán, R., Moreno, A., Adasme, S., Rovira, S., Jordán, V. y Poveda, L. (Coords.) (2026). Índice Latinoamericano de Inteligencia Artificial (ILIA) 2025. Hallazgos principales, IA aplicada y talento humano. *Documentos de Proyectos* (LC/TS.2026/1). Comisión Económica para América Latina y el Caribe y Centro Nacional de Inteligencia Artificial, pag. 84.
[301] Países como Argentina y Venezuela muestran una tendencia de pérdida neta acentuada.

los mercados avanzados. A diferencia de Estados Unidos, que concentra la propiedad intelectual del diseño de GPUs, la región participa casi exclusivamente en la etapa final de ensamblaje físico de servidores y dispositivos de *edge*, traccionada fuertemente por dinámicas de relocalización de cadenas de suministro y mitigación de riesgos geopolíticos.

La región carece en la actualidad de capacidad para competir en el diseño de chips de frontera para IA o en la manufactura avanzada de GPUs, segmentos liderados globalmente por firmas como NVIDIA, AMD y TSMC. Si bien existen esfuerzos gubernamentales recientes para integrarse a la cadena de valor de los semiconductores, estos no apuntan a los componentes de alta densidad requeridos para la IA generativa. Por ejemplo, en México se anunció recientemente la creación del Centro Nacional de Diseño de Semiconductores "Kutsari", una iniciativa con sedes en Jalisco, Puebla y Sonora enfocada en capacitar talento local y conectar el mundo académico con la industria mediante el uso de kits de diseño comercial [302]. Por otro lado, en Costa Rica, operan polos de ensamblaje, evidenciados por la reciente inversión de US$ 1,200 millones de Intel para modernizar sus instalaciones de ensamblaje, prueba e investigación y desarrollo [303]. No obstante, estas operaciones actúan como centros de empaquetado y validación, y no incluyen la fundición de los microprocesadores de frontera.

Es estrictamente en el segmento de ensamblaje de servidores donde la región logra una cierta captura de valor industrial. Impulsados por la necesidad estadounidense de establecer cadenas de suministro resilientes y con menor dependencia de Asia, los grandes fabricantes por contrato están desplegando capital en territorio latinoamericano. El caso más importante es el de la empresa Foxconn, la cual se encuentra construyendo en Guadalajara, México, una fábrica destinada a ensamblar los servidores equipados con los superchips GB200 de la familia Blackwell de NVIDIA 304. Este proyecto industrial se suma a los US$ 500 millones ya invertidos por la empresa en el país, y tiene como objetivo abastecer la demanda de los nuevos centros de datos norteamericanos. De este modo, aunque la región no captura los elevados márgenes de rentabilidad derivados del diseño de semiconductores, su competitividad manufacturera y la geografía le permite una oportunidad como sustituto o complemento de Asia.

2. Infraestructura de cómputo de la IA

El eslabón de infraestructura computacional en América Latina y el Caribe ha experimentado una expansión acelerada, traccionada por la convergencia entre la adopción corporativa de servicios en la nube y los crecientes requerimientos algorítmicos de la IA. Como fue detallado en el capítulo 10, a nivel agregado, la región transitó de operar 95 centros de datos en 2015 a consolidar 475 instalaciones para el año 2026. En cuanto a la infraestructura crítica para IA, el despliegue de zonas de hiperescaladores contabiliza 53 instalaciones en la región. Esta métrica reflejaba una alta concentración geográfica, favoreciendo a Brasil (21 zonas), Chile (13) y México (17) [305]. El incremento de estos despliegues a partir de 2019 responde a la convergencia entre la adopción corporativa de servicios de la nube y el aumento en los requerimientos de procesamiento algorítmico de IA.

Las operaciones del mercado de nube continúan siendo lideradas por proveedores estadounidenses —tales como AWS, Microsoft Azure, Google Cloud y Oracle—, a los que se han integrado firmas asiáticas como Huawei Cloud y Tencent Cloud. Simultáneamente, se identifica la aparición de proveedores de servicios de nube locales como Scala Data Center, de Brasil, ODATA (propiedad de *Aligned Data Centers*, pero con origen regional), y Kio Networks.

---

[302] Fuente: https://www.inaoep.mx/kutsari

[303] Fuente: https://procomer.com/intel-reafirma-su-compromiso-con-costa-rica-como-un-actor-clave-en-la-creciente-demanda-global-de-semiconductores/

[304] Fuente: https://www.redusers.com/noticias/foxconn-instalara-planta-mas-grande-mexico/ y https://mexicoindustry.com/noticia/foxconn-construira-en-mexico-la-fabrica-de-superchips-de-nvidia-mas-grande-del-mundo

[305] Fuente: Telegeography.

Los ingresos totales de los proveedores de servicios de la nube en la región alcanzan aproximadamente US$ 50 mil millones, de los cuales US$ 1.263 mil millones representan ventas de plataformas de IA (ver cuadro 24).

Cuadro 24
Ingresos de proveedores de infraestructura en América Latina, 2025
(En dólares)

País	Ingresos totales de servicios de nube	Ingresos por ventas de plataformas de IA
Argentina	3 116	71
Brasil	22 846	589
Chile	3 283	89
Colombia	3 839	94
México	10 937	262
Perú	2 157	48
Resto de América Latina	4 435	110
Total	50 163	1 263

Fuente: Katz, R., Jung, J. (2024). *Economic impact of cloud computing and artificial intelligence in Latin America*. New York: Telecom Advisory Services. Disponible en: https://www.teleadvs.com/economic-impact-of-cloud-computing-and-artificial-intelligence-in-latin-america/

En términos de la inversión de capital, los anuncios realizados entre el 2022 y el 2025 permiten estimar valores al 2025. Las inversiones anunciadas por los principales proveedores de servicios de nube en el periodo 2023-2025 consolidan un valor proyectado que supera los US$ 20,600 millones en infraestructura crítica, aunque parte de esas inversiones son anuncios de proyectos que se extienden hasta el 2040. Destacan por su magnitud y horizonte temporal, los planes de AWS en México y Chile, que conjuntamente representan US$ 9,000 millones, y la apuesta de Microsoft en Brasil y Chile, que suma más de US$ 6,000 millones. Si bien estas inversiones confirman la centralidad de Brasil, México y Chile, también evidencian la aparición de nuevos nodos estratégicos en el Cono Sur, como lo demuestra el reciente despliegue de Google en Uruguay (ver cuadro 25).

Si bien no se dispone de valores regionales en términos de costos operativos, y rentabilidad, el último balance publicado por Scala Data Networks reportó OPEX por $ 70.65 millones (63.42% de los ingresos), y un margen de EBITDA del 36.73%. (ver cuadro 26).

De manera similar, en el 2023 Kio Networks presentó márgenes EBITDA superiores al 33% [306] Esta asimetría entre ingresos e inversión es característica de los nodos de control físico de la cadena de valor, donde se prioriza la captura de capacidad instalada y potencia energética sobre la rentabilidad neta inmediata.

Cuadro 25
Inversiones en América Latina en el despliegue de centros de datos por proveedores de servicios de la nube, 2023-2025
(En millones de dólares)

País	Empresa	2023	2024	2025
Brasil	Microsoft Azure		2 700 (2024-2027)[a]	
	AWS		1 800 (2024-2034)[b]	
	Oracle		8[c]	
Chile	Oracle Cloud	100[d]		
	Microsoft Azure	317[e]		3 300 (2025-2029)[f]
	AWS			4 000 (2025-2040)[g]

[306] Fuente: https://www.spglobal.com/ratings/en/regulatory/article/-/view/sourceId/12652684

País	Empresa	2023	2024	2025
Colombia	Oracle	100[h]		
México	AWS			5 000 (2025-2040)[j]
	Microsoft Azure		1 300 (2024-2026)[j]	
Uruguay	Google Cloud		850 (2024-2026)[k]	
Regional	Google Cloud	1 200 (2022-2026)		

Fuente: Recopilación de los autores de fuentes públicas.
Nota: Si las inversiones son anunciadas, el valor es distribuido linealmente.
[a] Fuente: https://news.microsoft.com/source/latam/company-news-pt-br/microsoft-announces-14-7-billion-reais-investment-over-three-years-in-cloud-and-ai-infrastructure-and-provide-ai-training-at-scale-to-upskill-5-million-people-in-brazil/
[b] Fuente: https://dig.watch/updates/aws-announces-major-investment-plan-for-brazil
[c] Fuente: https://neofeed.com.br/negocios/oracle-esta-com-cheque-pronto-para-investir-em-ia-no-brasil-mas-ha-uma-barreira-no-caminho/
[d] Fuente: Presario Hub: Entrevista a VP de Telecomunicaciones de Oracle América Latina. https://www.prensariohub.com/oracle-abre-nueva-region-cloud-en-mexico/
[e] Fuente: Datacenter Dynamics. https://www.datacenterdynamics.com/en/news/microsoft-files-plans-for-chilean-data-center-region/
[f] Fuente: Microsoft. https://news.microsoft.com/es-xl/microsoft-inaugura-su-primera-region-de-datacenters-en-chile-para-acele rar-la-innovacion-y-el-desarrollo-economico-local
[g] Fuente: Diario Financiero. https://www.df.cl/empresas/telecom-tecnologia/amazon-web-services-anuncia-inversion-de-us-4-000-millones-en-santiago
[h] Fuente: El Colombiano. https://www.elcolombiano.com/negocios/oracle-trae-region-nube-publica-a-colombia-con-100-millones-de-dolares-JC23300911#
[i] Fuente: Gobierno de México. https://www.proyectosmexico.gob.mx/inversion-de-5-mmdd-de-amazon-web-services-en-mexico/
[j] Fuente: Microsoft. https://news.microsoft.com/es-xl/microsoft-announces-1-3-billion-usd-investment-in-cloud-and-ai-infrastructu re-supporting-inclusive-growth-through-technology-and-skilling-programs-in-mexico
[k] Fuente: https://www.infobae.com/america/america-latina/2024/08/30/google-confirmo-inversion-de-usd-850-millones-para-con struir-en-uruguay -el-segundo-data-center-de-la-region/

Cuadro 26
Dimensionamiento económico de Scala Data Centers, 2024
(En millones de dólares)

Métrica	Valor estimado	Fuente
Ingresos	111,39	Memoria y Balance 2024[a]
OPEX	70,65 (63,42% de los ingresos)	Memoria y Balance 2024
Margen EBITDA	36,73%	Memoria y Balance 2024, con un EBITDA de R$ 245,464 e ingresos de R$ 668,340
CAPEX	775,71 (696% los ingresos)	Memoria y Balance 2024
Valuación Estimada	4 000	Basado en el proceso de venta estratégica liderado por Deutsche Bank[b]

Fuente: elaboración propia sobre la base de Memoria y Balance de la empresa.
[a] Fuente: https://estadaori.estadao.com.br/wp-content/uploads/2025/04/scala-data-centers-sa-balanco-2025-04-09_22-42-38.pdf.
[b] Fuente: https://ionanalytics.com/insights/infralogic/digitalbridge-awaits-better-valuation-for-scala-data-centers-sale/.

En términos de capitalización de mercado, si bien ODATA no reporta información financiera, en el 2023 la misma tenía una valuación de US$ 1,850 millones[307]. El cuadro 26 reporta los montos compilados para cada una de las cinco métricas que permiten dimensionar el eslabón de infraestructura.

En conclusión, la infraestructura computacional en América Latina y el Caribe atraviesa un periodo de expansión, donde la rápida construcción de centros de datos responde a una necesidad crítica de capacidad de procesamiento local. Si bien el eslabón de infraestructura demuestra una importante presencia de operadores extranjeros, el desarrollo de la base material del ecosistema está siendo liderado por actores regionales que comprometen inversiones para garantizar la disponibilidad

[307] Fuente: https://www.halconesypalomas.com/2023/06/26/por-us1-850-millones-patria-investimentos-concreto-venta-a-estadounidense-aligned-data-centers-de-odata-incluye-colombia/

energética y técnica. Las empresas locales presentan márgenes de entre el 33% y el 37%, niveles inferiores que los alcanzados por las empresas de la misma etapa de la cadena de valor del mercado de Estados Unidos analizados en el capítulo II.

Cuadro 27
América Latina: Métricas de la cadena de valor de inteligencia artificial

Métrica	Total
Ingresos (2025)	Ingresos de nube: US$ 50.61 mil millones Ingresos de plataformas de IA: US$ 1.62 mil millones
OPEX	Scala Data Centers: 63.42% de ingresos (2024)
Inversión	US$ 3,018 millones
Rentabilidad (EBITDA)	Kio Networks: 33% (2023) Scala Data Center: 36.73% (2024)
Capitalización de mercado	ODATA: US$ 1.85 mil millones (2023) Scala Data Centers: US$ 4 mil millones (2024)

Fuente: Elaboración propia.

3. Preparación y gestión de datos

A diferencia del mercado estadounidense, donde este eslabón ha dado lugar a "pure plays" de escala como Scale AI, en América Latina y el Caribe, la preparación y gestión de datos opera bajo una lógica de servicios de capital humano intensivo y como un componente integrado en proyectos de soberanía tecnológica. Con relación al primer punto, la región se ha posicionado como un centro para servicios de etiquetado de datos, con empresas globales como TELUS International, expandiendo sus operaciones de soluciones de datos de IA en países como El Salvador y Guatemala para aprovechar la competitividad de costos y la diversidad lingüística [308].

Por otro lado, existen proyectos más orientados a la soberanía tecnológica. La captura de valor estratégico en este eslabón reside en la creación de datasets soberanos que mitiguen los sesgos cognitivos y lingüísticos de los modelos desarrollados en Estados Unidos o China. El esfuerzo más representativo es el proyecto LatamGPT, liderado por el CENIA en Chile. Paralelamente, en Brasil, la firma Maritaca AI utiliza un enfoque de Curriculum Learning sobre corpus brasileños para entrenar sus modelos, demostrando que la curaduría local permite superar el rendimiento de modelos globales en exámenes de dominio específico del área legal. Ambos proyectos serán desarrollados con mayor detalle en la próxima sección de modelos fundacionales, ya que los mismos integran ambas etapas de la cadena de valor.

En conclusión, la gestión de datos en América Latina se define por dos modelos: 1) el suministro de mano de obra para el etiquetado global y 2) la estructuración de información con valor cultural y legal propio. El valor económico del segundo modelo, no se mide por ingresos independientes sino por su rol como insumo para el desarrollo de modelos. La sostenibilidad de este eslabón dependerá de la integración de estos datasets en los flujos de trabajo corporativos, permitiendo a las empresas regionales operar con costos de inferencia reducidos, tema que será abordado en la siguiente sección de modelos fundacionales.

4. Modelos fundacionales

Bajo el paradigma de la autonomía tecnológica, América Latina y el Caribe han iniciado el desarrollo de modelos soberanos adaptados a las particularidades lingüísticas del castellano y el portugués. Esta transición responde a la necesidad de mitigar los sesgos cognitivos detectados en modelos desarrollados en Estados Unidos o China.

[308] Fuente: https://www.elsalvador.com/h-noticias/h-negocios/telus-inaugura-torre-emblema-con-una-inversion-de-29-2-millones-dolares/1125032/2024/

En este contexto, la región presenta dos enfoques divergentes para la construcción de modelos soberanos. En el ámbito privado, destaca la firma brasileña Maritaca AI, la cual emplea una arquitectura basada en Curriculum Learning y especialización de dominio sobre información legal nacional. Su modelo Sabiá ha demostrado un rendimiento equivalente o superior a GPT-4 en tareas específicas del idioma portugués y exámenes estandarizados como el de la Orden de Abogados de Brasil [309]. Más allá de la precisión técnica, esta iniciativa aborda el tema económico, ya que logra reducir los costos por token respecto a los modelos de frontera globales. Esto le permite ofrecer una oportunidad para proyectos de las etapas superiores de la cadena de valor de la IA que consumen tokens. Finalmente, la empresa no reporta datos financieros y tampoco posee una valuación pública de mercado (en el 2025 tuvo una oferta para adquirir el 25% de la empresa, pero no trascendieron los valores) [310].

Paralelamente, el proyecto LatamGPT, liderado por el CENIA (Chile), constituye el esfuerzo colaborativo público-privado más ambicioso de la región. Lanzado en junio de 2025, el modelo integra a 33 instituciones latinoamericanas para desarrollar una arquitectura de código abierto basada en Llama 3.1 [311]. El entrenamiento se sustenta en más de 70 mil millones de parámetros en 8 terabytes de datos regionales y se orienta a aplicaciones de política pública y la preservación de lenguas no mayoritarias como el mapuche y el guaraní [312]. A pesar de estos avances, la adopción de IA en la región continúa supeditada a plataformas internacionales, cuya dominancia establece una infraestructura difícil de sustituir.

La principal barrera para la consolidación de estos esfuerzos radica en una subinversión estructural. Mientras que el entrenamiento de modelos fundacionales en mercados desarrollados se apoya en inyecciones masivas de capital concentrado (ver capítulo de modelos fundacionales de la sección de Estados Unidos), el presupuesto inicial de LatamGPT sumó US$ 3 millones, de los cuales US$ 550, provenían de financiamiento directo y aportes del Banco de Desarrollo de América Latina y el Caribe (CAF) [313]. Esta cifra resulta reducida al contrastarla con la ronda de US$ 2,000 millones completada por la europea Mistral AI en septiembre de 2025 para sus modelos [314]. De igual forma, otras economías emergentes presentan compromisos superiores para el desarrollo de modelos soberanos: la India AI Mission invierte US$ 1,250 millones [315].

Finalmente, la fragmentación institucional impone una complejidad operativa que ralentiza la agilidad comercial. La coordinación de más de 100 profesionales en 15 países dentro de LatamGPT corre el riesgo de enfrentar fallos de gobernanza. Para que América Latina logre resultados viables, los objetivos de soberanía digital deben acompañarse de un esfuerzo inversor.

5. Plataformas de software y herramientas

En el eslabón de plataformas de software y herramientas, América Latina y el Caribe tienen una oportunidad estratégica para posicionarse como un exportador de propiedad intelectual de alto valor, operando bajo un modelo *asset-light* que se adapta a las restricciones de capital de la región. A diferencia de las capas de infraestructura física, este segmento depende primordialmente del talento de ingeniería y la capacidad de orquestación algorítmica. La región ha visto el surgimiento de plataformas especializadas que, aunque utilizan modelos fundacionales globales como base, desarrollan herramientas para la optimización de flujos de trabajo corporativos.

---

[309] Fuente: https://arxiv.org/abs/2603.10213
[310] Fuente: https://fusoesaquisicoes.com/acontece-no-setor/jusbrasil-investe-na-maritaca/
[311] Fuente: https://www.gob.cl/en/news/LatamGPT-learn-about-the-latin-american-ai-model-developed-in-chile/
[312] Fuente: https://www.techbuzz.ai/articles/latin-america-builds-rival-to-chatgpt-with-50b-parameter-model
[313] Fuente: https://www.rionegro.com.ar/tecnologia/LatamGPT-el-modelo-regional-que-busca-reducir-sesgos/
[314] Fuente: https://news.crunchbase.com/venture/europe-largest-ai-round-mistral-seriesc/
[315] Fuente: https://www.nature.com/articles/d44151-024-00035-5

Brasil se ha consolidado como el polo de mayor dinamismo en el desarrollo de herramientas de IA conversacional y automatización de la experiencia del cliente. La empresa Blip lidera este segmento, habiendo asegurado una inversión estratégica de US$ 60 millones de Warburg Pincus en 2024 para escalar su plataforma de "AI-First Conversational"[316]. Su plataforma permite a las empresas integrar modelos de lenguaje en canales de mensajería masivos, actuando como una capa de orquestación que gestiona la lógica de negocio y la seguridad de los datos[317]. La empresa ha reportado ingresos de US$ 100 millones anuales[318] en el 2022 y expandiendo su negocio a 40 países, con lo que logró aumentar los ingresos a $170 millones en el 2025[319]. Esta expansión lo logró con márgenes positivos[320], logrando dejar atrás la etapa de consumo de efectivo que caracteriza a los startups.

Un caso distintivo de Plataforma como Servicio en la región es el desarrollado por la empresa chilena The Not Company (NotCo). A través de su plataforma de IA denominada Giuseppe, la firma ha trascendido la producción de alimentos para licenciar su software a nivel global[321]. Giuseppe opera como una herramienta de descubrimiento molecular y formulación química que permite a los equipos de I+D acelerar el desarrollo de productos basados en plantas, ejemplificando cómo una herramienta de IA vertical nacida en la región puede escalar mediante un modelo de licenciamiento de software tradicional[322]. La empresa tiene actualmente una valoración de US$ 1,500 millones[323], con ingresos para el 2025 de US$ 170 millones[324].

En México, el enfoque se ha centrado en plataformas de orquestación para el comercio conversacional, donde Yalo (Yalochat) destaca como un actor predominante. En 2025, la empresa consolidó su plataforma de IA generativa para el sector de consumo masivo, permitiendo a las grandes distribuidoras automatizar la gestión de ventas y logística mediante agentes inteligentes. Esta herramienta actúa como un sistema operativo para el comercio en mercados emergentes, integrando capacidades de visión computacional y modelos de lenguaje para optimizar la toma de pedidos en tiempo real[325]. La empresa alcanzó ingresos anuales por US$ 150 millones en el 2024[326], sin reportar información financiera adicional.

Asimismo, empresas como Aivo (Argentina), ahora integrada en el ecosistema global tras su adquisición por parte de Engageware, han marcado el camino para las herramientas regionales de servicio al cliente autónomo. Su plataforma de orquestación de agentes permite a las instituciones financieras y de servicios en América Latina implementar asistentes con capacidades de razonamiento complejo, manteniendo la soberanía sobre el entrenamiento de los datos de los usuarios finales y operando con latencias optimizadas para la infraestructura regional[327].

Por último, en el segmento de la gestión de talento humano, destaca la brasileña Rocketmat. Esta plataforma utiliza algoritmos propios para la reducción de sesgos en el reclutamiento y la optimización de planes de carrera, resolviendo problemas de cumplimiento legal y cultural que las plataformas genéricas suelen omitir[328].

---

316 Fuente: https://techla.pro/2024/11/18/blip-recauda-us60-millones-en-ronda-liderada-por-softbank-y-microsoft/
317 Fuente: https://www.blip.ai/es/
318 Fuente: https://warburgpincus.com/2022/06/07/take-blip-announces-a-us70-million-series-b-financing-round/
319 Fuente: https://www.computing.es/mercado-ti/blip-duplicara-su-plantilla-en-espana-con-la-ia-conversacional-como-eje-estrategico/
320 Fuente: https://abes.org.br/en/para-alem-do-breakeven-blip-avanca-estrategia-de-negocio-gerando-caixa/
321 Fuente: https://www.foodbev.com/news/magnum-ice-cream-partners-with-notco-to-innovate-product-development-using-ai
322 Fuente: https://notco.com/cl/sobre/giuseppe
323 Fuente: https://nationaltoday.com/us/ca/san-francisco/news/2026/03/22/ai-startup-notco-helps-food-giants-future-proof-their-products/
324 Fuente: https://businessmodelcanvastemplate.com/products/notco-swot-analysis
325 Fuente: https://www.yalo.ai/blog/recomendaciones-inteligentes-para-cpg
326 Fuente: https://businessmodelcanvastemplate.com/blogs/growth-strategy/yalo-growth-strategy
327 Fuente: https://tracxn.com/d/companies/rocketmat/__90CDhJBbFYEuZHUPFPK44UwqtgICXoRX2zGInYwxpmo
328 Fuente: https://www.crunchbase.com/organization/rocketmat-ai

En conclusión, el eslabón de plataformas y herramientas representa el segmento de mayor madurez tecnológica propia en América Latina. A diferencia de los eslabones de infraestructura, donde predomina el gasto de capital extranjero, aquí la región ha logrado crear empresas que generan flujos de caja positivos y exportan propiedad intelectual. La sostenibilidad de estas plataformas radica en su capacidad para ofrecer una capa de "latinoamericanización" de la tecnología de frontera, adaptándola a las realidades transaccionales de WhatsApp y las normativas legales locales. Este ecosistema de herramientas es, en última instancia, el que habilita el siguiente paso de la cadena: la implementación masiva de soluciones verticales en la economía real, donde la consultoría y el desarrollo a medida transforman estas capacidades técnicas en valor de negocio directo.

6. Desarrollo e implementación de aplicaciones

En el último eslabón de la cadena de valor, correspondiente al desarrollo e implementación de aplicaciones, América Latina y el Caribe presentan su mayor capacidad de captura de valor económico real. Este segmento se enfoca en la resolución de problemas corporativos verticales y la personalización de modelos para sectores específicos. A diferencia de las etapas de hardware e infraestructura, este eslabón no requiere de una inversión intensiva en activos físicos, sino de una alta especialización en procesos de negocio y capital humano avanzado para la integración de sistemas.

El sector de la consultoría e integración tecnológica lidera esta fase, actuando como el puente necesario entre los modelos de frontera globales y las necesidades operativas de las empresas locales. Firmas globales con fuerte presencia regional, como Accenture, han reportado que una fracción creciente de sus ingresos ya son atribuibles directamente a la implementación de IA.

En el vertical de servicios financieros, destaca el caso de Nubank. Si bien la firma opera primordialmente como una institución bancaria digital y no como una empresa de IA en sentido estricto, ha integrado el aprendizaje automático en el núcleo de su modelo de negocio para la democratización del crédito y la prevención del fraude en tiempo real [329]. Al procesar miles de variables en milisegundos, el neobanco ha logrado servir a más de 100 millones de clientes en la región con perfiles de riesgo personalizados que la banca tradicional suele omitir [330]. Esta integración tecnológica no solo optimiza la rentabilidad operativa, sino que ha comenzado a reducir los costos de adquisición de clientes mediante el despliegue de asistentes financieros autónomos basados en IA generativa [331]. Para esta empresa, no se incluyen los datos financieros, ya que a pesar de que utiliza la IA para el desarrollo e implementación de aplicaciones, es sólo para adaptar el negocio original.

Similarmente, en el sector del comercio electrónico y la logística, Mercado Libre destaca por integrar la IA como un motor de eficiencia operativa, a pesar de no ser una empresa centrada exclusivamente en el desarrollo de software de IA. La compañía utiliza algoritmos predictivos y aprendizaje automático para anticipar la demanda de los consumidores, orquestar su red de distribución y optimizar la ubicación estratégica de inventarios en sus centros de almacenamiento [332]. Gracias a esta infraestructura tecnológica, la empresa ha logrado reducir sus tiempos de entrega con más del 50% de sus entregas realizadas en menos de 24 horas. Estas aplicaciones de IA, que incluyen la optimización de rutas y la automatización en depósitos, han mejorado la eficiencia logística, permitiendo capturar márgenes que compensan los complejos desafíos y costos estructurales del transporte en América Latina [333]. Al igual que en el análisis de Nubank, no se incluyen indicadores financieros

[329] Fuente: https://blog.nubank.com.br/nubank-artificial-intelligence-technology/
[330] Fuente: https://forbes.com.br/forbes-money/2024/05/nubank-ultrapassa-100-milhoes-de-clientes-na-america-latina/
[331] Fuente: https://www.bloomberglinea.com/negocios/modelo-de-ia-de-nubank-impulsa-expansion-del-ingreso-neto-por-intereses/
[332] Fuente: https://thelogisticsworld.com/tecnologia/hot-sale-mercado-libre-se-prepara-con-inteligencia-artificial-para-la-edicion-2024/
[333] Fuente: https://perspectivas.cohen.com.ar/articulos/tematicas-mundo-moderno-ia-y-logistica-reconectando-el-mundo

detallados para esta firma, dado que la implementación de IA actúa como una palanca de optimización de su infraestructura logística y comercial preexistente, y no como su oferta de mercado primaria.

Pasando a empresas cuyo núcleo del negocio esté ubicado en este eslabón de la cadena de valor de la IA, tenemos en el mercado brasileño a NeuralMind que se ha consolidado como un referente de "Pure Play" en el desarrollo de aplicaciones de IA de alto impacto corporativo. Fundada por académicos de la Universidad de Campinas (Unicamp), su plataforma principal, NeuralDoc, permite la automatización y el análisis inteligente de grandes volúmenes de documentos legales y financieros, logrando una reducción de costos operativos [334]. La empresa no presenta ingresos de modo independiente, pero está dentro del grupo de empresas del Parque Científico de la Unicamp que en su conjunto facturaron en 2024 US$ 25 millones [335].

En Argentina, la empresa Globant ha transformado su modelo de servicios tradicionales hacia la provisión de herramientas de productividad basadas en IA [336]. Durante 2024 y 2025, la compañía reportó que sus proyectos vinculados a la IA contribuyeron con más de US$ 350 millones a sus ingresos operativos. Globant ha desarrollado una suite de herramientas propias, incluyendo MagnifAI, una plataforma de testing visual asistido por IA, y GeneXus Next, que permite la generación automatizada de código a partir de descripciones en lenguaje natural, posicionándose como un competidor regional en el mercado de herramientas de desarrollo [337]. La empresa ha reportado en el 2025 US$ 2,455 millones totales de ingresos, de los cuales el 21.1% son de América Latina; de los mismos el 84.81% fueron destinados al OPEX (US$ 2,082 millones) y el 3.38% al CAPEX (US$ 83.1 millones), resultando en un margen EBITDA del 14.84%. No obstante, su valoración de mercado a marzo de 2026 se situó en US$ 1,857 millones, lo que representa una caída significativa desde su máximo histórico de US$ 15,000 millones. Esta contracción bursátil refleja la incertidumbre de los mercados ante el riesgo de desintermediación comercial, donde la capacidad de los propios modelos fundacionales para absorber tareas complejas de desarrollo y automatización amenaza con erosionar los márgenes de los integradores de sistemas (Ver cuadro 28).

Cuadro 28
Dimensionamiento económico de Globant, 2025
(En millones de dólares)

Métrica	Valor Estimado	Fuente
Ingresos	$ 2 455	Memoria y Balance 2025[a]
OPEX	$ 2 082 (84.81% de los ingresos)	Memoria y Balance 2025, considerando costos de ventas por US$ 1 596 y Costos Generales y Administrativos por $486
Margen EBITDA	14,84%	Memoria y Balance 2025 en base a EBITDA de US$ 364 e Ingresos de $2 455
CAPEX	$ 83,1 (3,38% de los ingresos)	Memoria y Balance 2025
Valuación estimada	$ 1 857 0,76 veces los ingresos anuales	Valuación de mercado a marzo de 2026 La empresa llego a tener una valuación de US$ 15 000[b]

Fuente: elaboración propia sobre la base de de la Memoria y el Balance de la empresa.
Nota: Considera la operatoria global de la empresa, donde América del Norte sigue siendo el principal generador de ingresos (53,8%), seguido por América Latina (21,1%), donde Argentina se mantiene como el principal polo de talento y facturación regional.
[a] Fuente: https://www.sec.gov/Archives/edgar/data/1557860/000162828026012910/glob-20251231.htm
[b] Fuente: https://www.infobae.com/economia/2026/02/12/la-accion-de-globant-volvio-a-desplomarse-y-cayo-hasta-18-cuales-son-las-variables-que-impactan-en-el-unicornio-argentino/

[334] Fuente: https://neuralmind.ai/
[335] Fuente: https://parque.inova.unicamp.br/eptv-empresas-vinculadas-a-unicamp-aumentam-33-e-faturamento-chega-a-r-129-milhoes/
[336] Esta empresa no es "Pure Play". Los productos MagnifAI GeneXus corresponden a Plataformas de Software y Herramientas. Al no poder separar los datos económicos por negocio, se incluyó en esta sección. Pero entendiendo que abarca las dos etapas de la cadena de valor.
[337] Fuente: https://www.globant.com/news/globant-launches-consumer-goods-manufacturing-ai-studio

Así también en Argentina, la consultora boutique Mutt Data se dedica a la implementación personalizada de Machine Learning y arquitectura de datos a gran escala. A diferencia de los grandes integradores generalistas, Mutt Data opera exclusivamente bajo la premisa de "Data-First", habiendo desarrollado e implementado sistemas de predicción de demanda y motores de recomendación para empresas regionales como Mercado Libre e instituciones financieras [338] [339].

Finalmente, la división de IA del grupo brasileño Stefanini, denominada Woopi, actúa como un implementador puro a escala global desde América Latina. Woopi se enfoca en el desarrollo de soluciones de IA generativa, visión computacional y asistentes virtuales inteligentes (como Sophie) para la automatización de flujos industriales y de servicios al cliente. La empresa en el 2026 va a realizar una inversión de US$ 200 millones en IA [340] [341].

En conclusión, el eslabón de desarrollo e implementación de aplicaciones constituye el epicentro de la captura de valor para América Latina y el Caribe, al transformar la infraestructura técnica global en soluciones productivas adaptadas a la realidad local. Mientras que los segmentos de hardware y modelos fundacionales operan bajo lógicas de alta concentración de capital extranjero, este segmento demuestra una vitalidad basada en el capital humano y el conocimiento de procesos verticales. La coexistencia entre integradores, empresas preexistentes que optimizan sus modelos de negocio (como Nubank y Mercado Libre) y firmas de desarrollo "Pure Play", configura un ecosistema capaz de exportar innovación.

El cuadro 31 reporta los montos compilados para cada una de las cinco métricas que permiten dimensionar esta etapa de la cadena de valor.

Cuadro 29
América Latina: métricas de eslabones de Plataformas de software y herramientas

Métrica	Total
Ingresos (2025)	Blip (Brasil): US$170 millones (2025) NotCo (Brasil): US$ 170 millones (2025) Yalo (México): US$ 150 millones (2024) Globant (Regional): US$ 350 millones en IA y US$ 2,455 totales (2025)
OPEX	Globant (Regional): 84.81% de ingresos (2024)
Inversión/Ingresos	Globant (Regional): 3,38% (2025)
Rentabilidad (EBITDA)	Globant (Regional): 14,84% (2025)
Capitalización de mercado	Notco (Brasil): US$ 1 500 millones (2025) Globant (Regional): US$ 1 857 millones (2026)

Fuente: Elaboración propia.

7. Conclusión

El análisis del dimensionamiento económico de la cadena de valor de la IA en la región revela un ecosistema en transición, caracterizado por una marcada asimetría entre la inversión en activos físicos y la generación de ingresos por propiedad intelectual, con márgenes más bajos que en el mercado estadounidense (Ver cuadro 30).

[338] Fuente: https://www.muttdata.ai/
[339] Fuente: https://pitchbook.com/profiles/company/520489-99#research
[340] Fuente: https://stefanini.com/en/insights/news/stefanini-group-leadership-and-innovation-in-the-era-of-generative-ai
[341] Fuente: https://www.globenewswire.com/news-release/2025/03/31/3051977/0/en/Stefanini-Group-focuses-on-global-leadership-with-AI-First-concept.html

Cuadro 30
Métricas por eslabón de la cadena de la IA en América Latina, 2025
(En millones de dólares)

Eslabón	Ingresos	OPEX/ ingresos	CAPEX	Margen EBITDA	Capitalización de mercado
1. Equipamiento	Sin información	Sin información	1 700[a]	Sin información	Sin información
2. Infraestructura de cómputo para IA	Nube: US$ 50.61 mil millones Plataforma de IA: US$ 1.62 mil millones	63.42%	3 018	33% 36.73%	Ascenty: US$ 1.8 mil millones (2018) ODATA: US$ 1.85 mil millones (2023) Scala Data Centers: US$ 4 mil millones (2024)
3. Preparación y gestión de datos	Sin información	Sin información	Sin información	Sin información	Sin información
4. Modelos fundacionales		Sin información	0,8[a]	Sin información	Sin información
5. Plataformas de software y herramientas	Blip (Brasil): US$170 millones (2025) NotCo (Brasil): US$ 170 millones (2025) Yalo (México): US$ 150 millones (2024) Globant (Regional): US$ 350 millones en IA (2025)	Globant (Regional): 84.81% de ingresos (2024)	Sin información	Globant (Regional): 14.84% (2025)	Notco (Brasil): US$ 1,500 millones (2025) Globant (Regional): US$ 1,857 millones (2026)
6. Desarrollo e implementación de aplicaciones			Globant (Regional): 3,38% (2025)		

Fuente: Elaboración propia.
[a] El valor no corresponde a Latam-GPT sin incluir los créditos otorgados por AWS para el entrenamiento del modelo.

A partir de los datos recopilados, se desprenden las siguientes conclusiones fundamentales sobre dónde reside el valor económico actual y potencial para América Latina:

- **Inserción incipiente en equipamiento:** la región mantiene una participación limitada en el eslabón de equipamiento, especialmente en diseño y fabricación de semiconductores avanzados, donde las barreras de capital, escala, propiedad intelectual y capacidades tecnológicas son muy elevadas. No obstante, existen oportunidades puntuales en ensamblaje, prueba, manufactura electrónica e integración de servidores para IA, particularmente en México y Costa Rica, apalancadas por procesos de "*nearshoring*" y capacidades acumuladas en electrónica y semiconductores. El desafío no es competir directamente en la frontera de chips avanzados, sino escalar hacia actividades de mayor valor agregado, como ensamblaje avanzado, integración de hardware, mantenimiento especializado, edge AI y desarrollo de proveedores vinculados a la expansión de centros de datos.
- **La Infraestructura de cómputo para IA como base del ecosistema:** la región ha consolidado una base material robusta, pasando de 95 a 475 centros de datos en una década. Las valoraciones de empresas como Scala Data Centers (US$ 4.000 millones) y ODATA (US$ 1.850 millones) demuestran que existe valor en esta etapa de la cadena, aunque este segmento opera con márgenes EBITDA (33%-37%) inferiores a los de sus pares en mercados de frontera.
- **La Brecha Crítica en Modelos Fundacionales:** la disparidad en la inversión es más aguda en el eslabón de modelos. Mientras proyectos soberanos como LatamGPT operan con presupuestos cercanos a los US$ 0.80 millones, otros modelos soberanos se benefician de

financiamientos sustancialmente más elevados (AMALIA[342] US$ 5.5 millones; ALIA[343] US$ 45 millones; SEA-LION[344] US$ 65 millones). Si bien es cierto que la versión actual de LatamGPT está más cerca de una infraestructura para desarrolladores que de un servicio desplegado al público y que sus características son menos desarrolladas (sin multimodalidad, falta de familia amplia de modelos, ausencia de pilotos públicos), el hecho es que la inversión debe ser aumentada sustancialmente para alcanzar un lanzamiento público. Esta subinversión estructural sugiere que la región no capturará valor en la creación de modelos de propósito general, sino en la "latinoamericanización" y especialización vertical (como el caso de Maritaca AI y el sector legal), donde la curaduría de datos locales genera una ventaja competitiva de nicho.

- **Captura de valor en capas intermedias y aplicaciones especializadas**: es en el eslabón de software donde la región muestra su mayor eficiencia en la captura de valor. Con empresas que ya superan la barrera de los US$ 100-150 millones de ingresos anuales (Blip, Yalo, NotCo), América Latina ha demostrado que puede exportar propiedad intelectual de alto valor. Estas empresas actúan como orquestadores que capturan el "margen intermedio" entre los modelos fundacionales y el usuario final, adaptando la tecnología a los canales de comunicación preferidos en la región (WhatsApp/Messenger).
- **La IA como Multiplicador de la Economía Real:** el mayor impacto económico no proviene necesariamente de las empresas que venden IA, sino de aquellas que la utilizan para transformar sectores tradicionales. Mercado Libre (logística) y Nubank (finanzas) han logrado optimizaciones operativas que generan márgenes de eficiencia inalcanzables sin IA. En este eslabón de Aplicación e Implementación, la región captura valor mediante la reducción de costos estructurales y la democratización del acceso a servicios.
- **El Rol del Integrador como Exportador de Talento:** firmas como Globant y Stefanini (Woopi) evidencian que el capital humano avanzado es el activo más resiliente de la cadena local. Con ingresos atribuidos a la IA que ya superan los US$ 350 millones en casos individuales y planes de inversión de US$ 200-350 millones hacia 2026, la consultoría tecnológica se posiciona como el puente indispensable que permite a la región no solo consumir IA, sino participar en la arquitectura de su implementación global.

América Latina y el Caribe capturan valor principalmente en los eslabones superiores (herramientas y aplicaciones). El desafío para el cierre de la década consiste en migrar desde una posición de "ensamblador físico" e "implementador de APIs externas" hacia un rol de desarrollador de soluciones verticales propietarias. **La verdadera soberanía económica no residirá en competir por la potencia de cómputo, sino en el control de las capas de software y datos que definen la productividad de las industrias clave de la región.**

[342] Modelo LLM en portugués, diseñado para la administración pública, la ciencia y aplicaciones de código abierto.
[343] Primer modelo europeo de IA en castellano y lenguas cooficiales.
[344] Familia de LLMs abiertos para el Sudeste Asiático, basado en modelos fundacionales abiertos Gemma, Llama y Qwen.

V. Elementos de política de desarrollo productivo para la inteligencia artificial en América Latina y el Caribe

El análisis de la estructura y dimensionamiento de la cadena de valor de la IA en América Latina y el Caribe evidencia un posicionamiento y peso económico de la región sesgado hacia los eslabones finales de la cadena de valor (herramientas, integración e implementación), mientras los eslabones intensivos en capital y factores de producción (chips de última generación y cómputo) quedan concentrados en los líderes globales. Al mismo tiempo, el análisis identifica desarrollos emergentes en manufactura de dispositivos de *edge computing* e IoT, ensamblaje de servidores, en algunos países, acompañado de un crecimiento de centros de datos y zonas cloud que provienen, principalmente, de proyectos de inversión de proveedores internacionales. Finalmente, se han identificado cuellos de botella sistémicos en escala y coordinación del ecosistema, que condicionan la capacidad regional de pasar de adopción a oferta exportable.

Este capítulo presenta recomendaciones de política pública para el desarrollo de las industrias asociadas a la IA en América Latina y el Caribe y está estructurado en seis secciones. La sección A establece como punto de partida que el objetivo para la región debe estar encaminado a escoger eslabones viables, coordinar instrumentos y escalar empresas. El posicionamiento estratégico en la cadena de valor implica reconocer que la región solo se encuentra preparada para competir en ciertos eslabones, es decir debe especializarse en aquellos donde ya existen capacidades observables como software, herramientas de integración y desarrollo de aplicaciones orientadas a sectores productivos y a la inclusión social [345].

La especialización en ciertos escalones de la cadena de valor se complementa con la necesidad de **impulsar el escalamiento empresarial y la consolidación productiva**, lo que se presenta en la sección B. Una de las principales restricciones regionales es la alta fragmentación del tejido empresarial tecnológico. La existencia de múltiples firmas pequeñas, con capacidades limitadas para invertir,

[345] En este segmento destacan actores como Globant en la implementación de soluciones de IA en industrias específicas, Stefanini en automatización empresarial y asistentes corporativos, y NotCo en aplicaciones de IA en la industria alimentaria.

innovar y exportar, exige políticas orientadas a la agregación de escala, la formación de clústeres, el fortalecimiento de campeones regionales y la especialización funcional dentro del ecosistema [346].

Estos esfuerzos deben estar acompañados por un eje orientado al **financiamiento y acceso a mercados** tal y como es presentado en la sección C. Esto permitirá superar las barreras existentes en etapas tempranas de emprendimiento, fases de crecimiento y su expansión internacional. La evidencia muestra que la región presenta déficits en capital de escalamiento y en mecanismos de inserción en cadenas globales de valor, lo que limita la capacidad de las empresas para convertir innovaciones en productos exportables.

En particular, el eje de **infraestructura de cómputo** para IA requiere aprovechar la inversión extranjera ya desplegada en centros de datos y servicios cloud para generar encadenamientos productivos locales y reducir las barreras de acceso a capacidades de procesamiento [347]. Esta recomendación es presentada en la sección D.

Finalmente, la promoción de **talento y desarrollo tecnológico aplicado,** presentados en la sección E, resulta transversal a toda la estrategia, ya que el crecimiento de la industria depende de la formación, retención y utilización efectiva de capital humano orientado a la resolución de problemas productivos concretos [348].

En general, estos cinco ejes de política de desarrollo productivo permiten articular una estrategia que podría responder a las limitaciones estructurales identificadas y capitalizar las oportunidades reales de la región para transitar hacia el desarrollo de una oferta competitiva de soluciones de inteligencia artificial.

El diagrama siguiente resume la lógica de intervención diferenciada que orienta las recomendaciones. En los eslabones iniciales, donde la presencia regional es limitada o depende de actores globales, la política debe ser selectiva y enfocarse en nichos viables, derrames, acceso a capacidades críticas y construcción de activos estratégicos. En los eslabones finales, donde la región cuenta con mayor base empresarial —software, herramientas, integración y aplicaciones—, el énfasis debe estar en escalamiento, especialización funcional, clústeres, adopción productiva, validación, acceso a mercados e internacionalización. Adicionalmente, esta priorización debe apoyarse en condiciones institucionales transversales habilitantes: capacidades TOPP [349], gobernanza de datos y regulación habilitante que se desarrollan más adelante en este capítulo.

---

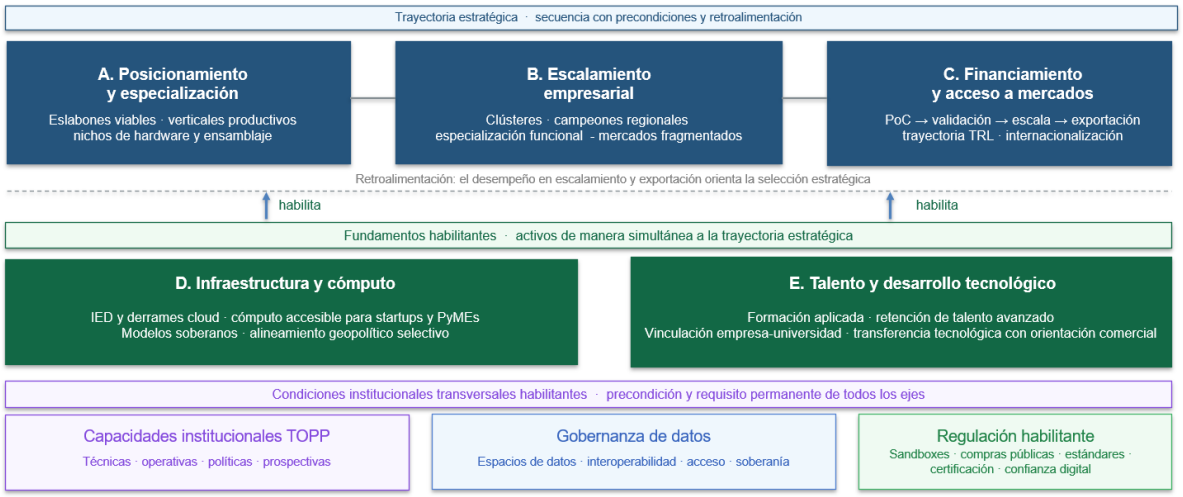
[346] En este contexto, el fortalecimiento de empresas con capacidad de expansión, siguiendo trayectorias similares a las de Accenture en integración de sistemas o replicando modelos regionales como Globant, se vuelve central para construir economías de escala y presencia internacional.

[347] En este ámbito, actores como Amazon Web Services, Microsoft Azure o Scala Data Centers juegan un rol clave como plataformas habilitantes.

[348] Iniciativas como CENIA o Maritaca AI muestran que la combinación entre academia y empresa puede generar capacidades diferenciales en la región.

[349] El concepto TOPP refiere a capacidades institucionales técnicas, operativas, políticas y prospectivas, entendidas por la CEPAL como funciones críticas para gestionar transformaciones complejas: producir evidencia, ejecutar políticas, articular legitimidad y anticipar escenarios. Véase Salazar-Xirinachs, J. M. y A. Boeninger Sempere (2025), *Capacidades institucionales técnicas, operativas, políticas y prospectivas (TOPP) para la gestión de las transformaciones: fundamentos para un nuevo paradigma*, Documentos de Trabajo, LC/TS.2025/56, CEPAL.

Diagrama 22
Cadena de valor de la IA en América Latina y el Caribe: posicionamiento regional y espacios de intervención de política



Fuente: Elaboración propia.

La matriz no debe interpretarse como una asignación rígida de instrumentos por eslabón, sino como una guía para ordenar prioridades de intervención. Los ejes de política —posicionamiento y especialización, escalamiento empresarial, financiamiento y acceso a mercados, infraestructura y cómputo, y talento y desarrollo tecnológico— se aplican con distinta intensidad según la posición de la región y países en cada eslabón. A su vez, las condiciones institucionales transversales —capacidades TOPP, gobernanza de datos y regulación habilitante— permiten que esas intervenciones se articulen de manera sostenida y generen confianza para la adopción y el escalamiento de soluciones de IA.

A. Posicionamiento estratégico y especialización de empresas digitales en la cadena de valor

1. Posicionamiento estratégico en eslabones con empresas regionales

La evidencia sobre la estructura de la cadena de valor de la IA en América Latina y el Caribe presentada en el capítulo anterior, muestra que la región parte desde una posición heterogénea, es decir sesgada hacia los eslabones finales, particularmente, en software, herramientas, desarrollo e implementación de aplicaciones, consultoría tecnológica e integración de sistemas. En estos segmentos existen firmas regionales con capacidades importantes (como Globant, Stefanini, NotCo o Nubank). Al mismo tiempo, se presenta un tejido de integradores y desarrolladores de menor escala que participan en consultoría y desarrollo de herramientas. Sin embargo, la presencia de estos actores no constituye el resultado de políticas productivas orientadas a consolidar sus actividades como un núcleo industrial regional coordinado y especializado. Por ello, esta línea estratégica contempla la promoción de políticas concentradas en aquellos eslabones donde existe base empresarial, aprendizaje acumulado y posibilidad de escalamiento. En un segundo plano, los segmentos de frontera, donde la región carece de masa crítica suficiente, deberían estar dirigidos a productos específicos para apoyar el despliegue y ensamblaje de soluciones de nicho como IoT o servidores *edge*.

En base a ello, se propone la **creación de una estrategia de posicionamiento industrial por eslabones de la cadena de valor** que permita identificar, clasificar y priorizar las actividades donde operan empresas regionales con la finalidad de distinguir desarrolladores de aplicaciones, integradores,

proveedores de software empresarial y empresas con capacidades de especialización sectorial. Esta estrategia podría traducirse en el desarrollo de hojas de ruta por segmento, construidas sobre diagnósticos empresariales y sectoriales, para concentrar instrumentos de financiamiento, innovación, asistencia técnica y promoción exportadora. En esa línea, se señala la necesidad de articular políticas digitales y de desarrollo productivo (PDPs), para superar la fragmentación institucional y la débil coordinación entre transformación digital y desarrollo empresarial [350]. Este direccionamiento coincide con lo articulado de manera general por instituciones multilaterales. Por ejemplo, el Banco Mundial, a través del marco DE4LAC propone organizar la intervención pública alrededor de infraestructura de datos, plataformas, negocios digitales, habilidades y confianza [351]. Asimismo, CAF ha mostrado que las hojas de ruta por cadenas productivas permiten ordenar la selección sectorial, la cuantificación de iniciativas y los métodos de financiamiento, evitando la dispersión de recursos en estrategias demasiado amplias [352].

En esa línea, América Latina y el Caribe requieren entender que el posicionamiento estratégico de eslabones viables para el desarrollo de la IA parte de **construir infraestructura institucional, vinculado a empresas con capacidades de exportación, incubación y desarrollo de producto**. Siguiendo las mejores prácticas en este área, resulta interesante la experiencia internacional de *Software Technology Parks of India* (STPI), organismo dependiente del Ministerio de Electrónica y Tecnologías de la Información, quien desplegó una estrategia focalizada en espacios específicos de la cadena como software, servicios de gestión de infraestructura tecnológica (IT) y habilitados por tecnologías de la información (ITES), así como la incubación, creación de producto e impulso exportador. [353] El esquema STPI fue concebido como un régimen sectorial orientado al desarrollo y exportación de software, con ventanilla única, infraestructura de conectividad, incubación y servicios de apoyo. Posteriormente, esta iniciativa evolucionó hacia una red de centros de emprendimiento y de tecnología aplicada en áreas como IA, IoT, *blockchain*, agritech o medtech [354]. Según STPI, la entidad ya cuenta con decenas de centros en el país, y ha desplegado 24 centros dedicados exclusivamente a la promoción de emprendimientos.

Asimismo, se recomienda complementar el posicionamiento empresarial con mecanismos de **escalamiento e internacionalización para las empresas de IA que ya operan en los eslabones priorizados**. En este objetivo, la experiencia de *Enterprise Singapore* demuestra cómo se articulan instrumentos para crecimiento empresarial y expansión internacional a través de programas como *Scale-Up* [355], orientado a compañías con alto potencial para convertirse en futuros campeones globales, y *Market Readiness Assistance Grant* [356]. Estas iniciativas cofinancian costos de promoción, desarrollo de negocios y establecimiento en nuevos mercados externos. Este tipo de intervención permite que la priorización de eslabones se viabilice a través de una trayectoria concreta de crecimiento empresarial, innovación y acceso a mercados. En consecuencia, una estrategia regional de posicionamiento debería incluir, al menos, tres líneas de acción: i) mapear empresas digitales existentes por eslabón y por nivel de madurez, ii) concentrar instrumentos en los segmentos donde ya existe presencia regional; y, iii) conectar esas empresas con programas de escalamiento e internacionalización.

---

[350] Fuente: CEPAL. Transformación Digital Productiva. Pg. 5 https://repositorio.cepal.org/server/api/core/bitstreams/678e47d8-cc9a-4691-8f38-eca4b74e0bb3/content

[351] Fuente: World Bank. https://www.worldbank.org/en/programs/lac-digital/de4lac

[352] Fuente: CAF. Estrategia para la transformación digital de los sectores productivos en América Latina. https://scioteca.caf.com/handle/123456789/1665

[353] Fuente: STPI. https://www.stpi.in/about-stpi

[354] Fuente: STPI. https://bengaluru.stpi.in/hi/directorate-statutory-services

[355] Fuente: Enterprise Singapore. https://www.enterprisesg.gov.sg/scale-up

[356] Fuente: Enterprise Singapore. https://www.enterprisesg.gov.sg/financial-support/market-readiness-assistance-grant

Este tipo de intervención tiene el beneficio de estar ya vinculada al esfuerzo latinoamericano para el desarrollo de industrias de software como la Ley de Economía del Conocimiento de Argentina [357], la Ley de Informática de Brasil [358] y el Programa SOFTEX [359] que combinan incentivos a la inversión, fortalecimiento de capacidades empresariales y promoción de la competitividad del sector tecnológico.

2. Especialización en verticales productivos

Si bien la evidencia disponible muestra que América Latina y el Caribe presentan limitaciones para disputar su presencia en los eslabones intensivos en capital de la cadena de valor de la IA, la región posee aspectos diferenciales en ciertos sectores productivos donde la tecnología puede ser adaptada para la resolución de problemas concretos de negocio y de coordinación entre eslabones. En particular, los informes de diagnóstico muestran que la especialización sectorial constituye una ruta para el desarrollo de soluciones de IA cuando confluyen cuatro elementos:

i) importancia económica del sector,

ii) disponibilidad de datos y conocimiento operativo,

iii) existencia de cuellos de botella productivos, y

iv) validación temprana en mercados reales.

Los capítulos precedentes de este estudio destacan que, para los países de la región, la trayectoria a considerar consiste en afinar modelos y construir aplicaciones para sectores donde existan datos, conocimiento operativo y ventajas comparativas, como agricultura, biodiversidad, logística, minería o fintech. De esa forma, este objetivo contempla la construcción de la política pública hacia capacidades empresariales en verticales productivos priorizados y que contengan una mayor posibilidad de diferenciación.

En este sentido, el caso agroindustrial argentino resulta ilustrativo con EIWA que representa una aplicación de tecnologías de big data y agricultura de precisión mediante una plataforma orientada a acelerar los procesos de innovación de grandes proveedores de insumos y ofrecer servicios a productores agrícolas [360]. En ese contexto, la digitalización, y especialmente la IA, representan herramientas para resolver problemas transversales de trazabilidad, logística, intercambio de información, abastecimiento y validación de nuevas soluciones.

Asimismo, en servicios financieros, Nubank representa un caso importante de aplicación de IA para la segmentación crediticia a partir del control de bases de datos sobre el comportamiento financiero de clientes latinoamericanos. En la industria alimenticia, NotCo evolucionó desde una empresa de consumo hacia una plataforma tecnológica B2B. En el ámbito corporativo, Globant ha construido unidades de negocio enfocadas en la implementación de IA en verticales específicas como aerolíneas, finanzas y automotriz.

Así, se propone la elaboración de estudios técnicos sectoriales que identifiquen cadenas productivas, casos de uso, demanda empresarial, disponibilidad de datos, barreras regulatorias y necesidades de financiamiento para el desarrollo de soluciones de IA en sectores específicos. En esa línea, el sector agropecuario presenta un alto potencial de desarrollo de aplicaciones de IA, donde las

---

[357] Fuente: Ministerio de Desarrollo Productivo. https://www.argentina.gob.ar/noticias/ley-de-economia-del-conocimiento-se-establecen-las-actividades-promovidas

[358] Fuente: Ministério do Planejamento e Orçamento. https://www.gov.br/planejamento/pt-br/acesso-a-informacao/participacao-social/conselhos-e-orgaos-colegiados/cmap/politicas/area/comunicacao-ciencia-e-tecnologia/subsidios/lei-informatica

[359] Fuente: Ministério da Ciência, Tecnologia e Inovação. https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/lei-de-tics/2022-programa-nacional-de-software-para-exportacao-2013-softex

[360] Fuente: (Katz, R., et al, 2025). EY. Competitividad de la cadena agroindustrial argentina: transformar para crecer. Pg. 39

estrategias regionales pueden articular políticas digitales con desarrollo productivo[361]. A nivel internacional, la Comisión Europea ha generado una política de especialización sectorial a través de la *Apply AI Strategy* que orienta medidas concretas hacia sectores estratégicos como salud y farmacéutica, movilidad, transporte y automoción, robótica, manufactura, ingeniería y construcción, agro-alimentos, energía, comunicaciones electrónicas y sector público[362]. Por ello, la especialización vertical representa una forma de ordenar la política de desarrollo productivo alrededor de mercados con demanda, y capacidades diferenciadas.

De manera similar, la referencia de *Scale AI* en Canadá describe una estrategia que se organiza alrededor de proyectos colaborativos aplicados a cadenas productivas concretas, especialmente, comercio minorista, manufactura, transporte, infraestructura, salud e ICT, con énfasis en cadenas de suministro inteligentes. Este enfoque resulta relevante porque **articula financiamiento, colaboración empresarial, desarrollo de producto y comercialización en torno a problemas sectoriales específicos**. El clúster *Scale AI* reúne sectores diversos para construir cadenas de suministro inteligentes y posicionar al país como exportador de productos y servicios basados en IA.

En conclusión, en el caso latinoamericano y caribeño la especialización en verticales productivos debe traducirse en al menos tres líneas de acción:

i) priorizar sectores donde existan activos diferenciales y demanda empresarial identificable,

ii) **crear mecanismos de validación y co-desarrollo** entre empresas tecnológicas y usuarios productivos; y,

iii) **convertir pilotos sectoriales en una base para la expansión comercial**, la generación de propiedad intelectual y la exportación de soluciones tecnológicas.

3. Consolidación de empresas regionales operando en mercados fragmentados

La evidencia presentada en el capítulo IV sugiere que las principales limitantes para el desarrollo de empresas con foco en IA en América Latina y el Caribe radica en la ausencia de capacidades consolidadas y en la forma dispersa en que se encuentran organizadas. En el eslabón de desarrollo e implementación de aplicaciones, la región incluye un tejido amplio de pequeñas firmas de consultoría, software e integración, pero presenta una reducida capacidad de atender proyectos complejos, invertir en desarrollo de producto o expandirse regionalmente. El caso uruguayo resulta ilustrativo porque si bien se identifican 243 empresas en la cadena de IA, el 55% tiene diez o menos empleados y el 39% se ubica en consultoría y desarrollo de herramientas[363]. De manera similar, la fragmentación se repite en muchos países de la región, aunque no dispone de estadísticas de dimensionamiento por empresa (ver cuadro 31).

Cuadro 31
América Latina: número de empresas en el ecosistema de IA por país, 2025

País	Número de empresas
Brasil	2 641
Chile	664
Colombia	584
México	742

Fuente: RankMyAI.

[361] Fuente: CEPAL. Transformación Digital Productiva. Pg. 43. https://repositorio.cepal.org/server/api/core/bitstreams/678e47d8-cc9a-4691-8f38-eca4b74e0bb3/content

[362] Fuente: Comisión Europea. Estrategia de uso de la inteligencia artificial. https://eur-lex.europa.eu/resource.html?uri=cellar:194ae542-a421-11f0-97c8-01aa75ed71a1.0012.02/DOC_1&format=PDF

[363] Weltevreden, J. (2025). Uruguay AI Ecosystem mapping 2025 – Methodology and key findings.

Este aspecto refleja una estructura empresarial fragmentada y de baja escala, con poca capacidad de financiamiento.

Por otro lado, en el mercado global de integradores de sistemas para IA, el jugador más importante (*Accenture*) concentra apenas 7.2% de participación, lo que confirma que se trata de un segmento naturalmente disperso, pero con ventaja para los actores de mayor tamaño que acceden a contratos corporativos, reputación y asociaciones tecnológicas. En conclusión, la fragmentación limita la posibilidad de crecimiento de firmas regionales y dificulta la formación de capacidades colectivas y economías de especialización. En esa línea, esta estrategia busca generar mecanismos de consolidación empresarial que permitan articular un conjunto disperso de pequeñas empresas con capacidad de competir por proyectos de mayor escala e insertarse en cadenas regionales.

Desde esta perspectiva, se propone **impulsar la formación de clústeres, consorcios empresariales, plataformas de innovación abierta y redes de colaboración** entre startups, integradores, firmas tecnológicas medianas y empresas ancla. Es decir, propender a la creación de estructuras de colaboración para compartir activos, capacidades comerciales, certificaciones, acceso a clientes, validación y puesta en marcha de soluciones, así como el desarrollo conjunto de proyectos. En esa línea, una experiencia internacional relevante es *BIND SME* del País Vasco, la cual corresponde a una iniciativa público-privada que conecta startups con pymes industriales para implementar soluciones innovadoras en entornos reales mediante proyectos financiados y con acceso directo al mercado. De ese modo, las pequeñas empresas tecnológicas pueden crecer a partir de contratos concretos, casos de éxito comprobados y experiencia verificable[364]. La plataforma también destaca su modelo de innovación abierta a la colaboración entre pymes y startups que impulsa nuevas oportunidades de cooperación en sectores industriales específicos[365].

Asimismo, la consolidación de empresas regionales a partir de mercados fragmentados requiere instrumentos para **promover el uso compartido de infraestructura, equipamiento, testeo (*test bed*) y validación tecnológica**. En este punto, otra referencia relevante es el *Basque Digital Innovation Hub* (BDIH) que es una red de activos y servicios que permite a las PyMEs acceder a recursos tecnológicos para formación, investigación, testeo y validación, bajo un esquema de trabajo colaborativo entre centros de I+D, formación profesional, universidades e instituciones públicas regionales. Esta experiencia resulta relevante porque un ecosistema fragmentado puede ganar escala cuando comparte recursos que, estarían acotados para empresa por si solas, infraestructura, asesoramiento y acceso a experimentación aplicada. De manera complementaria, la *Digital Hub Initiative* de Alemania ha sido concebida precisamente para fortalecer la cooperación entre startups y empresas establecidas, construyendo puentes sectoriales y territoriales para que la innovación no permanezca dispersa en nodos aislados.

En consecuencia, para América Latina y el Caribe esta estrategia debe traducirse en al menos tres líneas de acción:

i) impulsar clústeres y consorcios que permitan a firmas pequeñas atender conjuntamente proyectos de IA complejos y acceder a infraestructura como centros de datos equipados para apoyar el entrenamiento e inferencia de modelos de IA;

ii) crear programas de IA basados en innovación abierta y contratación con empresas tractoras que funcionen como primeros mercados; y,

[364] Fuente: BIND. https://bind.spri.eus/es/startup-sme/
[365] Fuente: Grupo SPRI. https://www.spri.eus/en/support/bind-sme

iii) habilitar redes de activos compartidos para testeo, validación y escalamiento, de forma tal que la fragmentación empresarial no derive en aislamiento productivo, sino en especialización cooperativa.

4. Madurez de mercados específicos para ensamblaje de hardware

Si bien América Latina y el Caribe no presentan condiciones para competir en la producción de semiconductores de frontera, el capítulo IV, en su primera sección, indica que existen en la región espacios viables en segmentos intermedios del eslabón de equipamiento, particularmente, en ensamblaje de servidores, placas, racks, dispositivos de edge AI e IoT industrial. En este punto, la evidencia regional sugiere que la oportunidad reside en una estrategia de consolidación de mercados específicos donde confluyen proximidad logística, capacidades industriales preexistentes y demanda internacional o regional. Así, por ejemplo, México ya concentra actividades de Foxconn, Inventec y Quanta en ensamblaje de GPUs, placas y racks para servidores de IA. Igualmente, Costa Rica mantiene capacidades de ingeniería avanzada vinculadas a Intel; y, Brasil presenta nichos especializados en soluciones System-in-Package para Edge AI, módulos de conectividad e IoT industrial, así como dispositivos ligados a la gestión inteligente de energía. De esta forma, el objetivo de política se orienta a fortalecer posiciones selectivas en mercados donde la región ya ha mostrado una inserción inicial.

En este terreno, la experiencia de Malasia resulta importante porque muestra que una estrategia de ensamblaje puede evolucionar hacia actividades de mayor valor agregado dentro de nichos seleccionados. La *National Semiconductor Strategy* reconoce como punto de partida la fortaleza previa del país en Ensamblaje y Prueba de Semiconductores Subcontratados (*Outsourced Semiconductor Assembly and Test* – OSAT) y plantea avanzar, de manera gradual, hacia empaquetado avanzado, manufactura de mayor complejidad y equipamiento especializado [366].

En ese marco, la lección de esta experiencia para América Latina y el Caribe debe enfocarse en **consolidar operaciones precisas** (por ejemplo, racks para servidores, packaging, pruebas, edge devices o módulos industriales) **y construir una trayectoria de sofisticación progresiva y maduración productiva a partir del ensamblaje de componentes** que permita atraer inversión, construir volumen, desarrollar proveedores, aprender estándares de calidad y generar capacidades de ingeniería aplicada. Una de las ventajas comparativas de la región sería la proximidad logística al mercado de Estados Unidos; y, en el caso de México, el capítulo 10 señala que la presencia de Foxconn, Inventec y Quanta responde a la reducción de riesgos políticos asociados a Asia y que la localización mexicana reduce costos de transporte hacia los centros de datos desplegados en Estados Unidos. A ello se suma que México estructuró una estrategia pública de semiconductores a través del proyecto *Kutsari* (integrado en Plan México), con un enfoque inicial en diseño y con la intención de consolidar capacidades de fabricación a escala industrial [367]. En consecuencia, el caso mexicano sugiere que la madurez de mercados de ensamblaje puede servir no solo para capturar inversión de relocalización, sino también para sentar las bases de un escalamiento gradual hacia funciones de mayor valor agregado [368].

[366] Fuente: MIDA. https://www.mida.gov.my/mida-news/national-semiconductor-strategy-to-guide-industry-up-value-chain

[367] Fuente: Presidencia de la República de México. https://www.gob.mx/presidencia/prensa/presidenta-claudia-sheinbaum-anuncia-creacion-del-centro-nacional-de-diseno-de-semiconductores-kutsari

[368] Fuente: OCDE. https://www.oecd.org/en/publications/promoting-the-development-of-the-semiconductor-ecosystem-in-mexico_02c81 dec-en/full-report/understanding-the-policy-and-regulatory-landscape_01546ba3.html#chapter-d1e6565-e89f91328c

5. Articulación regional para la negociación estratégica de materiales críticos

Como punto de partida, es necesario reconocer que América Latina y el Caribe no presentan condiciones para disputar el liderazgo en la fabricación o desarrollo de semiconductores de frontera, aunque dispone de una dotación relevante de elementos críticos y materias primas [369] requeridas para la fabricación de circuitos, baterías, componentes electrónicos y equipamiento asociado con la infraestructura que sostiene la IA. En esa línea, el problema estructural radica en la relación primario-exportadora y extractiva que mantienen muchos de los países de la región, con baja captura de valor y una posición dependiente en el comercio internacional. Por otra parte, el proceso de extracción ha aumentado el deterioro de los términos de intercambio de los productos primarios, dado que estos mantienen un menor valor relativo frente a los productos manufacturados, lo cual obliga a exportar un mayor volumen para compensar menores precios. En esa misma lógica, el análisis sobre materiales críticos [370] indica que América Latina y el Caribe posee una participación significativa en reservas mundiales de cobre, litio, grafito y tierras raras, pero continúa concentrada en etapas iniciales de la cadena, con una inserción débil en segmentos de mayor complejidad tecnológica. En este contexto, el objetivo de la política pública no debería limitarse a atraer inversión minera o ampliar exportaciones, sino **avanzar hacia una articulación regional que permita negociar mejores condiciones de intercambio, acceso a tecnología, refinación, procesamiento y compromisos de industrialización asociados a los recursos estratégicos**. Para la región, donde la fragmentación regional es estructural y los intereses mineros nacionales son divergentes, sería conveniente diferenciar entre una *diplomacia bilateral de materiales* (que es factible a corto plazo, y basada en la coordinación entre países) y una *articulación regional de negociación conjunta* (que requiere más tiempo e incluso una arquitectura institucional). Por ejemplo, la iniciativa conjunta entre Zambia y la República Democrática del Congo demuestra que la estructuración de una zona económica especial transfronteriza orientada a producir precursores y baterías, incorporando estrategias de contenido local y transferencia tecnológica. Esta experiencia es relevante porque indica que la ventaja no surge únicamente de poseer el recurso, sino de negociar en bloque el paso desde la extracción hacia actividades intermedias y manufactureras [371].

Por otro lado, desde la demanda, Japón ha institucionalizado una estrategia de diplomacia de recursos a través del JOGMEC [372] (*Japan Security Energy and Metals for Organization*) que combina cooperación con países ricos en recursos, apoyo financiero, asistencia técnica, formación y almacenamiento para asegurar suministros críticos [373]. En esa línea, la lección para América Latina y el Caribe corresponde construir una estrategia equivalente desde la posición de la oferta. Esto consiste en la **implementación de una diplomacia regional de materiales críticos que use la dotación de recursos como mecanismo para negociar** refinanciación, procesamiento, transferencia tecnológica, compras de largo plazo, estándares, formación y acceso preferente a segmentos más avanzados de la cadena. Este enfoque plantea la necesidad de utilizar el alineamiento geopolítico como instrumento para mejorar su posición en la cadena de valor y evitar que la región permanezca limitada a funciones de baja captura de valor.

---

[369] Según datos de la U*.S. Geological Survey Mineral Commodity Summaries 2026*, por ejemplo, Brasil representa un 24.7% de las reservas mundiales de tierras raras (https://pubs.usgs.gov/periodicals/mcs2026/mcs2026-rare-earths.pdf), en níquel un 11.43% de las reservas (https://pubs.usgs.gov/periodicals/mcs2026/mcs2026-nickel.pdf), mientras que Chile posee un 24.86% y Argentina un 11.89% de las reservas mundiales de Litio (https://pubs.usgs.gov/periodicals/mcs2026/mcs2026-lithium.pdf).

[370] En resumen, entre cobalto, cobre, níquel, litio, grafito y tierras raras, América Latina posee, en promedio, cerca del 22.7% de las reservas de materiales críticos, mientras que China un 17.3% y Estados Unidos un 3.3%. Esta interpretación coincide con la literatura sobre intercambio ecológicamente desigual que se sostiene que los productos primarios tienen menor valor relativo frente a los manufacturados (Falconí, F., et al, 2026) *Extraction and dematerialization in Latin America*, 1970–2023. https://www.sciencedirect.com/science/article/abs/pii/S0921800926000327

[371] Fuente: Comisión Económica para África (CEPA). https://www.uneca.org/stories/zambia-and-drc-to-implement-an-innovative-transboundary-battery-and-electric-vehicle-special

[372] Fuente: JOMEC. https://www.jogmec.go.jp/content/300382557.pdf

[373] Fuente: Ministry of Economy Trade and Industry of Japan. https://www.jogmec.go.jp/content/300382557.pdf

B. Escalamiento empresarial y consolidación productiva

1. Promoción de la inversión, innovación y exportación

Si bien América Latina y el Caribe muestra avances en los eslabones de software, herramientas e implementación (con presencia de firmas regionales) y una base empresarial en consultoría e integración, ese posicionamiento continúa siendo limitado por menores niveles de inversión, reducida capacidad de comercialización y dificultad de transformar innovaciones en soluciones exportables. Así, el capítulo IV en su sección 2 señala que la región enfrenta limitaciones estructurales en inversión. Si bien el proyecto de LatamGPT no tiene objetivos comerciales, este demuestra que, aun cuando existan capacidades y articulación institucional, persisten restricciones en acceso a financiamiento. De esa forma, este objetivo contempla que la promoción empresarial articule inversión, innovación y exportación como parte de una misma trayectoria de escalamiento.

En el contexto de la restricción estructural del ecosistema regional que corresponde a la debilidad de la inversión en investigación, desarrollo e innovación, particularmente, del lado privado, Katz et al (2023)[374] señalan que la región invierte 0.63 % del PIB en I+D, frente a 2.2 % en la OCDE. De manera complementaria, el mismo estudio advierte que la banca comercial y los fondos privados se mantienen al margen de los procesos de inversión debido al perfil de riesgo de muchos emprendimientos innovadores. En ese sentido, **el Estado y la banca multilateral deberían asumir un papel más activo en el financiamiento de iniciativas tecnológicas, startups e infraestructura relacionada con IA.** La promoción de la inversión, la innovación y la exportación deben entenderse como una política de acompañamiento empresarial ante la insuficiente profundidad del ecosistema regional para su financiamiento.

Así, se propone **estructurar instrumentos de promoción que permitan movilizar inversión hacia empresas tecnológicas** con capacidad de generar productos y servicios comercializables en IA. Al mismo tiempo, se debe fortalecer las capacidades de innovación y acceso a mercados internacionales. Esto implica orientar la política pública hacia esquemas de apoyo a las empresas, desde el desarrollo de una solución tecnológica hasta su validación comercial e inserción externa[375]. En ese sentido, una buena práctica internacional a replicar es la de *Enterprise Ireland*,[376] agencia pública que integra de manera explícita las funciones de innovación, inversión y comercio exterior dentro de una sola arquitectura institucional. La agencia invierte y apoya a empresas irlandesas para que ganen escala y se conviertan en líderes globales, operando a través de una red internacional de 40 oficinas, y combina instrumentos de financiamiento para I+D, así como el desarrollo de producto con programas específicos de acceso a mercados, investigación comercial y asesoría exportadora[377]. La agencia reporta que las empresas que acceden a sus apoyos de investigación e innovación registran, en promedio, cuatro veces más ventas de exportación Esta experiencia resulta útil, ya que muestra que la promoción empresarial debe combinar las políticas de innovación y comercial en una secuencia que acompañe a las firmas desde el desarrollo tecnológico hasta su internacionalización.

Por otra parte, se propone elevar la inversión privada en innovación empresarial mediante incentivos fiscales y financieros orientados a actividades de investigación, desarrollo y mejora tecnológica en el terreno de IA. Esta medida responde a la necesidad de crear condiciones para que el esfuerzo privado aumente a través de mecanismos tributarios a la inversión en I+D y el acceso al crédito.

---

374 Fuente: Katz, R. et al (2023). *Estrategia y recomendaciones de política pública para el fortalecimiento del ecosistema de ciencia, tecnología e innovación en América Latina y el Caribe: Lecciones aprendidas desde el caso argentino*. CAF.

375 Esta línea de acción es consistente con lo que organismos como el BID y el Banco Mundial han subrayado en sus agendas de desarrollo productivo digital, dado que la transformación tecnológica empresarial requiere combinar innovación, financiamiento, articulación con mercados y vínculos internacionales con la finalidad de evitar que las firmas queden atrapadas en mercados locales con bajo crecimiento.

376 Fuente: Enterprise Ireland. https://www.enterprise-ireland.com/en/global/about-enterprise-ireland

377 Fuente: Enterprise Ireland. https://www.enterprise-ireland.com/en/supports/become-more-innovative

En este punto, una referencia internacional importante corresponde al sistema británico de *R&D tax relief*, administrado por la autoridad tributaria, de pagos y aranceles aduaneros del Reino Unido (HMRC) [378], que fue diseñado para apoyar a empresas que desarrollan proyectos en ciencia y tecnología. Esta iniciativa permite reducir la carga tributaria o acceder a un pago en efectivo vinculado al gasto en I+D. [379] La utilidad de esta experiencia para la región representa un estímulo para asumir el costo del riesgo en inversiones tecnológicas.

Asimismo, se recomienda desarrollar programas de apoyo a la innovación empresarial, que no se limiten a la entrega de recursos, sino que combinen financiamiento, asesoría técnica, validación y acompañamiento. Esta medida se vincula con el diagnóstico que advierte que la banca comercial y fondos privados tienden a mantenerse al margen de la inversión proactiva por el perfil de riesgo de las empresas tecnológicas. Este aspecto exige una participación del Estado y la banca multilateral a través del apoyo a iniciativas intensivas en tecnología. Una experiencia relevante en este sentido es el *NRC IRAP* de Canadá, que se presenta como el principal programa de apoyo a la innovación para PyMEs y combina asistencia financiera, servicios de consultoría y conexiones con redes técnicas y de negocio para ayudar a las empresas a convertir sus innovaciones en productos y servicios [380].

2. Consolidación de clústeres y alianzas para potenciar el tejido empresarial

Como fue mencionado arriba, una parte importante del tejido empresarial de empresas de IA en América Latina y el Caribe está compuesta por pequeñas firmas con limitada capacidad para desarrollar producto, atender proyectos complejos, financiar procesos de certificación o acceder a mercados internacionales. Esta estructura empresarial implica una dificultad para coordinar activos, conocimiento y demanda en escalas suficientes para competir con integradores globales en el mercado de IA. De esa manera, este objetivo persigue promover la articulación entre pequeñas firmas tecnológicas, empresas ancla, centros tecnológicos, universidades y asociaciones empresariales.

En esa línea, se propone la **creación y fortalecimiento de clústeres tecnológicos, consorcios sectoriales y plataformas de innovación abierta** que permitan la compartición de infraestructura, acceso a servicios especializados, desarrollo de proyectos colaborativos y agrupación estratégica para captar clientes o mercados externos. Esta estrategia está orientada a reducir costos de coordinación, facilitar la especialización y mejorar la capacidad de respuesta ante proyectos empresariales más sofisticados. Una referencia internacional en este terreno es la *European Digital Innovation Hubs* (EDIH) [381], concebida por la Comisión Europea como una red que apoya a empresas y organizaciones públicas en sus desafíos digitales mediante el acceso a conocimiento técnico, pruebas antes de invertir (*test before invest*), asesoría para financiamiento y formación. La Comisión también destaca que la red combina proximidad regional, lo que permite a empresas locales acceder a servicios especializados que no siempre están disponibles en su propio entorno. [382] Otra experiencia relevante es *Manufacturing USA* que corresponde a una red de institutos de manufactura avanzada que articula alianzas público-privadas entre empresas, academia y gobierno para desarrollar investigación y desarrollo aplicado (I+D), reducir costos y riesgos de adopción tecnológica, compartir equipamiento y fortalecer capacidades laborales [383]. A través de la aplicación de su modelo colaborativo se han podido resolver ciertos problemas de innovación y escalamiento, mediante la agrupación y consolidación de startups, pequeñas y medianas firmas manufactureras o tecnológicas.

---

[378] Fuente: HMRC. https://www.gov.uk/guidance/corporation-tax-research-and-development-rd-relief

[379] Fuente: HMRC. https://www.gov.uk/guidance/tell-hmrc-that-youre-planning-to-claim-research-and-development-rd-tax-relief

[380] Fuente: National Research Council Canada. https://nrc.canada.ca/en/support-technology-innovation/financial-support-technology-innovation y https://nrc.canada.ca/en/support-technology-innovation/nrc-irap-advisory-services

[381] Fuente: Comisión Europea. *European Digital Innovation Hubs*. https://digital-strategy.ec.europa.eu/en/policies/edihs

[382] Fuente: Comisión Europea. https://digital-strategy.ec.europa.eu/en/library/digital-innovation-hubs-helping-companies-across-economy-make-most-digital-opportunities-brochure

[383] Fuente: Manufacturing USA. https://www.manufacturingusa.com/about-us

En ese sentido, la recomendación para América Latina podría seguir tres líneas de acción:

i) establecer plataformas de articulación territorial o sectorial entre pequeñas empresas tecnológicas y productivas con foco en IA;

ii) habilitar servicios compartidos (test, validación, certificación, formación y financiamiento); y,

iii) promover proyectos colaborativos con empresas ancla que permitan a las pequeñas empresas acceder a ciclos de aprendizaje industrial y oportunidades de escalamiento en el ecosistema de IA.

Bajo este enfoque se busca crear un entorno cooperativo que propicie la evolución y participación de las PyMEs en la cadena de valor de la IA.

3. Desarrollo de campeones regionales en la. cadena de valor de la IA

El capítulo IV establece que en América Latina y el Caribe ya se han desarrollado algunas empresas tecnológicas que han logrado superar ciertas restricciones de posicionamiento en determinados segmentos de la cadena de valor. Casos como Globant en integración e implementación de soluciones de IA, Stefanini en asistentes corporativos, NotCo en aplicaciones de IA para la industria alimentaria o Nubank en servicios financieros demuestran que, si bien la región posee capacidades empresariales avanzadas, no presenta una masa crítica de firmas que logren escalar de manera sostenida para convertirse en líderes regionales. El mismo diagnóstico muestra que la cadena de valor presenta una predominancia de empresas pequeñas con dificultades para acceder a capital y baja capacidad para escalar sus procesos de innovación. De esa manera, este objetivo contempla que la política pública pueda identificar y acelerar el crecimiento de firmas con potencial para convertirse en campeones regionales. Es decir, empresas capaces de convocar una demanda importante, atraer talento, juntar proveedores, ganar reputación internacional y generar aprendizaje tecnológico para el resto del ecosistema.

Para ello, se propone la creación de mecanismos de apoyo a empresas tecnológicas con alto potencial de escalamiento, innovación e internacionalización en IA. Esto incluye diferenciar los instrumentos generales de apoyo (a startups o PyMEs) para abarcar mercados más extensos. En la práctica, esta línea de acción requiere la aplicación de criterios objetivos de selección relacionados con incremento de ingresos, sofisticación tecnológica, capacidad exportadora, innovación de producto o servicios, liderazgo en nichos específicos y potencial para generar encadenamientos con otras firmas locales y regionales. En ese sentido, la política debería traducir este enfoque por medio de la articulación de instrumentos de financiamiento para el crecimiento, acompañamiento estratégico, fortalecimiento de liderazgo empresarial, expansión internacional y desarrollo de capacidades de I+D aplicada. Es por ello que esta fase prioriza la transición de empresas con condiciones para ser compañías exitosas hacia plataformas empresariales con presencia regional o global.

En esa línea, una referencia internacional importante es la estrategia de *Enterprise Singapore* a través del programa *Scale-Up* y de la *Singapore Global Enterprises Initiative*. Estos programas fueron diseñados para promover el crecimiento de empresas locales de alto potencial y convertirlas en futuros campeones globales. Según *Enterprise Singapore*, las 80 empresas participantes de las primeras siete cohortes generaron en conjunto $ 2.5 mil millones de dólares singapurenses en ingresos adicionales en los tres años posteriores a su ingreso al programa, mientras que sus ingresos provenientes del exterior crecieron 53%[384].

[384] Fuente: Enterprise Singapore. https://www.enterprisesg.gov.sg/-/media/esg/files/media-centre/media-releases/2025/september/mr03525_scale-up-participants-increased-revenue-by-s$25-billion-in-three-years-after-joining-the-scale-up-programme.pdf

De manera complementaria, la *Singapore Global Enterprises Initiative*[385] busca desarrollar una nueva generación de empresas globales, líderes en innovación y con capacidad para articular actividades estratégicas por medio del financiamiento, apoyo a talento, promoción de la expansión internacional e innovación interna. Esta experiencia resulta relevante para la región dado que muestra que el desarrollo de campeones empresariales surge a través de políticas que **identifican a empresas con potencial, las acompañan con instrumentos diferenciados, y las conectan con metas explícitas de internacionalización e innovación**.

En este sentido, se recomienda que esta estrategia atienda al menos tres líneas de acción:

i) establecer programas de selección y acompañamiento para firmas tecnológicas con mayor potencial de escalamiento de IA, con la finalidad de acceder a instrumentos especializados de financiamiento;

ii) **vincular a estas empresas con programas de expansión internacional** (desarrollo de liderazgo, atracción de talento y fortalecimiento de capacidades de innovación), con la finalidad de articular varias etapas que conforman la trayectoria comercial; y

iii) **utilizar a estas firmas como nodos del ecosistema** con el objetivo de promover su relación con startups, universidades, proveedores tecnológicos y cadenas regionales de valor.

Con este enfoque, se busca crear un conjunto de firmas pioneras que amplifique el aprendizaje, eleve la sofisticación productiva y contribuya a posicionar a la región como oferente de soluciones digitales para su propio mercado o mercados más amplios.

4. Especialización funcional dentro de los eslabones priorizados

El análisis regional muestra que, dentro de los eslabones donde la región presenta una importante participación empresarial (software, herramientas, integración e implementación de aplicaciones), persiste todavía la tendencia a la operación tradicional de los negocios. La propia evolución de la cadena de valor de la IA muestra que los últimos eslabones de la cadena presentan funciones empresariales distintas. Por un lado, desarrolladores de aplicaciones orientados a sectores o casos de uso específicos; y, por otro, integradores de sistemas encargados de adaptar plataformas de IA a las necesidades del negocio de sus clientes. Al mismo tiempo, el capítulo I señala que el modelo tradicional de software está siendo desafiado por soluciones especializadas; así como, por el avance de agentes y herramientas de IA que está redefiniendo la propuesta de valor del software. De esa forma, este objetivo busca favorecer la especialización funcional de las empresas tecnológicas dentro de eslabones priorizados con la finalidad de construir ventajas competitivas.

Por ello, se recomienda **orientar la diferenciación funcional de las empresas tecnológicas** entre desarrolladores de aplicaciones verticales de IA, integradores de sistemas, proveedores de software empresarial, automatización de procesos, herramientas de evaluación; así como, soluciones basadas en agentes. Esta especialización enfrenta barreras, clientes, ciclos comerciales y necesidades de financiamiento distintas. En la práctica, esto implica que los instrumentos públicos deben tratar con estrategias específicas a cada empresa que desarrolla un producto o servicio. Además, esta especialización puede contribuir a resolver uno de los principales cuellos de botella señalados en los capítulos de diagnóstico y que corresponde a la integración organizacional[386]. En esa línea, una buena práctica internacional es el programa *IMDA Accreditation* de Singapur, que acredita a empresas de producto tecnológico con alto crecimiento para posicionarlas como oferentes calificados ante compradores públicos y grandes empresas[387]. Este aspecto busca establecer un licenciamiento para

[385] Fuente: Enterprise Singapore. https://www.enterprisesg.gov.sg/grow-your-business/boost-capabilities/growth-and-transformation/singapore-global-enterprises-initiative

[386] Según OMDIA el 44% de empresas reporta que las barreras principales para incorporar IA en sus procesos residen precisamente en la complejidad y dificultad de integración. OMDIA (2022). AI maturity study.

[387] Fuente: IMDA. https://www.imda.gov.sg/how-we-can-help/imda-accreditation

compañías innovadoras de producto y proveer confianza a los compradores sobre la funcionalidad y capacidad de entrega de esas empresas. Así ayuda a las empresas a escalar en mercados locales e internacionales. Además, el *Tech Acceleration Lab* complementa ese esquema por medio del acceso a pruebas y despliegues en entornos de desarrollo y producción, reduciendo la fricción entre validación tecnológica y contratación.

Asimismo, se propone que alrededor de esta estrategia se **establezcan tipologías empresariales** dentro de los eslabones priorizados con el objetivo de **identificar el tipo de función que cumple cada empresa en la cadena de valor de la IA**. En esta línea, la experiencia de *Je choisis la French Tech* muestra que, una vez identificadas y movilizadas empresas tecnológicas con perfiles diferenciados, la política puede intervenir sobre la cadena de compra mediante formación, encuentros comerciales y compromisos de adquisición por parte de compradores públicos y privados, con el objetivo de ampliar ingresos y mercado para startups y scale-ups tecnológicas [388].

De la misma forma, se requieren **diseñar instrumentos diferenciados según función empresarial** (SaaS, vertical AI, certificación y acceso a pruebas para proveedores de soluciones) **y programas de escalamiento comercial** para integradores y empresas de implementación. En la misma línea, se podría **vincular la diferenciación empresarial con el acceso a procesos de compras públicas**, mecanismos de acreditación, validación tecnológica y acceso a grandes clientes, de forma que se convierta en una vía concreta para aumentar sofisticación, reputación y capacidad exportadora. Este enfoque permite que la especialización funcional dentro de los eslabones priorizados mejore la organización y formalice el ecosistema, lo que permite pasar de una oferta dispersa a una estructura empresarial complementaria y competitiva. En esa línea, la guía australiana de "*Broader economic benefits in ICT procurement"* constituye una experiencia importante, ya que señala que, en adquisiciones públicas TIC, se pueden dividir proyectos grandes en secciones de trabajo más pequeñas con la finalidad de mejorar el acceso de las pymes a la contratación pública. Con ello, se permite que empresas con alta capacidad técnica pero menor escala participe en componentes especializados del proyecto [389].

C. Financiamiento, modelo de negocios y acceso a mercados

1. Brecha de financiamiento en desarrollo de producto

Una de las principales restricciones para el desarrollo de empresas de IA en América Latina y el Caribe radica en la disponibilidad de financiamiento para transformar capacidades técnicas en productos comercializables. Si bien la región ya muestra avances en sectores específicos de los últimos eslabones de la cadena (software, herramientas e implementación de sistemas), todavía persisten dificultades para sostener procesos de experimentación, prototipado, validación técnica, preparación comercial y propiedad intelectual. Así, este objetivo busca la creación de instrumentos que permitan cerrar la brecha entre idea, prototipo y producto, con la finalidad de evitar que las firmas regionales queden atrapadas por falta de recursos en etapas tempranas de experimentación o baja sofisticación.

Una primera línea de acción consiste en **crear fondos para pruebas de concepto y prototipos para emprendimientos tecnológicos** y startups que todavía no cuentan con validación suficiente para atraer capital privado. Esta fase permite cubrir el tramo más riesgoso del proceso innovador, donde la empresa necesita demostrar viabilidad técnica y factibilidad comercial antes de transitar hacia el escalamiento. Una referencia internacional apropiada es el *ERC Proof of Concept Grant*. Este instrumento de la Unión Europea está diseñado para cerrar la brecha entre resultados de investigación

[388] Fuente: La French Tech. https://lafrenchtech.gouv.fr/en/programme/je-choisis-la-french-tech

[389] Fuente: Department of Industry, Science and Resources. https://www.industry.gov.au/publications/broader-economic-benefits-ict-sector-procurement

y sus primeras fases de comercialización, así como para financiar actividades como test, experimentación, demostración, validación y estrategia [390]. En 2026, el esquema mantiene una subvención de €150,000 por 18 meses para convertir una idea o resultado técnico en una prueba de concepto con orientación al mercado [391].

Asimismo, otra referencia importante para la fase de validación pre-comercial y maduración tecnológica se encuentra en el *EIC Transition* de la Unión Europea [392]. Este programa financia la validación y demostración de tecnologías en entornos de aplicación con el objetivo de llevarlas aproximadamente desde niveles bajos de madurez tecnológica (TRL 3/4) hacia estadios diferenciadores (TRL 5/6) [393]. Asimismo, el programa ofrece subvenciones de hasta €2.5 millones y servicios de aceleración empresarial. Esta referencia resulta útil debido a que se concentra en el tramo donde el producto todavía no está listo para escalar, pero requiere pruebas, demostración y preparación previa a su introducción al mercado. Por ello, se propone **financiar la validación tecnológica en entornos de aplicación y la preparación para mercado**, con lo cual se apoya a las empresas que superaron la fase de concepto, pero requieren demostrar el uso de su tecnología en condiciones casi reales; así como, ajustar su modelo de negocio y construir una propuesta comercial robusta.

Otra línea de acción consiste en **estructurar el financiamiento para el desarrollo de producto a través de procesos secuenciales.** En ese sentido, las etapas de financiamiento deberán transitar desde factibilidad hasta prototipo; así como, comercialización parcial mediante fases sucesivas de evaluación. En este caso, una experiencia internacional es el modelo *Small Business Innovation Research - SBIR* de Estados Unidos, administrado a través de la *Small Business Administration* - SBA y agencias federales, que organiza el apoyo financiero en tres etapas diferenciadas: i) demostración de mérito técnico, factibilidad y potencial comercial, ii) continuación del esfuerzo de I+D; y, iii) comercialización con recursos no SBIR [394]. El sistema federal señala que fase 1 se orienta a pruebas de concepto y que la fase 2 profundiza el desarrollo sobre la base de resultados previos, mientras que además aplica métricas de transición y comercialización para asegurar que las firmas no queden atrapadas indefinidamente en subsidios tempranos [395]. Para la región, esta experiencia sugiere que los instrumentos de financiamiento deben diseñarse como una ruta progresiva e hitos verificables para transitar por las etapas de desarrollo.

2. Financiamiento para escalamiento

El análisis regional evidencia que, una vez superada la fase inicial de desarrollo de producto, muchas empresas tecnológicas enfrentan barreras críticas relacionadas con la escasez de capital para escalar operaciones, fortalecer equipos comerciales, ampliar capacidades de implementación, entrar a nuevos mercados o expandirse. En la práctica, este aspecto implica que empresas con soluciones validadas tienen dificultad para convertir su ventaja tecnológica en crecimiento sostenido debido a la ausencia de instrumentos de capital para crecimiento o coinversión. En ese sentido, este objetivo apunta al diseño de una política para financiar la etapa de escalamiento de productos de IA dirigida hacia empresas que han demostrado tracción comercial y potencial de expansión.

En una primera línea de acción se propone **crear mecanismos de coinversión público-privada para rondas de crecimiento**, dirigidos a empresas en la cadena de valor de IA intensivas en I+D que ya han alcanzado una validación suficiente y requieren montos mayores para expandirse. En esa línea, una

---

390 Fuente: European Research Council. https://erc.europa.eu/apply-grant/proof-concept

391 Fuente: European Research Council. https://erc.europa.eu/news-events/news/war-damage-heart-repair-erc-grantees-supported-turn-their-science-practice

392 Fuente: European Innovation Council. https://eic.ec.europa.eu/eic-funding-opportunities/eic-transition_en

393 Fuente: Ministerio de Industria de España. *Niveles de madurez de la tecnología*. https://www.mintur.gob.es/publicaciones/publicacionesperiodicas/economiaindustrial/revistaeconomiaindustrial/393/notas.pdf

394 Fuente: Small Business Innovation Research. https://www.sbir.gov/about

395 Fuente: Small Business Innovation Research. https://www.sbir.gov/performance-benchmarks

referencia internacional importante es *Future Fund: Breakthrough* del Reino Unido, que corresponde a un programa de 425 millones de libras administrado por British Business Bank que co-invierte con el sector privado en empresas innovadoras de alto crecimiento y alta intensidad en I+D [396]. El esquema exige rondas mínimas de 20 millones de libras, y se orienta a compañías en etapa de crecimiento, y busca cerrar la brecha de capital de etapas posteriores al inicio o prototipado. El propio banco reporta que el programa ha comprometido 134 millones de libras en 18 empresas, y ha movilizado 863 millones de co-inversión privada. Esta experiencia es muy relevante ya que muestra que el escalamiento de empresas tecnológicas no solo depende de subsidios o fondos semilla, sino de instrumentos que atraen capital privado hacia rondas más grandes y de mayor riesgo tecnológico [397].

Asimismo, se recomienda **desarrollar mecanismos especializados para crecimiento y etapas tardías** con el objetivo de **fortalecer el mercado de capital para scale-ups** tecnológicas. Así, las empresas con probabilidad de crecer puedan obtener el financiamiento necesario a través de herramientas regionales. Una práctica internacional es la German *Future Fund (GFF) – EIF Growth Facility*, que representa un instrumento con un volumen de hasta €3.3 mil millones que forma parte del fondo futuro alemán de €10 mil millones, y que fue diseñado para apoyar financiamiento de crecimiento y etapas posteriores [398]. El EIF señala que a través de esta iniciativa se busca respaldar a las empresas hacia su internacionalización y liderazgo global.

En la misma línea, se propone **habilitar instrumentos de deuda de riesgo** (*venture debt*) para empresas tecnológicas que ya cuentan con inversionistas, modelo de negocio y necesidad de financiar su expansión sin recurrir a nuevas ventas de participaciones (rondas de equity). En este punto, una experiencia relevante es el mecanismo de venture debt del Banco Europeo de Inversiones (EIB), que ofrece financiamiento de largo plazo a PyMES y empresas innovadoras de media capitalización (*mid-caps*). Los montos están entre €10 millones a €50 millones, principalmente, para promover su etapa comercial [399]. El EIB destaca que este instrumento complementa el *venture capital*, y funciona como señal de calidad para atraer inversores adicionales para ampliar operaciones, fortalecer infraestructura o financiar expansión comercial sin perder control societario. Así, en 2025, un ejemplo de aplicación fue la asignación de deuda de riesgo por valor de 35 millones de euros con *Sidekick Health*, una empresa líder en salud digital de Islandia [400].

3. Estrategia exportadora de servicios tecnológicos

Si bien América Latina y el Caribe ya cuenta con empresas posicionadas en software, integración, herramientas y aplicaciones de IA, su inserción internacional sigue siendo limitada. En ese contexto, la restricción no solo está ligada a la capacidad de innovar, sino a la dificultad para transformar dichas capacidades en exportaciones sostenidas de servicios tecnológicos; y, con ello, el acceso a mercados internacionales y presencia en cadenas globales de valor. De esa forma, este objetivo está enfocado en el diseño de una estrategia exportadora para empresas tecnológicas, anclada a resolver barreras de mercado, entrada comercial, construcción de redes y generación de demanda internacional.

En esa línea, se recomienda financiar la preparación exportadora y la exploración de nuevos mercados para empresas tecnológicas que ya cuentan con un producto, servicio o modelo de negocio en etapa de desarrollo, pero sin estructura para su internacionalización. Una referencia útil es *Market Explorer de Business Finland*, que representa un instrumento para financiar la adquisición de conocimiento experto

---

[396] Fuente: British Business Bank. https://www.british-business-bank.co.uk/finance-options/equity-finance/future-fund-breakthrough

[397] Fuente: British Business Bank. https://www.british-business-bank.co.uk/for-finance-providers/equity-finance/future-fund-breakthrough/factsheet

[398] Fuente: European Investment Fund. https://www.eif.org/equity-products/all/gff-eif-growth-facility

[399] Fuente: European Investment Bank. https://www.eib.org/en/products/equity/venture-debt/index

[400] Fuente: European Investment Bank. https://www.eib.org/en/press/all/2025-196-sidekick-health-secures-eur35-million-venture-debt-from-eib-to-accelerate-rd-and-global-expansion

sobre un nuevo mercado de exportación. Según la agencia finlandesa, el programa cubre entre €5,000 y €20,000, equivalentes al 50% de los costos del proyecto, y permite financiar análisis de mercado, ventajas competitivas, canales de venta, requisitos legales y necesidades de localización del producto o servicio [401]. Esta experiencia muestra que la internacionalización no depende solo de misiones comerciales o promoción de marca país, sino también de apoyo técnico especializado para evaluar si una solución digital (o de IA) tiene condiciones de competir en un mercado externo.

Por otra parte, resulta necesaria la **creación de servicios de apoyo integral para facilitar la expansión internacional de empresas y startups a un nuevo mercado** (*soft landing*) y la conexión comercial en ecosistemas tecnológicos internacionales. De ese modo, las empresas regionales podrían reducir los costos de construcción de redes, validación de demanda o comprensión de las reglas de entrada en mercados altamente competitivos. En este punto, una experiencia internacional es Desafía [402] (iniciativa de ICEX junto con Red.es e INCIBE), que está orientada a facilitar la entrada y establecimiento de startups españolas en ecosistemas tecnológicos como San Francisco, Tel Aviv, Londres, Nueva York o Berlín. ICEX señala que el programa se encarga de asesorar, acompañar y conectar a las empresas con actores clave de cada mercado. Cada edición del programa se trabaja en grupos reducidos de 6 a 10 empresas para facilitar *networking* y contacto con actores locales relevantes [403].

En esa línea, otra experiencia importante representa la *Enterprise Europe Network* (EEN) [404] de la Comisión Europea. Esta corresponde a la mayor red de apoyo del mundo para PyMEs con posibilidades de internacionalización. La red ofrece servicios personalizados, bases de datos de oportunidades empresariales y apoyo para encontrar socios comerciales, tecnológicos o de I+D [405]. Este programa incluye acuerdos para manufactura, distribución, co-desarrollo y suministro de productos y servicios. Además, la EEN dispone de grupos sectoriales especializados y una plataforma abierta a socios internacionales para promover una política exportadora a través de la construcción de canales permanentes de articulación entre empresas, compradores e innovadores. Así, esta experiencia consolida una política aplicable a la región a través del **fortalecimiento de la inserción internacional por medio de redes de asociación empresarial, codesarrollo y acceso a oportunidades de negocio transfronterizas.**

En el caso particular de la región, que presenta fortalezas en los últimos eslabones de la cadena, también se propone la **construcción de una plataforma exportadora especializada para empresas de software y servicios tecnológicos**. De ese modo, la internacionalización depende de un entorno institucional que reduzca costos de entrada, simplifique trámites y provea servicios habilitantes para exportar. En esta dirección, la experiencia de *Software Technology Parks of India* (STPI) resulta interesante debido a que fue concebido como un esquema orientado a la exportación para software y servicios de software, con ciertas características replicables a la región: i) una ventanilla única de comercio exterior (*single window clearance*), ii) posibilidad de establecer unidades en cualquier parte del país, iii) 100% de inversión extranjera permitida; y, iv) exención arancelaria para importaciones de hardware y software [406]. Además, el STPI define entre sus objetivos promover el desarrollo y la exportación de software y servicios, proveer servicios de comunicación de datos, apoyar a MiPyMEs y crear un entorno favorable para el emprendimiento tecnológico [407].

---

401 Fuente: Business Finland. https://www.businessfinland.fi/en/services/funding/funding-services/explorer/market-explorer

402 Fuente: ICEX. https://www.icex.es/es/sala-prensa/notas-de-prensa/2025/02/nueva-iniciativa-de-colaboracion-transversal-de-icex-cdti-y-red-es-para-impulsar-la-internacionalizacion-de-startups-tecnologicas-espanolas

403 Fuente: ICEX. https://www.icex.es/es/servicios/consultoria-estrategica/rapido-crecimiento/desafia

404 Fuente: Enterprise Europe Network. https://een.ec.europa.eu

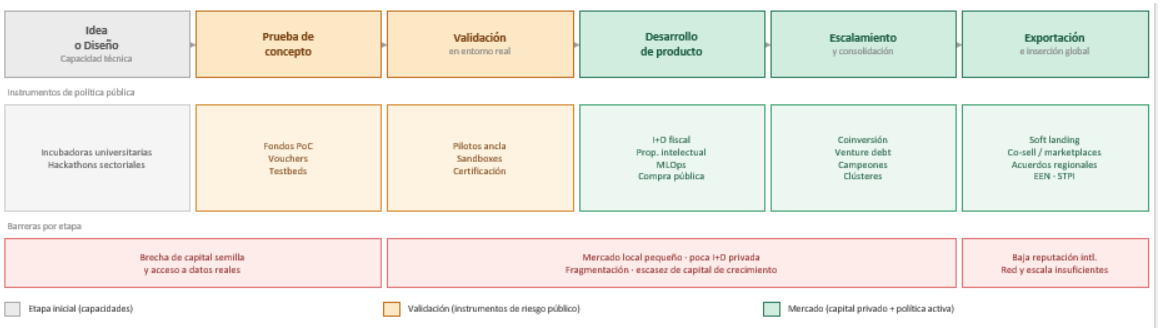
405 Fuente: Enterprise Europe Network. https://een.ec.europa.eu/news/open-invitation-join-enterprise-europe-network-international-network-partner

406 Fuente: STPI. https://stpi.in/index.php/en/stp-scheme

407 Fuente: STPI. https://stpi.in/en/stpi-objectives

Finalmente, también se requiere **promover la articulación estratégica de empresas regionales con plataformas tecnológicas globales**, de forma que la internacionalización incluya la posibilidad de integrarse a ecosistemas que concentran estándares técnicos, canales comerciales, entornos de prueba y acceso a grandes clientes. En esta dirección, una experiencia importante es la de Amazon Web Services (AWS) a través de *AWS Partner Central*, *Co-sell with AWS* y *AWS Marketplace Channel*[408], que permiten a socios tecnológicos gestionar oportunidades conjuntas con los equipos comerciales de AWS para acceder a financiamiento, demanda, colaboración, y vender o revender software y servicios profesionales.[409] Estos instrumentos buscan combinar las capacidades de nube de la empresa con la especialización de sus socios estratégicos con el objetivo de facilitar la creación de oportunidades comerciales, visibilidad y uso de herramientas integradas (co-sell y marketplace). Esta experiencia sugiere que las empresas regionales puedan obtener certificaciones, participar en programas articulados para socios, integrarse a marketplaces tecnológicos y posicionarse como implementadoras, desarrolladoras de soluciones o proveedoras especializadas dentro de ecosistemas globales. En esta línea, por ejemplo, una empresa regional como Globant ha generado estas alianzas a través de anuncios de *Strategic Partner Agreement* con Google Cloud[410] para impulsar transformaciones en AI y cloud, y también informa su estatus como AWS *Premier Tier Partner* dentro del AWS Partner Network[411], mientras que, con Microsoft, Globant se presenta como *Digital Agency & AI Inner Circle Partner*[412].

Diagrama 23
De la adopción a la oferta exportadora



Fuente: Elaboración propia.

D. Infraestructura, cómputo y derrames

Como detallado en el capítulo I, un insumo clave para el desarrollo de capacidades locales de IA es el acceso a los recursos de cómputo necesarios para el entrenamiento y validación de modelos, así como para la capacidad de operar inferencias. En ese sentido, es importante desplegar mecanismos y programas que permitan a emprendimientos locales acceder a esta capacidad.

1. Aprovechamiento estratégico de infraestructura internacional

Una parte sustancial de la infraestructura crítica para el desarrollo de tecnologías digitales e inteligencia artificial en América Latina y el Caribe (centros de datos, cloud y servicios de procesamiento) depende

[408] Fuente: AWS. https://aws.amazon.com/es/partners/co-sell-with-aws
[409] Fuente: AWS. https://aws.amazon.com/es/partners/partner-central
[410] Fuente: Globant. https://www.globant.com/news/globant-strategic-partner-agreement-google-cloud
[411] Fuente: Globant. https://www.globant.com/news/globant-aws-premier-tier-partner
[412] Fuente: Globant. https://www.globant.com/about/partners

de inversiones y plataformas de proveedores internacionales. Sin embargo, esta dependencia no implica una posición pasiva por parte de los países de la región. Por el contrario, la política puede orientarse a transformar esa infraestructura en una plataforma de desarrollo que se vincule con encadenamientos productivos, formación de capacidades, articulación con empresas nacionales, y acceso para startups y firmas tecnológicas. De esa forma, esta estrategia busca **atraer infraestructura internacional y aprovecharla estratégicamente para ampliar la base productiva del ecosistema digital regional**.

En ese sentido, una primera línea de acción consiste en **orientar la atracción de inversión en infraestructura digital hacia la generación de derrames económicos locales**. En este punto, una referencia internacional está vinculada con la expansión de la Región AWS Europa (España), donde AWS anunció una inversión de €15,700 millones hasta 2033, con una estimación de 17,500 empleos al año en negocios locales, y una contribución de €21,600 millones al PIB español. Además, la propia empresa señala que la expansión se realiza con apoyo del gobierno regional de Aragón y que los empleos esperados se distribuyen en construcción, mantenimiento de instalaciones, ingeniería y telecomunicaciones. [413] Así, la infraestructura internacional puede convertirse en un activo de desarrollo territorial cuando la política de atracción de inversión incluye la autorización de uso de instalaciones, el empleo, dinámica con proveedores y actividades complementarias en el entorno local.

Asimismo, se propone desarrollar programas de integración de emprendimientos, estándares y talento alrededor de la infraestructura internacional ya instalada. En esta dirección, una práctica internacional relevante es la impulsada por la *Malaysian Investment Development Authority (MIDA) y GDS* IDC Services en Malasia, donde el *Data Centre Supply Chain Ecosystem Summit* fue concebido para fortalecer el ecosistema local de centros de datos a través de acuerdos entre el operador internacional y actores nacionales. La autoridad de inversión de Malasia destaca que el encuentro incluyó un memorando de entendimiento entre GDS y SIRIM (Instituto de Normas e Investigación Industrial de Malasia) para fortalecer calidad y estándares de la industria, acuerdos con proveedores globales para robustecer la cadena de suministro local y programas de talento para formar profesionales digitales. Además, esas alianzas fueron asociadas a una inversión estimada de RM1,000 millones en fábricas y centros de servicio integrados; así como, una contribución anual de más de RM5,000 millones a la economía y la creación de hasta 5,000 empleos locales [414].

2. Acceso eficiente y asequible a capacidades de cómputo para IA

Los principales cuellos de botella para el desarrollo de empresas de tecnologías digitales e IA en América Latina y el Caribe se relacionan con la disponibilidad de infraestructura, y las condiciones de acceso a esa capacidad de cómputo por parte de startups, integradores y desarrolladores de producto. El diagnóstico regional muestra que la infraestructura de la nube en la región continúa altamente concentrada en pocos países y operadores. Al mismo tiempo, el capítulo I en su sección 3 señala que los neo-clouds surgieron para responder a una demanda insatisfecha de acceso a GPU y que, en ciertos casos, pueden ofrecer precios más bajos que los hiperescaladores. En la misma línea, la experiencia de modelos locales como Maritaca AI en Brasil muestra que la reducción de costos de inferencia es un factor crítico para empresas que facturan en moneda local, pero deben pagar APIs y servicios de cómputo en dólares, quedando expuestas a volatilidad cambiaria y a mayores barreras de entrada. De esa forma, este objetivo propone una política para **facilitar el acceso eficiente y asequible a capacidades de cómputo orientándola a reducir los costos de uso, experimentación y escalamiento para empresas tecnológicas regionales**.

[413] Fuente: AWS. https://aws.amazon.com/es/local/spain/

[414] Fuente: MIDA. https://www.mida.gov.my/media-release/setting-the-stage-gds-summit-showcases-data-centre-supply-chain-ecosystem-further-strengthening-malaysias-digital-infrastructure/

Por ello, se propone **establecer mecanismos de acceso compartido y subsidiado a capacidad de cómputo para startups y PyMEs tecnológicas**. Así, las empresas podrían recibir financiamiento parcial para el entrenamiento, *fine-tuning*, inferencia o pruebas avanzadas de sus modelos. En esta dirección, la experiencia de las *AI Factories* de EuroHPC resulta importante porque fue diseñada para abrir el acceso a supercomputación e infraestructura avanzada a PyMEs. El esquema prevé modalidades diferenciadas para innovación industrial y establece que el acceso para startups y PyMEs de IA puede ser gratuito para fines de innovación. Aquí se combinan capacidades computacionales con apoyo técnico y entrenamiento [415].

Asimismo, se plantea **acompañar el acceso a cómputo con servicios de asesoría, formación e infraestructura operativa**, de manera que las empresas dispongan de apoyo para utilizar la capacidad de procesamiento de forma eficiente. En este punto, una experiencia internacional representa el *AI Service Centre Berlin Brandenburg* del *Hasso Plattner Institute*. Esta iniciativa corresponde a uno de los cuatro centros alemanes financiados por el gobierno federal, que ofrece de forma gratuita tres tipos de servicios integrados: consultoría, formación e infraestructura. La relevancia de esta práctica para la región muestra que el acceso asequible a cómputo debe entenderse como una combinación de infraestructura y capacidades de uso para empresas que todavía no dominan por completo el ciclo de desarrollo, operación y despliegue de soluciones de IA [416].

Finalmente, se recomienda articular el acceso a cómputo con alianzas público-privadas que acerquen a startups y PyMEs a infraestructura de frontera y servicios especializados. En esa línea, la experiencia de *AI Factory France* resulta interesante. Bajo el liderazgo de GENCI, la iniciativa anunció en 2025 la primera colaboración europea entre una AI Factory y los programas *NVIDIA Inception* y *NVIDIA Connect*, con el objetivo de ampliar el acceso de startups y PyMEs a infraestructura de cómputo y servicios especializados. Además, GENCI reporta que desde 2019 el supercomputador *Jean Zay* ha apoyado a más de 250 startups y facilitado más de 1,400 proyectos de IA por año [417]. Asimismo, como fue documentado en el capítulo IV, la experiencia de LatamGPT que recibió el apoyo de Amazon Web Services para el entrenamiento de la primera versión del modelo es emblematico del apoyo que podrían proporcionar los operadores internacionales en la región.

3. Desarrollo de modelos soberanos

La construcción de modelos soberanos en la región no se enmarca como un intento de competir de forma directa con los modelos fundacionales de frontera desarrollados por Estados Unidos o China, sino como una estrategia de adaptación tecnológica orientada a resolver problemas concretos de idioma, sesgo cognitivo, contexto productivo, confidencialidad de datos, autonomía operativa, y, por sobre todo, construcción de capacidades. El capítulo III en su primera sección muestra que la necesidad de esta adaptación a los modelos surge, entre otros, por los sesgos lingüísticos, culturales y cognitivos incorporados en modelos entrenados para otros mercados; y, por otro, para que los gobiernos y sectores regulados puedan procesar información sensible en entornos seguros y de alto desempeño. En ese marco, los casos de Maritaca AI y LatamGPT muestran dos rutas para la región: i) especialización privada sobre desarrollos locales de alta calidad; y, ii) la colaboración público-privada de escala regional sin fines comerciales.

En esa línea, se propone **promover la adaptación y especialización de modelos abiertos a dominios y necesidades económicas de la región**, en lugar de priorizar, desde el inicio, modelos generales de gran escala que están fuera del alcance de los recursos locales. Esta política permitiría

[415] Fuente: The European High Performance Computing Joint Undertaking (EuroHPC JU). https://www.eurohpc-ju.europa.eu/ai-factories/ai-factories-access-modes_en

[416] Fuente: Hasso Plattner Institute. https://hpi.de/en/ai-service-centre

[417] Fuente: Grand Equipement National de Calcul Intensif. https://www.genci.fr/actualites/fast-track-success-ai-factory-france-supercharges-ai-adoption-french-startups-and-smes

orientar recursos hacia el *fine-tuning*, desarrollo de conjuntos de datos depurados, evaluación de corpus multilingüe y la construcción de modelos específicos para sectores como administración pública, justicia, salud, finanzas, agricultura o educación. De esta manera, la experiencia europea de *OpenEuroLLM*[418] resulta útil, ya que fue concebida como una familia de modelos abiertos que cubre todas las lenguas oficiales de la Unión Europea. Esta iniciativa se apoya en supercomputación europea y en un enfoque de transparencia, apertura y cumplimiento regulatorio. La Comisión Europea indica que el proyecto cuenta con un presupuesto total de €37.4 millones, de los cuales €20.6 millones provienen del programa Europa Digital, y señala que su objetivo es facilitar que startups, PyMEs e innovadores europeos adopten y desplieguen IA de manera costo-efectiva[419]. Para la región, esta parte demuestra que la soberanía tecnológica puede construirse sobre modelos abiertos, multilingües y adaptables, sin exigir que cada país replique por sí solo todo el esfuerzo de frontera.

Por otro lado, otro objetivo importante radica en cerrar la brecha de financiamiento de los modelos soberanos mediante esquemas combinados de subsidio, créditos de cómputo, cofinanciamiento y apoyo a consorcios. En ese punto, resulta importante al señalar que, dado el perfil de riesgo de este tipo de emprendimientos, corresponde al Estado, la banca multilateral e instrumentos especializados asumir una parte central del financiamiento de iniciativas innovadoras y startups; así como, desarrollar oportunidades de fondeo para proyectos intensivos en tecnología[420]. Este apoyo es fundamental en el despliegue de IA soberana, ya que el propio diagnóstico regional muestra una fuerte asimetría entre el presupuesto inicial de LatamGPT (US$ 800,000) y los montos movilizados por otros países. Una referencia internacional destacada es *IndiaAI Mission*[421], aprobada con una asignación de $1.25 mil millones, estructurada en pilares que incluyen capacidad de cómputo, modelos fundacionales, datasets, aplicaciones, talento, financiamiento de startups y seguridad[422]. Además, esta recomendación gana sustento empírico, dado que América Latina y el Caribe no enfrenta únicamente una carencia de recursos para acompañar proyectos soberanos, sino que mantiene una brecha profunda entre el tipo de modelo que busca construir y el volumen de financiamiento que ese objetivo requiere para ser técnicamente viable y comparable con experiencias más robustas.

Asimismo, este análisis obliga entender que el problema no se resuelve, únicamente, ampliando fondos de arranque para startups o creando mecanismos generales de apoyo a la innovación, porque la comparación internacional muestra que los **proyectos más robustos financian el entrenamiento del modelo, así como capas complementarias de multimodalidad, infraestructura pública de cómputo, producto, pilotos, verificación regulatoria y ecosistema institucional**. En otras palabras, el caso de LatamGPT no evidencia únicamente una subinversión cuantitativa, sino también una subinversión en la densidad programática que acompaña a los modelos soberanos más avanzados. Por ello, la premisa inicial debería sostener que **para cerrar la brecha de financiamiento se requiere, efectivamente, subsidios, créditos de cómputo, cofinanciamiento y apoyo a consorcios, pero bajo una arquitectura escalonada que distinga entre financiamiento del núcleo de entrenamiento, financiamiento de infraestructura y cómputo soberano, y financiamiento de capas posteriores de despliegue y validación**.

De la misma forma, es importante mencionar la necesidad de **crear entornos soberanos de despliegue para el sector público y actividades reguladas** con la finalidad de que los modelos adaptados localmente puedan ejecutarse en infraestructuras controladas (con trazabilidad, seguridad y cumplimiento normativo), sin depender de APIs cerradas o de nubes externas no alineadas con los requerimientos

---

[418] Fuente: Strategic Technologies for Europe Platform: https://strategic-technologies.europa.eu/step-results/step-stories/ openeurollm_en

[419] Fuente: Comisión Europea. https://digital-strategy.ec.europa.eu/en/news/pioneering-ai-project-awarded-opening-large-language-models-european-languages

[420] Katz, R. et al (2023). *Estrategia y recomendaciones de política pública para el fortalecimiento del ecosistema de ciencia, tecnología e innovación en América Latina y el Caribe: Lecciones aprendidas desde el caso argentino.* CAF

[421] Fuente: India AI. https://indiaai.gov.in/

[422] Fuente: Government of India Press Information Bureau. https://www.pib.gov.in/PressReleaseIframePage.aspx?PRID=2030838&reg=3&lang=2

institucionales. Esta línea resulta altamente pertinente porque las organizaciones públicas procesan información que requiere entornos seguros, así como el desempeño y disponibilidad elevados. Una experiencia internacional reciente que promueve esta política es la alianza anunciada por Francia y Alemania con *Mistral AI* y SAP para lanzar una IA soberana destinada a la administración pública [423]. Esta iniciativa está concebida como una asociación público-privada para mejorar la eficiencia y resiliencia de los servicios públicos europeos. Asimismo, la asociación entre *Dassault Systèmes* y *Mistral AI* lleva soluciones empresariales de IA a una nube soberana con los más altos estándares de seguridad, lo que muestra su utilidad para industrias reguladas y sector público que priorizan confidencialidad, control y cumplimiento [424].

De todas maneras, es importante remarcar que, en lugar de *soberanía fundacional* (donde se entrenan modelos generales competitivos a nivel internacional), la región debe orientarse al Desarrollo de modelos en el marco de una *soberanía estratégica* (donde se mantiene el control sobre casos de uso críticos, datos sensibles y modelos sectoriales).

4. Alineamiento geopolítico selectivo

Como demostrado en el capítulo III, el desarrollo de la IA se encuentra cada vez más condicionado por una competencia geopolítica entre grandes bloques que controlan factores de producción (chips avanzados, infraestructura de cómputo, modelos fundacionales, software especializado y estándares regulatorios). En ese contexto, la tercera sección del capítuloIII explica que Estados Unidos busca eliminar barreras a la innovación, mientras que China procura desplegar una cadena de valor paralela con factores de producción soberanos, y Europa combina competitividad tecnológica con gobernanza, seguridad y regulación. Para América Latina y el Caribe, la principal implicancia es que difícilmente podrá acceder de manera autónoma a todos estos factores, por lo que requiere una estrategia de alineamiento selectivo que le permita acceder a infraestructura, conocimiento, financiamiento y mercados sin comprometer su margen de maniobra. De esa forma, se busca que esta estrategia no se conciba como una adhesión completa a un solo bloque, sino como una política diferenciada según eslabón, función tecnológica y capacidad productiva.

En ese sentido, se propone **adoptar una estrategia de diversificación de alianzas tecnológicas según el tipo de capacidad que se busca desarrollar**. La experiencia de India resulta importante porque muestra que un país de gran escala puede construir autonomía relativa sin aislarse de los principales polos tecnológicos. Así, la *India Semiconductor Mission* [425] fue concebida para posicionar al país como un *hub* global de manufactura y diseño, pero al mismo tiempo su propia hoja de ruta oficial muestra una estrategia abierta a acuerdos con múltiples socios. Asociaciones con Singapur, la Unión Europea, Japón y Estados Unidos, así como, vínculos con IBM, Lam Research, Purdue University y otros actores tecnológicos indican su margen de independencia en la adhesión por bloques [426]. Este enfoque asegura el acceso a diseño, cadena de suministros, investigación, talento y manufactura mediante una red plural de acuerdos. Para la región, esta selección podría organizarse por países o segmentos donde se busque cooperación con proveedores estadounidenses en infraestructura de cómputo y modelos, combinada con socios asiáticos, en componentes y manufactura, y con Europa en estándares, interoperabilidad y gobernanza.

Otro enfoque importante radica en **utilizar el alineamiento geopolítico como instrumento para ascender en la cadena de valor**. En este punto, la experiencia de Malasia resulta relevante. La *National Semiconductor Strategy* (NSS) parte de una fortaleza previa en actividades de ensamblaje y prueba de

---

[423] Fuente: Ministerio de Economía, Finanzas y Soberanía Industrial, Energética y Digital. https://presse.economie.gouv.fr/france-and-germany-join-forces-with-mistral-ai-and-sap-se-to-launch-a-sovereign-ai-for-public-administration

[424] Fuente: Dassault Systèmes'. https://www.3ds.com/newsroom/press-releases/new-era-sovereign-ai-dassault-systemes-and-mistral-ai-deepen-their-partnership

[425] Fuente: India Science, Technology and Innovation Portal. https://www.indiascienceandtechnology.gov.in/st-visions/national-mission/india-semiconductor-mission-ism

[426] Fuente: Government of India Press Information Bureau. https://www.pib.gov.in/PressNoteDetails.aspx?NoteId=157237&ModuleId=3&reg=3&lang=1

semiconductores subcontratados (*Outsourced Semiconductor Assembly and Test*-OSAT) y propone avanzar hacia ensamblaje avanzado, diseño de circuitos electrónicos, manufactura de equipos terminales y otras funciones de mayor valor[427]. Así, el gobierno malayo, en febrero de 2025, anunció que la diplomacia de semiconductores pasaría a ser un pilar estratégico de la NSS, con el objetivo de fortalecer cooperación con socios internacionales y con la Asociación de Naciones del Sudeste Asiático (ASEAN)[428] para asegurar una cadena de suministro sostenible y resiliente. A esto se suma que la estrategia ha sido acompañada por atracción de alianzas concretas, como la colaboración entre NVIDIA y YTL para infraestructura de IA, y por programas de MIDA orientados a integrar proveedores malasios en cadenas globales de empresas como Micron, ASML y Ferrotec[429].

Finalmente, es necesario **diferenciar el alineamiento por tipo de país y por nivel de capacidad**, con el objetivo de evitar una política regional uniforme. Los países de desarrollo medio como Brasil, Chile o Uruguay pueden apalancar sus sistemas educativos y ecosistemas de innovación para jugar un papel activo en el escenario competitivo internacional, mientras que otros países deberían concentrarse en aprovechar la inversión extranjera en centros de datos, formar talento y especializarse en adaptación de tecnologías para su aparato productivo. En ese sentido, el alineamiento selectivo debería reconocer que no todos los países necesitan el mismo tipo de socio ni persiguen el mismo objetivo. Así, unos países pueden priorizar el acceso a capacidad de cómputo, mientras que otros pueden anticipar alianzas para ensamblaje, dispositivos de borde (IoT) o manufactura intermedia, mientras que otros pueden abarcar la cooperación para adaptar modelos lingüísticos, gobernanza o despliegue público seguro. Bajo este enfoque, el alineamiento geopolítico se podría convertir en una herramienta de política de desarrollo productivo, orientada a reducir dependencias críticas, ampliar acceso a factores de producción y fortalecer posiciones regionales en aquellos espacios donde existen oportunidades viables.

La estrategia de alineamiento geopolítico selectivo adquiere mayor precisión cuando se organiza por eslabón o insumo crítico de la cadena de valor. Cada segmento presenta requerimientos distintos en términos de socios, modalidades de acceso y condiciones habilitantes, por lo que una política regional uniforme resulta ineficaz. En materiales críticos, el alineamiento prioritario apunta hacia la Unión Europea y hacia países de Asia-Pacífico que demandan suministros estratégicos, con el objetivo de negociar condiciones de industrialización y transferencia tecnológica como contrapartida a la exportación de recursos. En la manufactura de equipamiento y ensamblaje, la proximidad logística al mercado norteamericano convierte al alineamiento con Estados Unidos en una oportunidad concreta de inserción en cadenas globales de *nearshoring*, mientras que la cooperación con Taiwán y Corea del Sur resulta relevante para la adopción de estándares técnicos y redes de proveedores de componentes. En la infraestructura de cómputo y nube, el objetivo no reside en el control de la infraestructura sino en la negociación de las condiciones de su despliegue: derrames locales de empleo, formación de proveedores, créditos de cómputo para empresas nacionales. En los modelos fundacionales, la estrategia más factible para la región combina el acceso a modelos de pesos abiertos —provenientes tanto de iniciativas estadounidenses como europeas— con la cooperación regional para el desarrollo de datasets multilingües, capacidades de fine-tuning sectorial e infraestructura compartida de supercómputo. **Finalmente, en los eslabones de software y aplicaciones, donde la región ya dispone de presencia** empresarial, el alineamiento se orienta hacia la integración en ecosistemas globales de distribución y la obtención de certificaciones que amplíen la reputación y el acceso a mercados externos.

El cuadro 32 sintetiza esta diferenciación, identificando para cada eslabón la capacidad buscada, los socios estratégicos prioritarios, la modalidad de acceso recomendada y las principales condiciones y

[427] Fuente: MIDA. https://www.mida.gov.my/mida-news/national-semiconductor-strategy-to-guide-industry-up-value-chain
[428] Fuente: MIDA. https://www.mida.gov.my/mida-news/nss-framework-passed-to-develop-110-local-high-tech-companies-miti/
[429] Fuente: MIDA. https://www.mida.gov.my/media-release/malaysia-seeks-global-investors-to-advance-its-semiconductor-value-chain-at-semicon-sea-2025

riesgos asociados. La lectura del cuadro debe hacerse de manera horizontal —eslabón por eslabón— y no como una política uniforme aplicable al conjunto de la cadena.

Cuadro 32
Matriz de alineamiento geopolítico selectivo por eslabón de la cadena de valor de la IA en ALC

Eslabón/Insumos de la cadena	Capacidad objetivo	Socio(s) estratégico(s) y bloque	Modalidad de acceso	Riesgo / condición habilitante
Materiales críticos y minería	Refinación, procesamiento y mayor valor agregado sobre litio, cobre, níquel y tierras raras	UE (CRMA, cadenas de baterías y vehículos eléctricos) · Japón/ Rep de Corea (, compras de largo plazo, asistencia técnica) · Multilateral (BID, CAF, mecanismos de industrialización)	Condiciones de industrialización en contratos de IED · diplomacia regional coordinada · transferencia tecnológica como contrapartida · zonas económicas especiales para refinación	Requiere coordinación entre países productores de la región (la fragmentación genera competencia interna) · capacidad institucional minera para negociar · riesgo de tensiones con inversores si se imponen restricciones a exportación de materias primas brutas
Equipamiento y ensamblaje de hardware	Ensamblaje de servidores, racks, dispositivos edge AI e IoT industrial; packaging de semiconductores en nichos seleccionados	EE.UU. (nearshoring, Foxconn/Inventec/Quanta) · Taiwán/Rep de Corea del Sur (estándares OSAT, redes de componentes, manufactura avanzada)	Atracción de IED con contenido local progresivo · zonas francas tecnológicas · programas de proveedores locales con empresas globales · adopción de estándares de calidad internacionales	Dependencia de decisiones de deslocalización global · riesgo de quedar atrapado en ensamblaje de bajo valor agregado · requiere infraestructura industrial preexistente (México, Costa Rica, Brasil)
Infraestructura de cómputo y nube	Acceso a capacidad de cómputo, reducción de latencia regional, derrames locales de inversión extranjera	EE.UU. (AWS, Azure, Google Cloud: inversión en centros de datos, créditos para startups) · UE (GAIA-X, estándares de interoperabilidad, European Alliance for Industrial Data and Cloud)	Negociación de derrames locales en licencias de operación · acuerdos de interoperabilidad y estándares abiertos · créditos de cómputo subsidiados para PyMEs y startups · programas de talento y proveedores locales vinculados a la inversión extranjera	Riesgo de vendor lock-in en plataformas de un único proveedor · fuga de valor fiscal hacia sedes externas · requiere regulación nacional de datos y definición de soberanía digital
Modelos fundacionales y soberanos	Acceso a modelos de pesos abiertos, fine-tuning sectorial y multilingüe, supercómputo para entrenamiento y adaptación	EE.UU. (modelos abiertos: Llama) · UE (OpenEuroLLM, EuroHPC, Mistral) · Cooperación regional ALC (LatamGPT, CENIA, RedCLARA, datasets sectoriales)	Adaptación de modelos de pesos abiertos para dominios y lenguas regionales · cofinanciamiento multilateral de proyectos soberanos · acuerdos de acceso a créditos de cómputo · cooperación en datasets multilingües y sectoriales entre instituciones regionales	Brecha de financiamiento vs. laboratorios globales de frontera · riesgo de dependencia de APIs cerradas para casos de uso críticos · requiere coordinación sostenida entre universidad, empresa y gobierno más allá de ciclos electorales
Software, herramientas y plataformas	Integración en ecosistemas globales, acceso a canales de distribución internacionales, certificación y adopción de estándares	EE.UU. (AWS Partner Network, Google Cloud Partner Advantage, Azure Marketplace, programas de co-sell) · UE (estándares de interoperabilidad, contratación pública europea, Enterprise Europe Network)	Programas de socios tecnológicos en hiperescaladores · acreditación y presencia en marketplaces globales · adopción de estándares UE para acceso a mercados europeos · co-sell y alianzas estratégicas con plataformas globales	Riesgo de dependencia de plataformas privadas para el acceso al mercado · condiciones contractuales y comisiones restrictivas · requiere producto validado con tracción comercial y equipo técnico certificado
Aplicaciones e implementación sectorial	Exportación de servicios tecnológicos, acceso a mercados externos, reconocimiento y reputación internacional	EE.UU. (ecosistemas de startups, Silicon Valley, programas de soft landing) · UE (Enterprise Europe Network, contratos públicos europeos, programas de coinnovación) · Regional ALC (acuerdos de reconocimiento mutuo, plataformas exportadoras conjuntas)	Programas de soft landing en ecosistemas tecnológicos internacionales · acuerdos de reconocimiento mutuo de certificaciones · redes de asociación empresarial (EEN, equivalentes regionales) · plataformas exportadoras especializadas en software y servicios (modelo STPI)	Baja reputación internacional de empresas tecnológicas regionales · ciclos comerciales largos en mercados externos · requiere estrategia articulada de marca-país tecnológica y financiamiento de la etapa de entrada a nuevos mercados

Fuente: Elaboración propia.
Nota: a) El alineamiento propuesto no implica adhesión exclusiva a un bloque geopolítico, sino una estrategia diferenciada según el tipo de capacidad buscada en cada eslabón. Los países de la región deben calibrar su posicionamiento según su nivel de desarrollo institucional y productivo. Países con mayor madurez de ecosistema (Brasil, Chile, Uruguay) pueden apalancar alianzas más complejas en eslabones superiores, mientras que otros países pueden concentrarse en los eslabones de infraestructura, ensamblaje y aplicaciones. b) China no se incluye como socio prioritario en ningún eslabón dado el contexto de restricciones tecnológicas vigentes, aunque algunos países de la región pueden acceder a infraestructura cloud o modelos de origen chino en condiciones de autonomía tecnológica.

E. Talento, conocimiento aplicado y desarrollo tecnológico

1. Formación de talento aplicado a perfiles de diseño y desarrollo

El desarrollo de empresas de inteligencia artificial en América Latina y el Caribe depende tanto de la disponibilidad de profesionales (ingenieros o programadores) en términos agregados, como de la capacidad del sistema educativo y de formación para generar perfiles aplicados a diseño, desarrollo, integración y transferencia tecnológica hacia el aparato productivo. En esa línea, una de las fallas de coordinación más significativas a nivel regional es la ausencia de canales que permitan anticipar las necesidades de capital humano por parte del sistema productivo. Las mejores prácticas internacionales combinan, al menos, cuatro mecanismos para resolver esta falla: i) estimación de demanda de habilidades, ii) actualización curricular permanente en formación terciaria y vocacional, iii) programas de formación dual y continua articulados con institutos especializados y universidades, y iv) creación de centros interuniversitarios de investigación y transferencia en tecnologías avanzadas. Por ello, esta línea estratégica busca la formación de talento aplicado como un instrumento para reducir la brecha entre oferta y demanda laboral para fortalecer la base de recursos humanos en actividades de diseño y desarrollo tecnológico.

De esta manera, se recomienda **desarrollar sistemas de anticipación y monitoreo de la demanda de perfiles de diseño y desarrollo** con el objetivo de que la política de formación se organice desde la oferta educativa disponible y desde las necesidades de las empresas y sectores productivos. Esta recomendación identifica como práctica central la preparación de estudios de estimación de demanda de innovación en IA, utilizando observatorios, encuestas de empleadores, grupos focales y ejercicios prospectivos. Una experiencia internacional en esta dirección es la de Irlanda a través del *Expert Group on Future Skills Needs* (EGFSN), que representa el órgano oficial que asesora al gobierno sobre requerimientos futuros de habilidades y que ha producido estudios específicos para sectores estratégicos [430]. En particular, su informe *Together for Design* [431] se concentra en tres disciplinas consideradas clave para el país (digitalización, producción y diseño estratégico). Además, considera que estos perfiles son cada vez más demandados por la transformación tecnológica e industrial. Esta experiencia es importante para América Latina, ya que muestra que la formación en diseño y desarrollo en IA debe partir de análisis prospectivos de la demanda, de modo que los programas académicos y vocacionales respondan a oportunidades reales de especialización.

En esa misma línea, también se requiere **fortalecer programas de formación dual y trayectorias prácticas de transición al sector productivo**, especialmente, para perfiles ligados a desarrollo de software, integración de sistemas, diseño de productos digitales, automatización y tecnología. El de asistencia técnica puede extenderse a programas de formación dual y continua, los cuales deben ser articulados con institutos especializados y centros de transferencia de las universidades. Adicionalmente, los procesos de transición de graduados al sistema productivo representan un componente central para corregir brechas de habilidades. En este punto, una referencia internacional está vinculada con el sistema dual de formación profesional de Alemania, donde la capacitación combina aprendizaje práctico en una empresa con formación teórica en una escuela vocacional. La Agencia Federal de Empleo alemana destaca que en este modelo la práctica y la teoría se alternan durante varios años para que el alumno adquiera capacidades en el entorno productivo [432].

[430] Fuente: Expert Group on Future Skills Needs. https://enterprise.gov.ie/en/what-we-do/workplace-and-skills/skills-for-enterprise
[431] Fuente: Department of Enterprise, Tourism and Employment. https://enterprise.gov.ie/en/publications/together-for-design.html
[432] Fuente: Federal Employment Agency. https://www.arbeitsagentur.de/int/en/vocational-training

Otra línea de acción consiste en crear y fortalecer centros aplicados de investigación, desarrollo y transferencia vinculados a universidades, empresas y clústeres tecnológicos, con el objetivo de formar talento para resolver problemas reales de diseño, prototipado, integración y desarrollo de producto. Esta recomendación incluye la creación de centros interuniversitarios de investigación en tecnologías avanzadas como un mecanismo de transferencia entre el sector académico y el sistema productivo. Además, utiliza la experiencia argentina (Instituto Balseiro, CETRI de la Universidad del Litoral e INTIA de Tandil) como ilustración del papel que cumplen estos centros en la transferencia de conocimiento al aparato productivo. Otra experiencia internacional relevante es la de los centros *Fraunhofer* en Alemania, cuya misión es transferir conocimiento e innovación hacia la economía y la sociedad mediante proyectos industriales, asociaciones público-privadas, formación continua y estandarización [433].

2. Retención de talento y masa crítica regional

Considerando que las necesidades de la cadena de valor de IA son heterogéneas según el tipo de vertical y el tipo de eslabón, la retención de talento especializado resulta un aspecto crítico a tomar en cuenta para promover el despliegue tecnológico. En los segmentos más avanzados, como el desarrollo de modelos fundacionales, herramientas base y capacidades soberanas, la región requiere investigadores de alto nivel, equipos de investigación y perfiles doctorales capaces de trabajar en entrenamiento, evaluación, eficiencia y adaptación de modelos. En los eslabones de software, integración e implementación empresarial, las necesidades se vinculan con la presencia de ingenieros aplicados, perfiles de producto e ingenieros de despliegue e integración, capaces de adaptar modelos y herramientas a procesos reales de negocio (denominados *forward deployment engineer*). Los propios insumos de diagnóstico señalan que América Latina y el Caribe debería desarrollar recursos humanos para adaptar la IA a las necesidades del aparato productivo, y que las iniciativas educativas y la creación de centros de ingeniería de alto nivel deben contribuir a retener una masa crítica de investigadores y desarrolladores como base del ecosistema futuro.

Así, una línea de acción podría estar enfocada en **concentrar y financiar talento avanzado para investigación y desarrollo de capacidades base** en áreas vinculadas a modelos, herramientas y componentes tecnológicos de mayor complejidad. El enfoque de esta medida responde a focalizar recursos en polos, consorcios y centros capaces de alcanzar masa crítica. En esta línea, la experiencia de *India AI Future Skills* [434] resulta interesante porque incluye un esquema explícito de apoyo a talento avanzado con becas doctorales, además de trayectorias diferenciadas para posgrados y pregrados. Hacia diciembre de 2025, el programa reportaba apoyo a 500 doctorandos, 5,000 posgraduados y 8,000 estudiantes de pregrado, dentro de una arquitectura nacional orientada a construir una oferta escalonada de talento para IA [435]. Esta experiencia muestra que la retención de talento avanzado se logra con instrumentos específicos de financiamiento y prestigio para investigadores de alto nivel en áreas tecnológicas priorizadas.

Tomando en cuenta la referencia de *AI Apprenticeship Programme* (AIAP) de *AI Singapore* [436], creado para la atracción de talento local para trabajar en proyectos de IA aplicada, es necesario aplicar esta iniciativa en la región para **desarrollar programas intensivos de formación aplicada para ingenieros de producto, integración y despliegue empresarial.** Esta medida es importante porque el propio diagnóstico de cadena de valor muestra que la integración de IA en organizaciones enfrenta barreras de adaptación operativa, lo que incrementa la necesidad de perfiles capaces de trabajar con datos, modelos, despliegue y procesos de negocio en entornos reales.

---

[433] Fuente: Fraunhofer. https://www.fraunhofer.de/en/research/range-of-services/technology-transfer.html

[434] Fuente: India AI. https://indiaai.gov.in/article/indiaai-fellowship-under-the-indiaai-mission

[435] Fuente: India AI. https://indiaai.gov.in/article/indiaai-mission-announces-fellowships-for-phd-scholars

[436] Fuente: AI Singapore. https://aiap.sg/apprenticeship Nota: El programa combina dos meses de deep-skilling con siete meses de entrenamiento en proyectos reales, y está diseñado para formar AI engineers con capacidad de construir modelos, pipelines y desplegar soluciones en producción para empresas y organismos.

En esa línea, también se requiere crear trayectorias de carrera, redes regionales y entornos institucionales capaces de retener talento en clústeres y centros tecnológicos, con el objetivo de evitar que la región forme recursos humanos que luego se desvinculen por falta de oportunidades, financiamiento o proyección internacional. Esta medida se asocia con las recomendaciones del informe de fortalecimiento de ecosistema de ciencia y tecnología sobre concentración de recursos humanos, creación de centros interuniversitarios y fortalecimiento de redes entre clústeres. A nivel internacional, una experiencia relevante es el marco europeo de atracción y retención de talento investigador, reforzado por la iniciativa *Choose Europe for Science*, que agrupa más de 100 esquemas nacionales y regionales para atraer, desarrollar y retener investigadores, complementados por cerca de €900 millones de apoyo a nivel europeo. [437] En ese sentido, es importante mencionar que la retención requiere carreras más atractivas, movilidad dentro de redes científicas y empresariales, condiciones de trabajo estables y conexión con proyectos de frontera. En otras palabras, la masa crítica se construye formando talento, y proveyendo razones concretas para quedarse y crecer dentro del ecosistema regional.

3. Vinculación entre empresa y universidad

La relación entre empresa y universidad representa uno de los mecanismos más importantes para transformar capacidades científicas y académicas en soluciones productivas, nuevos servicios; y, por tanto, en empresas de base tecnológica. En la región, la barrera para llegar a estas sinergias radica en la escasez de investigación; así como, una débil transferencia de conocimiento hacia el aparato productivo, ausencia de procesos colaborativos para desarrollar soluciones innovadoras y limitada articulación entre educación superior, centros tecnológicos y empresas. Uno de los problemas centrales es la falta de alineamiento entre el esfuerzo del sector académico con las necesidades del sistema productivo, mientras que destaca que una de las respuestas más relevantes consiste en financiar redes de centros tecnológicos y mecanismos de transferencia que conecten ambos actores. En esa misma línea, es relevante la experiencia argentina, donde instituciones como el Instituto Balseiro, el CETRI de la Universidad Nacional del Litoral o el INTIA en Tandil actuaron como plataformas de transferencia de conocimiento hacia sectores productivos concretos.

Por ello, es necesario **crear y fortalecer instituciones puente entre universidad y empresa,** con capacidad de ofrecer infraestructura compartida, servicios de transferencia, experimentación aplicada y desarrollo conjunto de soluciones. Esta recomendación plantea el financiamiento de redes de centros tecnológicos como uno de sus ejes de intervención para resolver la articulación entre conocimiento y producción. Entre las mejores prácticas internacionales están *Manufacturing USA* y los Institutos *Fraunhofer* porque reúnen recursos académicos y empresariales en torno a problemas productivos y acceso compartido a infraestructura de I+D. En esa misma línea, una experiencia internacional de referencia útil es la *Catapult Network* del Reino Unido, que representa una red de nueve centros de innovación con más de 65 ubicaciones que ofrece instalaciones avanzadas de I+D y desarrollo técnico para apoyar a las empresas en la transición desde concepto hasta adopción en mercado [438]. La red fomenta la colaboración entre academia e industria, y que su función es ser el puente para cerrar la brecha entre investigación e industria lo que la convierte en un referente para estructurar centros de transferencia orientados a tecnologías digitales, IA, manufactura avanzada o telecomunicaciones [439].

[437] Fuente: Comisión Europea. https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/choose-europe-science-over-100-national-and-regional-initiatives-aim-attract-global-research-talent-2026-01-30_mt

[438] Fuente: Catapult Network. https://catapult.org.uk

[439] Fuente: Catapult Network. https://catapult.org.uk/about-us/why-the-catapult-network

Otra línea de acción consiste en **desarrollar programas de codesarrollo y movilidad de conocimiento entre universidades y empresas**, de forma que la relación se traduzca en proyectos concretos, conformación de equipos multidisciplinarios y talento focalizado en desafíos productivos reales. En esta dirección, una experiencia internacional relevante es los *Knowledge Transfer Partnerships* (KTPs) del Reino Unido, que han ayudado a empresas a innovar mediante colaboraciones de largo plazo con universidades y organización de investigación. [440] El programa está diseñado para unir empresa, academia y graduados altamente calificados en proyectos de innovación de 12 a 36 meses de duración. Según información oficial se destaca que las empresas pueden acceder directamente a experiencia académica y talento humano, mientras que las universidades aplican investigación a problemas reales, lo cual abre variadas posibilidades comerciales. Esta experiencia, a través de *Innovate* UK, abrió recientemente una nueva ronda de transferencia de conocimiento (2026–2027) con financiamiento de hasta 10 millones de libras, que cubre hasta 67% del costo del proyecto para PyMEs [441]. Esta experiencia ofrece un modelo de arquitectura concreta para que la transferencia de conocimiento ocurra a través de la alineación entre los objetivos del sector productivo y las capacidades de investigación del centro tecnológicos y universitarios.

Es importante destacar la necesidad de institucionalizar la transferencia tecnológica, asesorar en el proceso de patentes y acompañar proyectos de cooperación con empresas. En ese sentido, se recomienda **fortalecer mecanismos universitarios de comercialización, spin-offs y creación de startups**, con el objetivo de que la investigación con potencial de acceder al mercado se vincule con los esfuerzos empresariales. Una referencia internacional valiosa es *Cambridge Enterprise* de la Universidad de Cambridge, que combina transferencia tecnológica, fondos para pruebas de concepto, fondos semilla y creación de startups [442]. Su *Technology Investment Fund* dispone de 10 millones de libras en cinco años para financiar experimentos que reduzcan riesgo tecnológico, mientras la unidad de inversión universitaria reporta 28 inversiones entre 2023 y 2024; así como, 6.47 millones de libras invertidos en 37 spinouts. Además, el *University of Cambridge Enterprise Fund* ha levantado 30.4 millones de libras y las empresas de su cartera han captado colectivamente más de 1,000 millones en capital sindicado [443]. La lección para la región radica en la vinculación entre la universidad y empresa a través de la investigación y gestión de patentes, así como la utilización de capacidades institucionales para validar, estructurar y financiar el tránsito de empresas y productos al mercado.

De manera complementaria, la vinculación entre empresa y universidad requiere una dimensión más amplia de **transformación institucional y ecosistémica**. En este punto, la *EIT Higher Education Initiative* resulta relevante porque no se limita a financiar proyectos individuales, sino que busca fortalecer la capacidad de innovación y emprendimiento de las universidades en asociación con empresas y otros actores [444]. La iniciativa reporta más de 700 instituciones involucradas, €69 millones adjudicados, más de 1,700 startups apoyadas y más de 90,000 personas formadas. Además, en 2025, lanzó una convocatoria de hasta €70 millones para proyectos que fortalezcan la cooperación entre educación superior, industria y ecosistemas de innovación. [445] Esta experiencia confirma que la relación entre universidad y empresa debe tratarse como una política de construcción de capacidades organizacionales y de gobernanza para que las universidades sean actores activos que influyen en el desarrollo empresarial.

---

[440] Fuente: Innovate UK. Knowledge Transfer Partnerships. https://iuk-business-connect.org.uk/programme/knowledge-transfer-partnerships

[441] Fuente: Innovate UK. https://iuk-business-connect.org.uk/opportunities/knowledge-transfer-partnership-ktp-2026-2027-round-1

[442] Fuente: Cambridge Enterprise. https://www.enterprise.cam.ac.uk/pioneering-research-innovations/develop-a-commercial-opportu nity/translational-funding/technology-investment-fund

[443] Fuente: Cambridge Enterprise. https://www.enterprise.cam.ac.uk/venture-building-investment

[444] Fuente: EIT Higher Education Initiative. https://eit-hei.eu/about-us

[445] Fuente: European Institute of Innovation and Technology. https://www.eit.europa.eu/news-events/news/eit-launches-eu70-million-call-boost-stem-innovation-and-strengthen-university

4. Desarrollo tecnológico con orientación comercial

El desarrollo tecnológico en la región enfrenta un problema de generación de conocimiento y conversión de dicho conocimiento en productos, servicios y empresas con capacidad de llegar al mercado. En esa línea, uno de los principales desafíos es la débil transferencia tecnológica hacia el aparato productivo, la baja contribución del sector privado a la I+D realizada por el sector académico y la limitada capacidad para financiar y escalar emprendimientos intensivos en tecnología. En este sentido, avances ocurren cuando la investigación se conecta con mecanismos de incubación, aceleración y creación de startups, y también cuando clústeres buscan avanzar hacia productos y soluciones con mayor valor agregado. Por ello, esta línea estratégica contempla que la política pública se oriente hacia resultados comercializables de IA, escalables y vinculados a demanda real.

En esta dirección, se propone **estructurar instrumentos de apoyo al desarrollo tecnológico en función de hitos de comercialización**, de manera que la política acompañe la transición desde el resultado científico hacia el desarrollo del producto y la consolidación empresarial. Este enfoque enfatiza la necesidad de resolver fallas de coordinación entre el sector académico y productivo [446], donde el desafío de empresas y clústeres pasa por agregar valor y convertir capacidades técnicas en resultados de mercado. Así, una experiencia internacional destacada es *EXIST Forschungstransfer* de Alemania, que representa un programa que apoya proyectos de emprendimiento basados en investigación con trabajo de desarrollo complejo y riesgoso. El programa se organiza en dos fases: i) los resultados de investigación con potencial de convertirse en startups se desarrollan hacia productos y procesos técnicos, junto con la elaboración del modelo de negocio; y, ii) se financia el lanzamiento de la actividad empresarial y la preparación para inversión externa [447]. La relevancia de esta experiencia para la región muestra que el apoyo público puede diseñarse alrededor de la transición hacia mercado, y no solo alrededor de la investigación o prototipo.

Asimismo, resulta importante **fortalecer incubadoras y *venture builders* especializados en tecnologías complejas**, con capacidad de financiar, aportar gestión, validación, laboratorios, y conexiones comerciales. Una referencia internacional es el *NIDHI–Technology Business Incubators* (NIDHI-TBI) de India, impulsada por el *Department of Science and Technology*, que apoya incubadoras tecnológicas ubicadas en torno a instituciones académicas y de I+D para acelerar la comercialización de tecnologías y la creación de empresas de base tecnológica. Estas incubadoras están diseñadas para promover startups basadas en conocimiento e innovación, facilitar la rápida comercialización de resultados de investigación y ofrecer servicios de valor agregado como mentoría, apoyo financiero, asistencia técnica, asesoría legal y servicios relacionados con propiedad intelectual [448].

Finalmente, otra línea de acción está vinculada con el **financiamiento de la transición desde prototipado y validación hacia la entrada al mercado** por medio de instrumentos orientados a proyectos con base científica que tienen potencial de convertirse en ofertas comercializables. En este punto, una experiencia internacional importante es *Innosuisse Start-up Innovation Projects* de Suiza, que corresponde a un programa dirigido a startups antes de su entrada al mercado y orientado a proyectos de innovación basados en investigación aplicada. Esta iniciativa señala que el objetivo es acelerar la entrada al mercado de nuevos productos y servicios, apoyando a startups en la transición desde el prototipado y la validación hacia el primer lanzamiento comercial [449]. Una vez superada esa etapa, el país cuenta con el *Swiss Accelerator*, orientado a comercializar de manera rápida y efectiva productos y servicios innovadores de startups y PyMEs ya presentes en el mercado. La utilidad de esta experiencia para la región es que permite

[446] Fuente: Katz, R. et al (2023). "Estrategia y recomendaciones de política pública para el fortalecimiento del ecosistema de ciencia, tecnología e innovación en América Latina y el Caribe: Lecciones aprendidas desde el caso argentino". CAF

[447] Fuente: Ministerio Federal de Asuntos Económicos. https://exist.de/en/programm/exist-forschungstransfer

[448] Fuente: NIDHI – Technology Business Incubators. https://nidhi.dst.gov.in/nidhitbi

[449] Fuente: Innosuisse Swiss Innovation Agency. https://www.innosuisse.admin.ch/en/start-up-innovation-projects

diferenciar entre validación pre-comercial y comercialización acelerada con el objetivo que las empresas queden atrapadas indefinidamente entre la prueba técnica y la venta efectiva.

F. Capacidades institucionales transversales habilitantes

La implementación de una política de desarrollo productivo para la inteligencia artificial en América Latina y el Caribe requiere algo más que instrumentos de financiamiento, programas de escalamiento empresarial, acceso a poder de cómputo, formación de talento o promoción de exportaciones. También exige condiciones institucionales transversales que permitan reducir la incertidumbre, generar confianza, habilitar el uso responsable de datos, coordinar actores y sostener la ejecución de políticas en el tiempo.

Estas condiciones son particularmente relevantes dado el posicionamiento heterogéneo de la región en la cadena de valor de la IA. Mientras los eslabones intensivos en capital, como semiconductores de última generación, poder de cómputo de frontera y modelos fundacionales generalistas, se encuentran altamente concentrados en líderes globales, América Latina y el Caribe presenta mayores capacidades relativas en software, herramientas, integración, implementación y desarrollo de aplicaciones orientadas a sectores productivos y sociales. Por ello, la política pública debe combinar intervenciones selectivas por eslabón con condiciones habilitantes comunes que permitan articular empresas, sectores usuarios, infraestructura, datos, regulación, financiamiento y capacidades públicas. Esta lógica es consistente con el enfoque del capítulo, que plantea seleccionar eslabones y verticales viables, escalar empresas, consolidar clústeres, resolver restricciones de financiamiento y sostener el esfuerzo con infraestructura, cómputo, talento y transferencia tecnológica.

En este marco, se identifican tres condiciones institucionales transversales, aunque no todas operan en el mismo nivel. La regulación habilitante y la gobernanza de datos son condiciones de entorno que el Estado construye hacia el ecosistema: definen reglas, reducen incertidumbre, habilitan mercados y facilitan el uso responsable de activos digitales. Las capacidades institucionales TOPP, en cambio, corresponden a capacidades internas del Estado para diseñar, implementar, coordinar, monitorear y adaptar esas políticas. En otras palabras, regulación y gobernanza de datos son ámbitos de intervención pública; TOPP es la capacidad estatal que permite que esas intervenciones sean técnicamente sólidas, operativamente viables, políticamente sostenibles y prospectivamente adaptables.

Estas tres condiciones son interdependientes y se refuerzan mutuamente. Una regulación habilitante requiere capacidades técnicas y prospectivas para comprender la evolución tecnológica y evaluar sus impactos; una gobernanza de datos efectiva requiere capacidades operativas para construir estándares, acuerdos de acceso e infraestructuras seguras; y los sandboxes regulatorios o las compras públicas innovadoras funcionan mejor cuando se apoyan en espacios de datos sectoriales que permiten probar soluciones con información pertinente, bajo reglas claras y con mecanismos de supervisión.

1. Regulación habilitante para la innovación y adopción confiable de IA

El desarrollo y la adopción productiva de la inteligencia artificial requieren marcos regulatorios que reduzcan la incertidumbre, protejan derechos, promuevan la confianza y, al mismo tiempo, habiliten la experimentación, la inversión y la innovación. La regulación no debe concebirse únicamente como un mecanismo de control o mitigación de riesgos, sino también como una herramienta de desarrollo productivo. Al establecer reglas claras, estándares, mecanismos de validación y condiciones de confianza, la regulación puede contribuir a crear mercados, reducir costos de adopción y facilitar la interacción entre proveedores tecnológicos, empresas usuarias, sector público, inversionistas y ciudadanía.

La región no parte de cero. Chile cuenta con una Política Nacional de Inteligencia Artificial, actualizada en 2024 y vigente hasta 2031, que organiza su acción en torno a factores habilitantes, desarrollo y adopción, y gobernanza y ética [450]. Brasil discute un marco legal específico sobre IA mediante el PL 2338/2023, en un contexto en el que ya cuenta con la Ley General de Protección de Datos como base relevante para la gobernanza digital [451]. Uruguay ha desarrollado una estrategia de IA para gobierno digital orientada al uso responsable de IA en la administración pública Agencia de Gobierno Electrónico y Sociedad de la Información y del Conocimiento (AGESIC), *Estrategia de Inteligencia Artificial para el Gobierno Digital*, Gobierno de Uruguay [452]. Además, varios países de la región han avanzado en mecanismos de innovación regulatoria en el sector financiero, como los esquemas de modelos novedosos o sandboxes en México y Colombia. Estas experiencias muestran que existe una base institucional parcial, pero todavía fragmentada, que debe conectarse con objetivos de desarrollo productivo.

En América Latina y el Caribe, este enfoque resulta especialmente relevante porque muchas empresas tecnológicas enfrentan barreras para validar y escalar soluciones de IA en sectores regulados de alta importancia productiva y social, como servicios financieros, salud, agroindustria, energía o administración pública. La ausencia de reglas claras sobre uso de datos, responsabilidad algorítmica, explicabilidad, seguridad, privacidad, propiedad intelectual, contratación pública, certificación y auditoría puede limitar la adopción de IA, especialmente en organizaciones de mayor escala o en sectores donde los riesgos son más sensibles.

Una regulación habilitante debería combinar, al menos, cuatro orientaciones. En primer lugar, establecer marcos proporcionales al riesgo, de modo que las aplicaciones de bajo riesgo no enfrenten cargas regulatorias excesivas, mientras que los usos de alto impacto estén sujetos a mayores exigencias de seguridad, transparencia, evaluación, trazabilidad y supervisión. El AI Act europeo constituye una referencia comparada relevante en esta materia, en la medida en que adopta un enfoque basado en riesgos y combina objetivos de seguridad, derechos fundamentales, innovación e inversión; sin embargo, su experiencia debe analizarse de manera crítica y no trasladarse mecánicamente a la región.

En segundo lugar, se requiere crear entornos controlados de experimentación, como sandboxes regulatorios, pilotos supervisados o espacios de prueba sectoriales, que permitan validar soluciones de IA en condiciones reales o casi reales antes de su despliegue masivo. En tercer lugar, es necesario desarrollar mecanismos de certificación, auditoría, evaluación y estándares técnicos que permitan a empresas regionales demostrar calidad, seguridad, interoperabilidad y confiabilidad. En cuarto lugar, las compras públicas pueden operar como instrumento habilitante si incorporan criterios de innovación, interoperabilidad, protección de datos, transparencia, confianza digital y beneficios productivos en la contratación de soluciones de IA.

Este enfoque permitiría que la regulación contribuya a crear demanda sofisticada para soluciones regionales de IA. Las compras públicas innovadoras, los sandboxes y los mecanismos de certificación pueden funcionar como primeros mercados para startups, pymes tecnológicas, integradores y empresas de software, reduciendo la brecha entre prototipo, validación y escalamiento. De esta forma, la regulación habilitante no sólo protege frente a riesgos, sino que también contribuye a formar mercados, generar confianza y acelerar la adopción productiva.

En consecuencia, esta línea de acción debería considerar: i) desarrollar marcos regulatorios proporcionales al riesgo y compatibles con la innovación; ii) crear sandboxes regulatorios y pilotos supervisados en sectores estratégicos; iii) establecer mecanismos de certificación, auditoría y

---

[450] Ministerio de Ciencia, Tecnología, Conocimiento e Innovación de Chile, *Política Nacional de Inteligencia Artificial*, actualización 2024, Gobierno de Chile.

[451] Senado Federal de Brasil, *Projeto de Lei n.º 2338, de 2023*, sobre uso de inteligencia artificial en Brasil.

[452] Agencia de Gobierno Electrónico y Sociedad de la Información y del Conocimiento (AGESIC), *Estrategia de Inteligencia Artificial para el Gobierno Digital*, Gobierno de Uruguay.

estándares para soluciones de IA; iv) incorporar criterios de innovación, interoperabilidad, protección de datos y confianza digital en compras públicas; y v) promover capacidades regulatorias especializadas en IA dentro de agencias sectoriales.

2. Gobernanza de datos y espacios de datos sectoriales

Los datos constituyen un insumo crítico para el desarrollo, adaptación y adopción de soluciones de IA. Sin embargo, su valor productivo no depende únicamente de su volumen, sino de su calidad, disponibilidad, interoperabilidad, representatividad, trazabilidad, seguridad y condiciones de acceso. En América Latina y el Caribe, uno de los principales desafíos consiste en transformar datos dispersos, fragmentados o subutilizados en activos productivos que permitan desarrollar modelos, aplicaciones y soluciones adaptadas a los contextos sectoriales, lingüísticos, regulatorios e institucionales de la región.

También en este ámbito la región cuenta con avances parciales. Uruguay se ha consolidado como referente regional en gobierno digital, interoperabilidad y gobernanza de datos, apoyado en continuidad institucional y capacidades públicas acumuladas. A nivel regional, iniciativas como ILDA han puesto énfasis en la gobernanza de datos para IA, destacando que los datos son un insumo central para fortalecer el ecosistema de gobernanza de estas tecnologías en América Latina, el Caribe y el Sur Global. Sin embargo, la agenda regional ha estado frecuentemente más asociada a datos abiertos, transparencia o gobierno digital que a la construcción de activos de datos para innovación productiva, desarrollo de modelos y especialización sectorial.

La gobernanza de datos debe entenderse como una condición habilitante para toda la cadena de valor de la IA. En los eslabones de modelos, permite entrenar, ajustar, evaluar y especializar sistemas con información pertinente. En software y herramientas, facilita el desarrollo de plataformas de analítica, automatización, búsqueda semántica, procesamiento documental y agentes inteligentes. En aplicaciones, permite construir soluciones verticales para sectores como agricultura, salud, educación, logística, minería, finanzas, biodiversidad, manufactura o gestión pública. En integración e implementación empresarial, reduce costos de adopción, mejora la interoperabilidad con sistemas existentes y permite que las soluciones de IA se inserten en procesos productivos reales.

Una política de gobernanza de datos para IA productiva debería avanzar desde una lógica centrada exclusivamente en apertura de datos hacia una lógica más amplia de espacios de datos sectoriales. Estos espacios pueden establecer reglas, estándares, mecanismos de acceso, acuerdos de uso, modelos de gobernanza y arquitecturas de interoperabilidad para compartir datos entre actores públicos, privados, académicos y sociales bajo condiciones seguras y confiables. Esto es especialmente relevante en sectores donde ningún actor individual posee todos los datos necesarios para desarrollar soluciones robustas, pero donde la colaboración puede generar valor colectivo.

La gobernanza de datos debe resolver tensiones entre apertura, soberanía, privacidad, seguridad, competencia y uso productivo. El objetivo no debería ser maximizar la circulación indiscriminada de datos, sino habilitar mecanismos diferenciados de acceso responsable —abierto, compartido, controlado o seguro— según el tipo de dato, el riesgo, el uso previsto y el interés público involucrado. Bajo esta lógica, la apertura de datos es sólo una modalidad dentro de una arquitectura más amplia de gobernanza.

En el contexto regional, los espacios de datos sectoriales pueden cumplir una función estratégica. Primero, pueden facilitar la especialización de modelos y aplicaciones en sectores donde la región posee activos diferenciales. Segundo, pueden reducir la dependencia de datos externos que no siempre reflejan adecuadamente las características lingüísticas, institucionales, culturales o productivas de América Latina y el Caribe. Tercero, pueden generar condiciones para que empresas regionales desarrollen soluciones de IA con mayor pertinencia local y potencial de escalamiento. Cuarto, pueden

contribuir a una mejor gobernanza pública, al permitir que los datos sean utilizados bajo reglas claras de seguridad, protección de derechos, trazabilidad y rendición de cuentas.

En consecuencia, esta línea de acción debería considerar: i) identificar datos estratégicos para verticales productivos prioritarios; ii) desarrollar espacios de datos sectoriales con reglas de acceso, interoperabilidad y seguridad; iii) fortalecer estándares de calidad, metadatos, trazabilidad y documentación; iv) crear mecanismos de acceso controlado a datos públicos y privados para innovación; v) promover infraestructuras seguras para entrenamiento, evaluación y validación de modelos; y vi) articular la gobernanza de datos con protección de datos personales, competencia, ciberseguridad, propiedad intelectual y regulación sectorial.

3. Capacidades institucionales TOPP para implementar una política productiva de IA

La complejidad de una política productiva de IA exige capacidades institucionales para coordinar actores, movilizar recursos, ejecutar instrumentos, aprender de la implementación y adaptar prioridades ante cambios tecnológicos acelerados. En este sentido, el enfoque TOPP de la CEPAL —capacidades técnicas, operativas, políticas y prospectivas— ofrece un marco útil para evaluar si los Estados cuentan con las condiciones necesarias para implementar una agenda de IA orientada al desarrollo productivo. La CEPAL plantea este enfoque como un marco analítico y operativo para gestionar transformaciones complejas, en el que las capacidades técnicas, operativas, políticas y prospectivas se entienden como dimensiones interrelacionadas de una arquitectura institucional dinámica.

TOPP debe entenderse como una condición de segundo orden: no es una política sectorial adicional, sino el conjunto de capacidades que permite diseñar, coordinar, ejecutar, evaluar y adaptar las políticas de regulación, datos, infraestructura, financiamiento, talento, adopción productiva y desarrollo empresarial. Su incorporación es especialmente importante en una región caracterizada por distintos niveles de madurez institucional. Países con capacidades digitales más consolidadas, como Chile o Uruguay, pueden avanzar hacia instrumentos más sofisticados de regulación, gobernanza de datos, compras públicas innovadoras o espacios sectoriales de datos; otros países podrían requerir primero fortalecer capacidades básicas de coordinación, información, interoperabilidad, talento público y ejecución programática. El diagnóstico TOPP puede servir precisamente para identificar ese punto de partida y graduar la implementación según la realidad institucional de cada país.

La capacidad técnica refiere a la posibilidad del Estado de producir, interpretar y utilizar evidencia para la toma de decisiones, y de contar con funcionarios y tomadores de politicas con conocimientos suficientes sobre IA, datos, infraestructura digital, modelos, riesgos y usos productivos. Esto implica disponer de diagnósticos actualizados sobre la cadena de valor, mapeos de empresas y capacidades, información sobre talento, disponibilidad de datos, acceso a poder de cómputo, demanda sectorial, barreras regulatorias y oportunidades de exportación. También supone formar equipos públicos capaces de dialogar técnicamente con empresas, universidades, proveedores tecnológicos, reguladores y organismos internacionales, como instituciones multilaterales y bancos de desarrollo.

La capacidad operativa remite a la posibilidad de convertir prioridades en instrumentos concretos. Esto incluye diseñar y ejecutar fondos de prueba de concepto, programas de validación tecnológica, acceso subsidiado a cómputo, clústeres, compras públicas innovadoras, certificaciones, sandboxes, programas de internacionalización, *testbeds*[453] y servicios de extensionismo tecnológico. Esta capacidad es crítica porque muchas políticas fallan no por falta de diagnóstico, sino por problemas de implementación, coordinación interinstitucional, continuidad presupuestaria o capacidades de gestión.

[453] Los testbeds son entornos controlados de experimentación y validación tecnológica que permiten probar soluciones en condiciones reales o cercanas a la realidad antes de su despliegue a mayor escala.

La capacidad política se refiere a la posibilidad de construir legitimidad, sostener acuerdos y coordinar intereses entre actores públicos, privados, académicos y sociales. Una política productiva de IA involucra ministerios de economía, ciencia y tecnología, telecomunicaciones, educación, trabajo, comercio exterior, reguladores sectoriales, bancos de desarrollo, empresas tecnológicas, sectores usuarios, universidades, centros de investigación y gobiernos subnacionales. Sin capacidades políticas, los instrumentos pueden operar de manera aislada, carecer de respaldo presupuestario o perder continuidad frente a cambios institucionales.

La capacidad prospectiva es especialmente relevante en IA por la velocidad del cambio tecnológico, la incertidumbre geopolítica y la reorganización de cadenas globales de valor. No se limita a producir escenarios, sino que permite detectar señales tempranas sobre cambios en costos de cómputo, disponibilidad de modelos abiertos, estándares técnicos, restricciones regulatorias, concentración de plataformas, reconfiguración de cadenas de suministro y nuevas oportunidades de especialización regional. Esta capacidad debe traducirse en mecanismos permanentes de vigilancia tecnológica, diálogo con actores de frontera, evaluación periódica de instrumentos y actualización de prioridades. Sin ella, la política puede quedar anclada en una fotografía estática de la cadena de valor, sobredimensionar expectativas tecnológicas o depender excesivamente de proveedores externos para definir prioridades estratégicas.

Incorporar TOPP en una estrategia productiva de IA permite reforzar que la agenda no depende sólo de definir qué hacer, sino de construir las capacidades estatales para hacerlo. En términos operativos, los países podrían utilizar TOPP como herramienta de diagnóstico institucional antes de implementar programas de IA productiva, evaluando si cuentan con evidencia suficiente, agencias ejecutoras capaces, mecanismos de coordinación, legitimidad política y capacidades prospectivas para ajustar la estrategia. También podría utilizarse como marco de seguimiento, asociando cada recomendación de política con la capacidad institucional necesaria para llevarla adelante.

En consecuencia, esta línea de acción debería considerar: i) aplicar diagnósticos TOPP a la institucionalidad responsable de IA, digitalización y desarrollo productivo; ii) fortalecer la formación de tomadores de políticas y equipos técnicos públicos en IA, datos, modelos, infraestructura, regulación y adopción productiva; iii) crear mecanismos de coordinación interministerial para IA productiva; iv) fortalecer unidades técnicas de datos, IA, evaluación y seguimiento de políticas; v) dotar a agencias ejecutoras de capacidades para operar instrumentos complejos; vi) establecer espacios permanentes de diálogo público-privado-académico; y vii) desarrollar capacidades de prospectiva tecnológica y actualización periódica de la estrategia.

G. Síntesis de recomendaciones de política

El análisis de la cadena de valor de la inteligencia artificial confirma que América Latina y el Caribe parte desde una posición rezagada y con una base productiva limitada a la de los mercados líderes en la cadena de valor de la IA. La región se encuentra mejor posicionada en los eslabones finales de la cadena (software, integración, implementación, consultoría tecnológica y aplicaciones sectoriales), mientras que los segmentos intensivos en capital, cómputo de frontera y semiconductores avanzados permanecen concentrados en Estados Unidos, China, Europa y algunos polos asiáticos. Esta situación obliga a la región abandonar estrategias de cobertura completa de la cadena y, en cambio, precisa orientar la política pública hacia espacios concretos con factibilidad de captura de valor, donde ya existe presencia empresarial, conocimiento acumulado y posibilidades reales de escalamiento. En términos de políticas de desarrollo productivo, el objetivo primordial del documento es que la región debe construir capacidades en eslabones donde presenta cierta masa crítica.

A partir de ese diagnóstico, este estudio plantea que la política debe organizarse alrededor de una secuencia coherente. En primer lugar, es importante seleccionar eslabones y verticales viables. A partir de ello, se requiere escalar empresas, consolidar clústeres y fortalecer campeones regionales. Posteriormente, se necesita resolver las restricciones de financiamiento, internacionalización y acceso a mercados y, finalmente, sostener todo el esfuerzo con infraestructura, poder de cómputo, talento y mecanismos de transferencia tecnológica. Esta continuidad es importante porque evita tratar la innovación, exportación, financiamiento y formación de talento como políticas aisladas. Por el contrario, las vincula como parte de una misma trayectoria de desarrollo empresarial en la que una empresa tecnológica necesita pasar desde la validación técnica y comercial hasta la inserción internacional, apoyada por instrumentos diferenciados según etapa, función y nivel de sofisticación.

Un hallazgo central del análisis es que la principal debilidad regional corresponde a la escasa densidad institucional y productiva para convertir capacidades dispersas en oferta exportable. La fragmentación del tejido empresarial, la baja inversión privada en I+D, la limitada escala de los mecanismos de financiamiento, la falta de articulación entre universidad y empresa, así como la concentración externa de infraestructura crítica explica el lento avance de la región en la transición desde adopción hacia producción tecnológica. En ese sentido, las recomendaciones apuntan a corregir fallos de coordinación mediante hojas de ruta por eslabón, consorcios y clústeres, inserción de las empresas en la compra pública innovadora, aplicación de mecanismos de coinversión, acceso subsidiado a cómputo, centros de transferencia tecnológica y plataformas de internacionalización.

Por otra parte, la infraestructura internacional y las plataformas globales pueden representar una oportunidad estratégica. Así, respecto de centros de datos, la región difícilmente podrá alcanzar los niveles de inversión necesarios para construir una infraestructura de cómputo regional con resultados redundantes a la ya existente. Por el contrario, la oportunidad reside en adoptar políticas para capturar derrames locales, formar proveedores, desarrollar capacidades de integración, aprovechar ecosistemas de socios globales y construir modelos soberanos adaptados a sus necesidades lingüísticas, productivas y regulatorias.

Adicionalmente, estas estrategias requieren condiciones institucionales transversales habilitantes. La regulación habilitante, la gobernanza de datos y las capacidades institucionales TOPP no constituyen un objetivo sectorial adicional, sino una capa común que permite que los instrumentos de política operen de manera coordinada, generen confianza y se adapten al cambio tecnológico. A diferencia de los ejes anteriores, estas condiciones no se dirigen a eslabones específicos de la cadena de valor, sino que atraviesan simultáneamente todos los espacios de intervención identificados. El cuadro 33 contiene un resumen de las estrategias y líneas de acción que se podrían considerar en los países de la región.

Cuadro 33
Resumen de objetivos, estrategias y líneas de acción de políticas enfocadas al desarrollo de las industrias asociadas a la inteligencia artificial

Objetivo	Estrategia	Línea de acción
Posicionamiento estratégico y especialización de empresas digitales en la cadena de valor	Posicionamiento estratégico en eslabones con empresas regionales	Creación de una estrategia de posicionamiento industrial por eslabones
		Construcción de infraestructura institucional, vinculado a empresas con exportación, incubación y desarrollo de producto
		Escalamiento e internacionalización para las empresas digitales que ya operan en los eslabones priorizados
	Especialización en verticales productivos	Elaboración de estudios técnicos que identifiquen cadenas, casos de uso, demanda empresarial, disponibilidad de datos, barreras y financiamiento
		Articulación entre financiamiento, colaboración empresarial, desarrollo de producto y comercialización en torno a problemas sectoriales específicos
		Priorizar sectores donde existan activos diferenciales
		Crear mecanismos de validación y co-desarrollo
		Convertir pilotos sectoriales en una base para la expansión comercial

Objetivo	Estrategia	Línea de acción
	Consolidación de empresas regionales operando en mercados fragmentados	Impulsar la formación de clústeres, consorcios empresariales, plataformas de innovación abierta y redes de colaboración
		Promover el uso compartido de infraestructura, equipamiento, testeo (test bed) y validación tecnológica
	Madurez de mercados específicos para ensamblaje de hardware	Consolidar funciones precisas y construir una trayectoria de sofisticación progresiva y maduración productiva a partir del ensamblaje
Escalamiento empresarial y consolidación productiva	Promoción de la inversión, innovación y exportación	Estructurar instrumentos que permitan movilizar inversión hacia empresas tecnológicas con capacidad de generar productos y servicios comercializables
		Fortalecer las capacidades de innovación y acceso a mercados internacionales
		Elevar la inversión privada en innovación empresarial mediante incentivos fiscales y financieros
		Desarrollar programas de apoyo a la innovación empresarial que combinen financiamiento, asesoría técnica, validación y acompañamiento
	Consolidación de clústeres y alianzas para potenciar el tejido empresarial	Creación y fortalecimiento de clústeres tecnológicos, consorcios sectoriales y plataformas de innovación abierta
		Establecer plataformas de articulación territorial o sectorial
		Habilitar servicios compartidos
		Promover proyectos colaborativos con empresas ancla
	Desarrollo de campeones regionales	Creación de mecanismos de apoyo a empresas tecnológicas con alto potencial de escalamiento, innovación e internacionalización
		Articulación de instrumentos de financiamiento para el crecimiento, acompañamiento estratégico, fortalecimiento de liderazgo empresarial, expansión internacional y desarrollo de capacidades de I+D aplicada
		identificar a empresas con potencial y acompañar con instrumentos diferenciados; así como, conectar con metas de internacionalización e innovación
		Establecer programas de selección y acompañamiento para firmas tecnológicas con mayor potencial de escalamiento
		Vincular empresas con potencial de escalar con programas de expansión internacional y utilizar a estas firmas como nodos del ecosistema
	Especialización funcional dentro de los eslabones priorizados	Orientar la diferenciación funcional de las empresas tecnológicas
		Establecer tipologías empresariales e identificar el tipo de función que cumple cada empresa en la cadena
		Diseñar instrumentos diferenciados según función empresarial y programas de escalamiento comercial
		Vincular la diferenciación empresarial con el acceso a procesos de compras públicas
Financiamiento, modelo de negocio y acceso a mercados	Brecha de financiamiento en desarrollo de producto	Crear fondos para pruebas de concepto y prototipos para emprendimientos tecnológicos
		Financiar la validación tecnológica en entornos de aplicación y la preparación para mercado
		Estructurar el financiamiento para el desarrollo de producto a través de procesos secuenciales
	Financiamiento para escalamiento	Crear mecanismos de coinversión público-privada para rondas de crecimiento
		Desarrollar mecanismos especializados para crecimiento y etapas tardías con el objetivo de fortalecer el mercado de capital para scale-ups
		Habilitar instrumentos de deuda de riesgo
	Estrategia exportadora de servicios tecnológicos	Financiar la preparación exportadora y exploración de mercados para empresas tecnológicas que cuentan con producto, servicio o modelo de negocio
		Creación de servicios de apoyo integral para facilitar la expansión internacional de empresas y startups a un nuevo mercado
		Fortalecimiento de la inserción internacional por medio de redes de asociación empresarial, co-desarrollo y acceso a oportunidades de negocio transfronterizas
		Construcción de una plataforma exportadora especializada para empresas de software y servicios tecnológicos
		Promover la articulación estratégica de empresas regionales con plataformas tecnológicas globales
Infraestructura, cómputo y	Aprovechamiento estratégico de	Orientar la atracción de inversión en infraestructura digital hacia la generación de derrames económicos locales

Objetivo	Estrategia	Línea de acción
dependencia estratégica	infraestructura internacional	Desarrollar programas de integración de proveedores, estándares y talento alrededor de la infraestructura internacional instalada
		Construir capas de interoperabilidad, experimentación y acceso sobre infraestructuras distribuidas
	Acceso eficiente y asequible a capacidades de cómputo	Establecer mecanismos de acceso compartido y subsidiado a capacidad de cómputo para startups y PyMEs tecnológicas
		Acompañar el acceso a cómputo con servicios de asesoría, formación e infraestructura operativa
		Articular el acceso a cómputo con alianzas público-privadas que acerquen a startups y PyMEs a infraestructura de frontera y servicios especializados
	Desarrollo de modelos soberanos	Promover la adaptación y especialización de modelos abiertos a lenguas, dominios y necesidades institucionales de la región
		Cerrar la brecha de financiamiento de los modelos soberanos mediante esquemas combinados de subsidio, créditos de cómputo, cofinanciamiento y apoyo a consorcios
		Crear entornos soberanos de despliegue para el sector público y actividades reguladas
	Alineamiento geopolítico selectivo	Adoptar una estrategia de diversificación de alianzas tecnológicas según el tipo de capacidad que se busca desarrollar
		Utilizar el alineamiento geopolítico como instrumento para ascender en la cadena de valor
		Diferenciar el alineamiento por tipo de país y por nivel de capacidad
Talento, conocimiento aplicado y desarrollo tecnológico	Formación de talento aplicado a perfiles de diseño y desarrollo	Desarrollar sistemas de anticipación y monitoreo de la demanda de perfiles de diseño y desarrollo
		Fortalecer programas de formación dual y trayectorias prácticas de transición al sector productivo
		Crear y fortalecer centros aplicados de investigación, desarrollo y transferencia vinculados a universidades, empresas y clústeres tecnológicos
	Retención de talento y masa crítica regional	Concentrar y financiar talento avanzado para investigación y desarrollo de capacidades base
		Desarrollar programas intensivos de formación aplicada para ingenieros de producto, integración y despliegue empresarial
		Crear trayectorias de carrera, redes regionales y entornos institucionales capaces de retener talento en clústeres y centros tecnológicos
	Vinculación entre empresa y universidad	Crear y fortalecer instituciones puente entre universidad y empresa
		Desarrollar programas de co-desarrollo y movilidad de conocimiento entre universidades y empresas
		Fortalecer mecanismos universitarios de comercialización, spin-offs y creación de startups
		Vinculación entre empresa y universidad a través de transformación institucional y ecosistémica
	Desarrollo tecnológico con orientación comercial	Estructurar instrumentos de apoyo al desarrollo tecnológico en función de hitos de comercialización
		Fortalecer incubadoras y *venture builders* especializados en tecnologías complejas
		Financiar la transición desde prototipado y validación hacia la entrada al mercado
Condiciones institucionales transversales habilitantes		
Condiciones institucionales transversales habilitantes	Regulación habilitante para la innovación y adopción confiable de IA	Desarrollar marcos regulatorios proporcionales al riesgo y compatibles con la innovación
		Crear sandboxes regulatorios y pilotos supervisados en sectores estratégicos
		Establecer mecanismos de certificación, auditoría y estándares técnicos para soluciones de IA
		Incorporar criterios de innovación, interoperabilidad, protección de datos y confianza digital en compras públicas
		Promover capacidades regulatorias especializadas en IA dentro de agencias sectoriales
	Gobernanza de datos y espacios de datos sectoriales	Identificar datos estratégicos para verticales productivos prioritarios
		Desarrollar espacios de datos sectoriales con reglas de acceso, interoperabilidad y seguridad
		Fortalecer estándares de calidad, metadatos, trazabilidad e interoperabilidad
		Crear mecanismos de acceso controlado a datos públicos y privados para innovación
		Articular la gobernanza de datos con protección de datos personales, competencia, ciberseguridad y propiedad intelectual

Objetivo	Estrategia	Línea de acción
	Capacidades institucionales TOPP	Aplicar diagnósticos TOPP a la institucionalidad responsable de IA, digitalización y desarrollo productivo
		Fortalecer la formación de policy makers y equipos técnicos públicos en IA, datos, modelos, infraestructura y regulación
		Crear mecanismos de coordinación interministerial para IA productiva
		Fortalecer unidades técnicas de datos, IA, evaluación y seguimiento de políticas
		Establecer espacios permanentes de diálogo público-privado-académico
		Desarrollar capacidades de prospectiva tecnológica y actualización periódica de la estrategia

Fuente: Elaboración propia.

Finalmente, la principal conclusión de política pública es que el desarrollo de una industria regional de IA dependerá de la articulación simultánea entre políticas de desarrollo productivo, innovación tecnológica, regulación habilitante, gobernanza de datos y formación de talento. La región tiene oportunidades reales en los eslabones de software, servicios tecnológicos, integración, verticales productivos, ensamblaje intermedio de hardware y especialización por funciones empresariales. Sin embargo, estas oportunidades solo podrán volverse sostenibles si los gobiernos construyen entornos donde los emprendimientos de IA puedan innovar, acceder a financiamiento, usar infraestructura, desarrollar talento, vincularse con universidades, utilizar datos bajo reglas claras, validar soluciones en marcos regulatorios habilitantes y entrar a mercados externos. En otras palabras, la política pública no debe aspirar a reproducir la frontera global, sino a construir un ecosistema regional con densidad productiva e institucional suficiente para escalar capacidades ya existentes y convertirlas en una base estable de competitividad tecnológica.

VI. Conclusiones

La irrupción de la inteligencia artificial constituye un cambio en el centro de gravedad de la organización industrial del ecosistema digital. La transición desde una estructura dominada por plataformas de agregación y OTTs hacia otra centrada en semiconductores de frontera, poder de cómputo, modelos fundacionales y herramientas de IA modifica, de forma simultánea, los nodos donde se captura valor, requisitos de inversión, cuellos de botella tecnológicos y capacidades necesarias para competir. En la cadena de la IA los fabricantes de semiconductores de frontera, los proveedores de servicios de cómputo, y aquellos que operan modelos fundacionales, se convierten en actores clave, posicionándose en nuevo eslabón.

En este contexto, la distribución de los ingresos, rentabilidad y valoración de activos en la cadena de valor de la IA revela una asimetría estructural. La captura de valor se concentra en los eslabones que controlan activos tecnológicos escasos, propiedad intelectual crítica, capacidad de cómputo, mercados de escala y poder de fijación de precios. Por ello, los segmentos iniciales —especialmente equipamiento especializado, semiconductores avanzados e infraestructura de cómputo— concentran una parte significativa de la facturación y de la rentabilidad, mientras otros eslabones enfrentan procesos de comoditización y presión sobre márgenes. Finalmente, la cadena de valor de la IA revela una disparidad profunda en la valoración de los activos físicos frente a la propiedad intelectual, donde inversionistas asignan valores superiores basados en la capacidad de diseñar GPUs y desarrollar modelos fundacionales.

El esfuerzo para desarrollar y desplegar IA representa una de las fronteras de competencia entre naciones, con implicancias geoestratégicas. Países y regiones compiten para controlar el acceso a ventajas comparativas para asegurar la preeminencia económica, política y militar derivadas de la IA en un contexto internacional fragmentado y polarizado que enfrenta Estados Unidos, la República Popular China y el continente europeo.

En este nuevo contexto, América Latina y el Caribe no parte de una posición equivalente a la de los países o regiones que lideran este mercado, aunque la región dispone de empresas, capacidades y trayectorias en los eslabones finales de la cadena. La cadena de valor de la IA en la región demuestra un

avance en los eslabones de desarrollo local de software, herramientas e implementación de sistemas, con una presencia principalmente extranjera en infraestructura, y un movimiento embrionario en modelos fundacionales soberanos.

De ello se desprende que el principal desafío regional está relacionado con la articulación entre las capacidades efectivamente existentes y los eslabones donde se concentra la captura de valor económico. En América Latina y el Caribe la cadena de valor está marcada por una asimetría determinada por la falta de acceso a activos físicos de última generación, una capacidad limitada en la manufactura de equipamiento y una amplia brecha en el desarrollo de modelos fundacionales de frontera. Al mismo tiempo, las mayores capacidades existen en el desarrollo de software y herramientas y la integracion de aplicaciones, últimos eslabones de la cadena. Por ello, el reto del desarrollo latinoamericano consiste en elaborar una estrategia regional equilibrada, que se enfoque en puntos específicos de la cadena de valor, donde se puedan mostrar las ventajas desarrolladas a través de la experiencia adquirida en eslabones finales de la cadena. De igual forma, la región posee también una base notoria en desarrollo e implementación de aplicaciones, así como en ciertos verticales donde existen datos, conocimiento operativo y demanda productiva.

Asimismo, la región debe aprovechar la tenencia de un importante porcentaje de reservas en materias primas críticas de las que depende la manufactura de semiconductores. En otras palabras, si bien no es viable una política orientada a reproducir de manera integral la cadena de valor de los países líderes, tampoco resulta suficiente una estrategia basada en adoptar herramientas controladas externamente, lo que conllevaría una salida de recursos hacia países desarrollados y un ampliado desequilibrio en la balanza de pagos tecnológica.

Con ello, se infiere que la política debe partir de la selección adecuada de eslabones y verticales viables, organizar instrumentos de medición y construir densidad productiva en conjunto con empresas que demuestren potencial para llegar a líderes regionales con una trayectoria viable de exportación de servicios.

En sentido estricto, hay que considerar que la adopción empresarial de la IA debe entenderse, en su conjunto, como una política amplia de transformación productiva que incluye aspectos de política tecnológica que son transversales. En esa línea, el diagnóstico sobre adopción muestra que su difusión en la región sigue siendo incipiente y desigual. Es decir, las empresas grandes y pertenecientes a segmentos tecnológicos han avanzado más rápido en su adopción dado que ya disponen de capacidades organizacionales, financieras y formativas superiores. Esto significa que la velocidad de incorporación de la IA depende de condiciones previas para identificar casos de uso, ordenar procesos, gestionar datos, rediseñar funciones y sostener cambios organizacionales. Por ello, se concluye que el cuello de botella regional se ubica entre la combinación de varios aspectos como: i) la medición insuficiente, ii) débil coordinación institucional, iii) escaso financiamiento; y, iv) brechas de capacidades. La consecuencia central radica en una política de adopción productiva con la finalidad de evitar que la IA profundice heterogeneidades preexistentes entre empresas, sectores y territorios, lo cual reforzaría la distancia entre firmas que ya están cerca de la frontera digital y aquellas que permanecen fuera de ella.

Dentro de esa misma lógica, el desarrollo de la industria regional de IA depende del fortalecimiento de firmas que muestran capacidades en segmentos específicos de la cadena de valor. El diagnóstico identifica una base empresarial relevante en integración, consultoría tecnológica, software, desarrollo e implementación de aplicaciones y verticales productivos con conocimiento operativo acumulado. Sin embargo, esa base todavía se encuentra fragmentada y con limitada escala regional. Por ello, la política de desarrollo productivo debe concentrarse en firmas con potencial para transformarse y capaces de expandirse, atraer talento y generar encadenamientos con startups, proveedores, universidades y centros tecnológicos. Así, el objetivo es formar un núcleo empresarial que permita la conversión de la experiencia técnica en plataformas de crecimiento e internacionalización.

Esta orientación también exige mayor claridad sobre la función de los instrumentos de política pública que deben servir para ejecutar una trayectoria de escalamiento empresarial. Este aspecto incluye financiamiento para crecimiento, apoyo a internacionalización, acceso a mercados, compras públicas, asistencia para innovación aplicada, vinculación con grandes clientes y mecanismos que faciliten alianzas entre firmas regionales. En definitiva, aparte de la selección de eslabones viables, se requiere construir condiciones para que las empresas que operan en esos espacios aumenten tamaño, complejidad tecnológica y presencia internacional.

Desde otra arista, es importante mencionar que la infraestructura internacional de cómputo debe ser aprovechada estratégicamente para generar sinergias locales, formar proveedores, desarrollar capacidades de integración, ampliar acceso a cómputo y construir modelos adaptados a necesidades lingüísticas y sectoriales. Este aspecto plantea que la política puede diseñar mecanismos intermedios de captura de beneficios, interoperabilidad, acceso compartido, formación de talento y adaptación local sobre infraestructura ya desplegada. La conclusión alrededor de la soberanía tecnológica regional no debe entenderse como reproducción integral de todos los eslabones de frontera, sino como capacidad de decisión sobre los esquemas de aprovechamiento de los activos controlados por actores externos.

En esa misma línea, la dotación regional de materias primas críticas agrega una dimensión estratégica adicional. América Latina y el Caribe concentra una proporción importante de reservas de litio, cobre, níquel, grafito y tierras raras, que representan insumos esenciales para la infraestructura asociada a la inteligencia artificial. Esa posición crea una oportunidad concreta para mejorar términos de negociación, promover refinación y procesamiento, atraer inversión en etapas intermedias y construir una diplomacia regional de materiales críticos. El principal objetivo radica en evitar que la demanda mundial de estos minerales reproduzca una especialización primaria y represente una trampa de baja productividad. En ese sentido, la región cuenta con una palanca importante para fortalecer su estrategia industrial, a través de la articulación de estrategias conjuntas alrededor de sus objetivos productivos, transferencia tecnológica y formación de capacidades locales.

Desde el desarrollo de los modelos soberanos o regionales de IA hay que reconocer que existe una brecha de financiamiento frente a los laboratorios globales que resulta demasiado amplia para generar un programa competitivo con modelos fundacionales generales. Por ello, la viabilidad regional radica en la adaptación de soluciones lingüísticas, afinamiento sectorial, especialización vertical, despliegue en sectores regulados y uso de datos locales donde existan ventajas de nicho. En esa línea, la finalidad es que los modelos diseñados regionalmente ayuden a reducir la dependencia en usos críticos, mejorar desempeño en dominios concretos y crear capacidades acumulables para empresas e instituciones públicas y que se encuentren conectados con el aparato productivo regional.

En general, el desarrollo de una posición regional competitiva en inteligencia artificial depende más de la capacidad de construir un ecosistema con institucionalidad estable, articulación productiva y gobernanza que dicte certidumbre. Así, la región requiere, entre otros, la adopción empresarial con objetivos de reconversión laboral, el diseño de una política de desarrollo de talento, y, la creación de instrumentos de adopción para la creación de oferta potencial. Por ello, la estrategia para América Latina y el Caribe depende del cumplimiento de políticas de adopción, empleo, infraestructura, financiamiento, integración empresarial, formación y transferencia tecnológica; y, que estas sean tratadas como partes de una misma arquitectura. En ese sentido, el objetivo debería apuntar hacia la construcción gradual de una base regional de competitividad tecnológica capaz de transformar adopción en aprendizaje, especialización, capacidad exportable y empleo de mayor calidad.

Bibliografía

ABES. (2024, 5 de junio). Beyond breakeven: Blip advances business strategy by generating cash. https://abes.org.br/en/para-alem-do-breakeven-blip-avanca-estrategia-de-negocio-gerando-caixa/

Abonizio, H., Sales Almeida, T., Laitz, T., Malaquias Junior, R., Kerche Bonas, G. Nogeira, R. y Pires, R. (2024). Sabia-3 Technical Report. https://arxiv.org/abs/2410.12049

Ahuja, H. (2025, 18 de noviembre). "Pinecone vs. Weaviate: The Trade-offs you only discover in production". DZone. https://dzone.com/articles/pinecone-vs-weaviate-the-trade-offs-you-only-disco

AI CertsNews. (2026). European VC 2025 Investment Trends Reach New Heights. https://www.aicerts.ai/news/european-vc-2025-investment-trends-reach-new-heights/

AI Singapore. (s. f.). AI Apprenticeship Programme. https://aiap.sg/apprenticeship

Alfaro, X. (2024, 23 de febrero). Telus inaugura Torre Emblema con una inversión de $29.2 millones. El Diario de Hoy / ElSalvador.com. https://www.elsalvador.com/h-noticias/h-negocios/telus-inaugura-torre-emblema-con-una-inversion-de-29-2-millones-dolares/1125032/2024/

Amazon Web Services. (2023, 25 de septiembre). Amazon and Anthropic announce strategic collaboration to advance generative AI. https://www.aboutamazon.com/news/company-news/amazon-aws-anthropic-ai

Amazon Web Services. (s. f.). AWS in Spain. https://aws.amazon.com/es/local/spain/

Amazon Web Services. (s. f.). Co-sell with AWS. https://aws.amazon.com/es/partners/co-sell-with-aws

Amazon Web Services. (s. f.). Partner Central. https://aws.amazon.com/es/partners/partner-central

Anthropic. (2024, 22 de noviembre). Powering the next generation of AI development with AWS. https://www.anthropic.com/news/anthropic-amazon-trainium

Asisteapp. (2025). Startups maduras y rentables: IA, regulación y capital (noviembre 19). https://asisteapp.co/startups-maduras-y-rentables-ia-regulacion-y-capital/

Aussemsn M. (2026, 10 de febrero). "European data center market reaches tipping point due to AI and energy constraints", IT Daily. https://itdaily.com/news/datacenter/european-data-center-market-reaches-tipping-point-due-to-ai-and-energy-constraints/

Bank of England. (2025, 31 de octubre). All chips in! Would a fall in AI-related asset valuations have financial stability consequence?.

Big Data Value Association. (2025). EuroHPC JU selects six new AI factories. https://bdva.eu/news/eurohpc-ju-selects-six-new-ai-factories/

BIND. (s. f.). Startup SME. https://bind.spri.eus/es/startup-sme/

Blip. (s. f.). Blip: Plataforma de comunicación automatizada. https://www.blip.ai/es/

Bloomberg Intelligence. (2025). China AI software Outlook. https://assets.bbhub.io/professional/sites/41/China-AI-Deep-Dive.pdf

British Business Bank. (s. f.). Future Fund: Breakthrough factsheet. https://www.british-business-bank.co.uk/for-finance-providers/equity-finance/future-fund-breakthrough/factsheet

British Business Bank. (s. f.). Future Fund: Breakthrough. https://www.british-business-bank.co.uk/finance-options/equity-finance/future-fund-breakthrough

Bull, A. (2025, 25 de noviembre). The rise of the AI applications Layer. Medium. https://medium.com/b8125-fall2025/the-rise-of-the-ai-application-2a17aa712e33

Business Finland. (s. f.). Market Explorer. https://www.businessfinland.fi/en/services/funding/funding-services/explorer/market-explorer

Business Model Canvas Template. (s. f.). NotCo SWOT Analysis Template Research. https://businessmodelcanvastemplate.com/products/notco-swot-analysis

Business Wire. (2025, 21 de octubre). Latin America Data Center Colocation market industry outlook 2025-2030. https://www.businesswire.com/news/home/20251021184222/en/Latin-America-Data-Center-Colocation-Market-Industry-Outlook-2025-2030-Key-Players-Such-As-Ascenty-Equinix-ODATA-Scala-Data-Centers-and-KIO-Networks-are-Leading-Large-scale-Developments---ResearchAndMarkets.com

CAF. (s. f.). Estrategia para la transformación digital de los sectores productivos en América Latina. https://scioteca.caf.com/handle/123456789/1665

Cambridge Enterprise. (s. f.). Technology Investment Fund. https://www.enterprise.cam.ac.uk/pioneering-research-innovations/develop-a-commercial-opportunity/translational-funding/technology-investment-fund

Cambridge Enterprise. (s. f.). Venture building and investment. https://www.enterprise.cam.ac.uk/venture-building-investment

Cao, S., Wang, C.Y., y Yi, X. (2025). When LLMs Go abroad: Foreign bias in AI Financial predictions. Harvard Business School Working Paper 26-013.

Carayannis, T., Kane, N., Png, M., y Nelson, A. (2025). Summary Report: Workshop on the Geopolitics of Critical Minerals and the AI Supply Chain. Science, Technology, and Social Values Lab, Institute for Advanced Study.

Castro, E. (2024, 11 de diciembre). "Ministry invests R$220 million in CEITEC for semiconductor production". Canal Solar. https://canalsolar.com.br/en/ministry-invests-in-ceitec-semiconductor-production/

Catapult Network. (s. f.). About the Catapult Network. https://catapult.org.uk

Catapult Network. (s. f.). Why the Catapult Network. https://catapult.org.uk/about-us/why-the-catapult-network

Census Bureau. (2023). Annual Business Survey 2022.

Census Bureau. Reglamento provisional para la administración de servicios de inteligencia artificial generativa. https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

Central Cyberspace Administration. https://www.cac.gov.cn/2023-07/13/c_1690898326795531.htm

Central Cyberspace Administration. Reglamento sobre la evaluación de la seguridad de los servicios de información en Internet con capacidad de movilización social o de opinión pública. https://www.cac.gov.cn/2018-11/15/c_1123716072.htm

Challapally, A., Pease, C., Raskar, R., y Chari, P. (2025). The state of AI in business 2025. MIT Project Nanda.

Chelidze Group. (2026, 12 de noviembre). Artificial intelligence market in China in 2025. https://www.chelidze.group/en/post/ai-china-2025

Coase, R. (1937). The nature of the firm, 4 Economica 386.

Colle, A., Donaldson, K., y Dragone, M. (2025) "Co-design methodology for rapid prototyping of modular robots in care settings", Front Robot AI (May 22); 12 :1581506. https://pmc.ncbi.nlm.nih.gov/articles/PMC12137090/

Comisión Económica para América Latina y el Caribe. (s. f.). Transformación digital productiva. https://repositorio.cepal.org/server/api/core/bitstreams/678e47d8-cc9a-4691-8f38-eca4b74e0bb3/content

Comisión Europea. (2024). Ley de IA. https://digital-https://strategy.ec.europa.eu/es/policies/regulatory-framework-ai
Comisión Europea. (2025). Enfoque europeo de la inteligencia artificial. https://digital-https://strategy.ec.europa.eu/es/policies/european-approach-artificial-intelligence
Comisión Europea. (s. f.). Centro de soporte de Data Spaces. https://digital-https://strategy.ec.europa.eu/en/news/data-spaces-support-centre
Comisión Europea. (s. f.). Choose Europe for Science: Over 100 national and regional initiatives aim to attract global research talent. https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/choose-europe-science-over-100-national-and-regional-initiatives-aim-attract-global-rese arch-talent-2026-01-30_mt
Comisión Europea. (s. f.). Digital innovation hubs: Helping companies across the economy make the most of digital opportunities. https://digital-strategy.ec.europa.eu/en/library/digital-innovation-hubs-helping-companies-across-economy-make-most-digital-opportunities-brochure
Comisión Europea. (s. f.). Estrategia de uso de la inteligencia artificial. https://eur-lex.europa.eu/resource.html?uri=cellar:194ae542-a421-11f0-97c8-01aa75ed71a1.0012.02/DOC_1&format=PDF
Comisión Europea. (s. f.). European Digital Innovation Hubs. https://digital-strategy.ec.europa.eu/en/policies/edihs
Comisión Europea. (s. f.). Pioneering AI project awarded for opening large language models in European languages. https://digital-strategy.ec.europa.eu/en/news/pioneering-ai-project-awarded-opening-large-language-models-european-languages
Comisión Europea. (s. f.). Simpl. https://digital-strategy.ec.europa.eu/en/policies/simpl
Comisión Europea. Estrategia Europea de Datos. https://digital-https://strategy.ec.europa.eu/en/policies/strategy-data
Comisión Europea. European Digital Innovation Hubs. https://digital-https://strategy.ec.europa.eu/en/policies/edihs
Comisión Europea. https://digital-https://strategy.ec.europa.eu/en/news/public-access-european-language-data
Comité Técnico Nacional de Normalización de la Ciberseguridad. https://www.tc260.org.cn/upload/2024-03-01/1709282398070082466.pdf
Contreras, R. (2026, 21 de enero). Blip duplicará su plantilla en España con la IA conversacional como eje estratégico. Computing. https://www.computing.es/mercado-ti/blip-duplicara-su-plantilla-en-espana-con-la-ia-conversacional-como-eje-estrategico/
Crunchbase. (s. f.). Rocketmat AI. https://www.crunchbase.com/organization/rocketmat-ai
Dassault Systèmes. (s. f.). New era of sovereign AI: Dassault Systèmes and Mistral AI deepen their partnership. https://www.3ds.com/newsroom/press-releases/new-era-sovereign-ai-dassault-systemes-and-mistral-ai-deepen-their-partnership
Data intelligence company Hyperplane - Nubank, Disponible. https://cdn.nubank.com.br/newsletter-migration/campaigns/NU_HEADLINES_EN/NuHeadlines_25_EN_preview_migrated.html
Datacenter Dynamics. https://www.datacenterdynamics.com/en/news/microsoft-files-plans-for-chilean-data-center-region/
Department of Enterprise, Tourism y Employment. (s. f.). Skills for enterprise. https://enterprise.gov.ie/en/what-we-do/workplace-and-skills/skills-for-enterprise
Department of Enterprise, Tourism y Employment. (s. f.). Together for Design. https://enterprise.gov.ie/en/publications/together-for-design.html
Department of Industry, Science y Resources. (s. f.). Broader economic benefits in ICT sector procurement. https://www.industry.gov.au/publications/broader-economic-benefits-ict-sector-procurement
Diario Financiero. https://www.df.cl/empresas/telecom-tecnologia/amazon-web-services-anuncia-inversion-de-us-4-000-millones-en-santiago
Digital Watch Observatory. (2024, 16 de septiembre). AWS announces major investment plan for Brazil. https://dig.watch/updates/aws-announces-major-investment-plan-for-brazil
Digwatch. (2025, 18 de julio). Perplexity AI overtakes ChatGPT on India's App Store. https://dig.watch/updates/perplexity-ai-overtakes-chatgpt-on-indias-app-store

Djurickovic, T. (2026, 15 de enero). "The 25 tech companies that dominated European funding in 2025", TechEli. https://tech.eu/2026/01/15/the-25-tech-companies-that-dominated-european-funding-in-2025/

Douglas, S. (2025). Advancing AI opportunities at the Telecom edge. Spirent, August 15. https://www.spirent.com/blogs/advancing-ai-opportunities-at-the-telecom-edge

Dudarenok, A. (2026, 9 de febrero). High Tech Companies Leading AI, Robotics, and EVs in. https://chozan.co/chinese-high-tech/

Economic Commission for Africa. (2024, April 15). Zambia and DRC to implement an innovative transboundary battery and electric vehicle special economic zone. https://www.uneca.org/stories/zambia-and-drc-to-implement-an-innovative-transboundary-battery-and-electric-vehicle-special

Efrati, A. (2025). Trump to let NVIDIA sell advanced AI chips to China. The Information (12/9).

EIT Higher Education Initiative. (s. f.). About us. https://eit-hei.eu/about-us

El Colombiano. https://www.elcolombiano.com/negocios/oracle-trae-region-nube-publica-a-colombia-con-100-millones-de-dolares-JC23300911#

Ela Glowicka & Jan Malek: "Digital Empires Reinforced? Generative AI value chain". Network Law Review. https://www.networklawreview.org/glowicka-malek-generative-ai/

Enki. (2026). China Mobile's AI Data Center Energy Efficiency Strategy 2026. https://enkiai.com/china-mobile-ai-energy-strategy-2026-inside-the-blueprint

Enterprise Europe Network. (s. f.). Enterprise Europe Network. https://een.ec.europa.eu

Enterprise Europe Network. (s. f.). Open invitation to join the Enterprise Europe Network as an international network partner. https://een.ec.europa.eu/news/open-invitation-join-enterprise-europe-network-international-network-partner

Enterprise Ireland. (s. f.). About Enterprise Ireland. https://www.enterprise-ireland.com/en/global/about-enterprise-ireland

Enterprise Ireland. (s. f.). Become more innovative. https://www.enterprise-ireland.com/en/supports/become-more-innovative

Enterprise Ireland. (s. f.). Market Readiness Assistance Grant. https://www.enterprisesg.gov.sg/financial-support/market-readiness-assistance-grant

Enterprise Ireland. (s. f.). Scale-Up participants increased revenue by S$2.5 billion in three years after joining the Scale-Up programme. https://www.enterprisesg.gov.sg/-/media/esg/files/media-centre/media-releases/2025/september/mr03525_scale-up-participants-increased-revenue-by-s$25-billion-in-thre e-years-after-joining-the-scale-up-programme.pdf

Enterprise Ireland. (s. f.). Scale-up. https://www.enterprisesg.gov.sg/scale-up

Enterprise Ireland. (s. f.). Singapore Global Enterprises Initiative. https://www.enterprisesg.gov.sg/grow-your-business/boost-capabilities/growth-and-transformation/singapore-global-enterprises-initiative

EPOCHAI. (2025). Can AI scaling continue through 2030?.

Escarzaga, A. (2025). "Record breakers: Europe's top 10 AI funding rounds of 2025". EU-Startups November 6). https://www.eu-startups.com/2025/11/europe-top-10-ai-funding-rounds-of-2025/

Esential News. (2022). "In the last two years INTEL invested close to US$ 1 billion in Costa Rica and officially inaugurated an Assembly and Test plant." August 9. https://www.cinde.org/en/essential-news/in-the-last-two-years-intel-invested-close-to-usd-1billion-in-costa-rica-and-officially-inaugurated-an-assembly-and-test-plant

EUR-Lex. Gobernanza de datos europea. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=legissum%3A4604122

EUR-Lex. https://eur-lex.europa.eu/EN/legal-content/glossary/data-governance.html

EuroHPC Joint Undertaking. (s. f.). AI Factories access modes. https://www.eurohpc-ju.europa.eu/ai-factories/ai-factories-access-modes_en

European Commission. (2025). A pionering AI Project awarded for opening Large Language Models to European languages I Shaping Europe's digital future. https://strategy.ec.europa.eu/en/news/pioneering-ai-project-awarded-opening-large-larnguage-models-european-languages

European Commission. (2025). AI Factories. https://digital-https://strategy.ec.europa.eu/en/policies/ai-factories

European Innovation Council. (s. f.). EIC Transition. https://eic.ec.europa.eu/eic-funding-opportunities/eic-transition_en

European Institute of Innovation y Technology. (s. f.). EIT launches €70 million call to boost STEM innovation and strengthen university cooperation. https://www.eit.europa.eu/news-events/news/eit-launches-eu70-million-call-boost-stem-innovation-and-strengthen-university

European Investment Bank. (s. f.). Sidekick Health secures EUR 35 million venture debt from EIB to accelerate R&D and global expansion. https://www.eib.org/en/press/all/2025-196-sidekick-health-secures-eur35-million-venture-debt-from-eib-to-accelerate-rd-and-global-expansion

European Investment Bank. (s. f.). Venture debt. https://www.eib.org/en/products/equity/venture-debt/index

European Investment Fund. (s. f.). GFF-EIF Growth Facility. https://www.eif.org/equity-products/all/gff-eif-growth-facility

European Research Council. (s. f.). From war damage to heart repair: ERC grantees supported to turn their science into practice. https://erc.europa.eu/news-events/news/war-damage-heart-repair-erc-grantees-supported-turn-their-science-practice

European Research Council. (s. f.). Proof of Concept Grant. https://erc.europa.eu/apply-grant/proof-concept

Falconí, F., Burbano, R., Cango, P., y Intriago, R. (2026). Extraction and dematerialization in Latin America, 1970–2023. Ecological Economics, 244, Article 108947. https://doi.org/10.1016/j.ecolecon.2026.108947

Fastweb y Vodafone. (2025). Fastweb+Vodafone and Cineca: an alliance to accelerate the adoption of Generative AI in businesses. https://www.fastwebvodafone.it/var/storage_feeds/cms20/media/fwb_wbce_cms20_fv_comun/cf1/cf1cab42c5a5841f5a7513e4262c3622/comunicato_en/2025-02-18-cs-eng-cineca.pdf

Federal Employment Agency. (s. f.). Vocational training. https://www.arbeitsagentur.de/int/en/vocational-training

Foxconn revenue hits $27 billion in November, driven by AI server growth (5 december, 2025). https://www.techinasia.com/news/foxconn-revenue-hits-27b-in-nov-driven-by-ai-server-growth

Fraunhofer. (s. f.). Range of services. https://www.fraunhofer.de/en/research/range-of-services.html

Fraunhofer. (s. f.). Technology transfer. https://www.fraunhofer.de/en/research/range-of-services/technology-transfer.html

Future Bridge Americas. (2025). The Untold Story of NVIDIA. https://future-bridge.us/the-rise-of-nvidia-how-ai-hardware-became-the-hottest-sector-in-us-semiconductors/

García Retuerta, O; Garcia-Retuerta, D. (2026) La cadena de valor de la inteligencia artificial: estrategias de autonomía para España. Documento de Opinión IEEE Enero/12. https://www.defensa.gob.es/documents/2073105/3095923/cadena_de_valor_ia_2026_dieee003.pdf/9201d73c-38c0-efa7-023b-4bf8433c0f56?t=1765972424619

Gardizy, A. (2026, 22 de febrero). "Inside OpenAI's scramble to get computing power after Stargate stalled". The Information. https://www.theinformation.com/articles/inside-openais-scramble-get-computing-power-stargate-stalled?utm_campaign=article_email&utm_content=article-16643&utm_medium=email&utm_source=sg&rc=exdbp6

Gehan, A. (2026, 8 de enero). "OpenAI's shopping ambitions hit messy reality". The Information.

Gereffi, G. y Sturgeon, T. (2013), "Global value chain-oriented industrial policy: the role of emerging economies", Global value chains in a changing world.

Ghosh, S. (2024) "India's US$ 1.25 billion push to power AI", NatureIndia (March 17). https://www.nature.com/articles/d44151-024-00035-5

Glasner, J. (2025, 9 de septiembre). "Mistral's $2B series C is Europe's largest AI round by a lot", Crunchbase News. https://news.crunchbase.com/venture/europe-largest-ai-round-mistral-seriesc/

Globant. (2025). Investors. https://investors.globant.com/The-AI-opportunity-we-dare

Globant. (s. f.). Globant named AWS Premier Tier Partner. https://www.globant.com/news/globant-aws-premier-tier-partner

Globant. (s. f.). Globant partners. https://www.globant.com/about/partners

Globant. (s. f.). Globant signs strategic partner agreement with Google Cloud. https://www.globant.com/news/globant-strategic-partner-agreement-google-cloud

Gobierno de México. https://www.proyectosmexico.gob.mx/inversion-de-5-mmdd-de-amazon-web-services-en-mexico/

Google Cloud. (2024, 5 de diciembre). Google Cloud announces 41st cloud region in Mexico. Google Cloud Blog. https://cloud.google.com/blog/products/infrastructure/google-cloud-announces-41st-cloud-region-in-mexico

Gordon, D. (2025, 27 de octubre). "Open-source vs. closed-source AI: which model should your enterprise trust", VKTR. https://www.vktr.com/ai-platforms/open-source-vs-closed-source-ai-which-model-should-your-enterprise-trust/

Grand équipement national de calcul intensif. (s. f.). Fast-track success: AI Factory France supercharges AI adoption for French startups and SMEs. https://www.genci.fr/actualites/fast-track-success-ai-factory-france-supercharges-ai-adoption-french-startups-and-smes

Grandview Research. (2025). AI Agents Market (2026-2033).

Green, T. (2025). "The Great AI subsidy squeeze". Rawveg.substack. https://dev.to/rawveg/the-great-ai-subsidy-squeeze-4pjf

Gretel. (2026). What is synthetic data creation?. https://www.gretel.ai/technical-glossary/what-is-synthetic-data-creation

Grupo SPRI. (s. f.). BIND SME. https://www.spri.eus/en/support/bind-sme

HAI. (2025). Artificial Intelligence Index Report 2025.

Halcones y Palomas. (2023, 26 de junio). Por US$1.850 millones, Patria Investimentos concretó venta, a estadounidense Aligned Data Centers, de Odata (incluye Colombia). https://www.halconesypalomas.com/2023/06/26/por-us1-850-millones-patria-investimentos-concreto-venta-a-estadounidense-aligned-data-centers-de-odata-incluye-colombia/

Hasso Plattner Institute. (s. f.). AI Service Centre. https://hpi.de/en/ai-service-centre

HM Revenue y Customs. (s. f.). Corporation Tax: Research and Development (R&D) relief. https://www.gov.uk/guidance/corporation-tax-research-and-development-rd-relief

HM Revenue y Customs. (s. f.). Tell HMRC that you're planning to claim Research and Development (R&D) tax relief. https://www.gov.uk/guidance/tell-hmrc-that-youre-planning-to-claim-research-and-development-rd-tax-relief

Ho-Chun, H., Weener, T., Chen, Y-C., Noh, S., Zha, M., y Lo, H. (2025). Do language models favor their home countries? Asymmetric propagation of positive misinformation and foreign influence audits. Harvard Kennedy School (HKS) Misinformation Review, 6(5).

Holmes, A. y Muppidi, S. (2025, 25 de junio). "OpenAI takes a page from Palantir, doubles down on consulting services". The Information.

HT Micron - STMicroelectronics. https://www.st.com/content/st_com/en/partner/partner-program/partner page/HT-Micron.html

https://news.microsoft.com/es-xl/microsoft-inaugura-su-primera-region-de-datacenters-en-chile-para-acelerar-la-innovacion-y-el-desarrollo-economico-local

Huang, K. (2025, 28 de junio). "Meta hires four more OpenAI researchers". The Information y Muppidi, S. (2026). "OpenAI reserves $50 billion for stock grants". The information (February 8).

ICEX España Exportación e Inversiones. (s. f.). Desafía. https://www.icex.es/es/servicios/consultoria-estrategica/rapido-crecimiento/desafia

ICEX España Exportación e Inversiones. (s. f.). Nueva iniciativa de colaboración transversal de ICEX, CDTI y Red.es para impulsar la internacionalización de startups tecnológicas españolas. https://www.icex.es/es/sala-prensa/notas-de-prensa/2025/02/nueva-iniciativa-de-colaboracion-transversal-de-icex-cdti-y-red-es-para-impulsar-la-internacionalizacion-de-startups-tecnologicas-espanolas

INAOE. (s. f.). Kutsari. https://www.inaoep.mx/kutsari

India Science, Technology y Innovation Portal. (s. f.). India Semiconductor Mission (ISM). https://www.indiascienceandtechnology.gov.in/st-visions/national-mission/india-semiconductor-mission-ism

IndiaAI. (s. f.). IndiaAI fellowship under the IndiaAI Mission. https://indiaai.gov.in/article/indiaai-fellowship-under-the-indiaai-mission

IndiaAI. (s. f.). IndiaAI Mission announces fellowships for PhD scholars. https://indiaai.gov.in/article/indiaai-mission-announces-fellowships-for-phd-scholars

IndiaAI. (s. f.). IndiaAI. https://indiaai.gov.in/
Infobae. (2024, 30 de agosto). Google confirmó inversión de USD 850 millones para construir en Uruguay el segundo data center de la región. https://www.infobae.com/america/america-latina/2024/08/30/google-confirmo-inversion-de-usd-850-millones-para-construir-en-uruguay-el-segundo-data-center-de-la-region/
Infobae. (2026, 12 de febrero). La acción de Globant volvió a desplomarse y cayó hasta 18%: cuáles son las variables que impactan en el unicornio argentino. https://www.infobae.com/economia/2026/02/12/la-accion-de-globant-volvio-a-desplomarse-y-cayo-hasta-18-cuales-son-las-variables-que-impactan-en-el-unicornio-argentino/
Infocomm Media Development Authority. (s. f.). IMDA accreditation. https://www.imda.gov.sg/how-we-can-help/imda-accreditation
Innosuisse Swiss Innovation Agency. (s. f.). Start-up Innovation Projects. https://www.innosuisse.admin.ch/en/start-up-innovation-projects
Innovate UK Business Connect. (s. f.). Knowledge Transfer Partnership (KTP) 2026-2027 Round 1. https://iuk-business-connect.org.uk/opportunities/knowledge-transfer-partnership-ktp-2026-2027-round-1
Innovate UK Business Connect. (s. f.). Knowledge Transfer Partnerships. https://iuk-business-connect.org.uk/programme/knowledge-transfer-partnerships
Intuition Labs. (2026, 13 de enero). Top 20 Medtech Companies Leveraging AI in 2025. https://intuitionlabs.ai/articles/top-20-medtech-companies-using-ai-2025
IOT Analytics. (2025, 4 de marzo). The leading generative AI companies. https://iot-analytics.com/leading-generative-ai-companies/
Jacamon, V., Dallard, J. Spenser, T. (2025, 16 de noviembre). Overcoming energy constraints is key to delivering on Europe's data center goals. International Energy Agency. https://www.iea.org/commentaries/overcoming-energy-constraints-is-key-to-delivering-on-europe-s-data-centre-goals
Jacobson, A. (2025). "Microsoft, OpenAI admit to core problems internally and with users", The Information.
Japan Organization for Metals y Energy Security. (2025, August). About JOGMEC. https://www.jogmec.go.jp/content/300382557.pdf
Joskow, P. "Vertical integration", Issues in Antitrust Law and Policy 273.
Kadia, H. (2025, 3 de diciembre). "AI in India: Why Airtel-Perplexity and Jio-Gemini offer free premium access". Tecknexus. https://tecknexus.com/ai-in-india-why-airtel-perplexity-and-jio-gemini-offer-free-premium-access/
Katz, R., et al, (2025). EY. Competitividad de la cadena agroindustrial argentina: transformar para crecer. Pg. 39.
Katz, R., et al,. (2023). Estrategia y recomendaciones de política pública para el fortalecimiento del ecosistema de ciencia, tecnología e innovación en América Latina y el Caribe: Lecciones aprendidas desde el caso argentin". CAF.
Kementerian Energi dan Sumber Daya Mineral Republik Indonesia. (2024, August 5). Hilirisasi nikel hasilkan nilai tambah industri baterai kendaraan listrik. https://www.esdm.go.id/id/media-center/arsip-berita/hilirisasi-nikel-hasilkan-nilai-tambah-industri-baterai-kendaraan-listrik
Khosla, P. (2024, 30 de octubre). "OpenAI to build its first AI chip with Broadcom and TSMC", AI Harware & Edge AI Summit. https://www.kisacoresearch.com/blog/openai-build-its-first-ai-chip-broadcom-and-tsmc
Korst, J., Puntoni, S. y Tambe, P. (2025). Accountable acceleration: Gen AI Fast track into the enterprise (October).
Kr-Asia. (2026). "China cloud market tops USD 13 billion". https://kr-asia.com/pulses/159929
Kruppa, M. (2025, 31 de diciembre). 2026 "Predictions: Cash-strapped Oracle issues chip-backed Debt". The Information.
La French Tech. (s. f.). Je choisis la French Tech. https://lafrenchtech.gouv.fr/en/programme/je-choisis-la-french-tech
Lagarde, C. (2025, 24 de noviembre). The transformative power of AI: Europe's moment to act. European Central Bank. https://www.ecb.europa.eu/press/key/date/2025/html/ecb.sp251124_1~c239fb4a7f.en.html
Laitz, T., Sales Almeida, T., Abonizio, H., Malaquias Junior, R., Kerche Bonás, G., Piau, M., Larcher, C., Pires, R., y Nogueira, R. (2026). Sabiá-4 Technical Report. arXiv. https://arxiv.org/abs/2603.10213

Latam GPT: Learn about the Latin American AI model developed in Chile (February 13, 2025). https://www.gob.cl/en/news/LatamGPT-learn-about-the-latin-american-ai-model-developed-in-chile/

Liu, Q. (2025, 12 de agosto). "China Demands Companies to Halt Nvidia Chip Orders Over Security Concerns". The Information.

Liu, Q. (2025, 26 de noviembre). "China is slowly but surely breaking free from NVIDIA". The Information.

Liu, Q. (2026, 13 de enero). "China restricts NVIDIA chip purchases under special circumstances". The Information.

Liu, Q. (2026, 7 de enero). "China Tells Tech Companies to halt Nvidia H200 Chip Orders". The information.

Livedocs. "GPU cloud providers in 2026". Live Docs (Nov, 24, 2025). https://livedocs.com/blog/cloud-gpu-providers-analysis

Lotus Trading Academy. (2026). Las mejores acciones de semiconductores y chips. https://expertosentrading.com/las-mejores-acciones-de-semiconductores-y-chips/

Malaysian Investment Development Authority. (s. f.). Malaysia seeks global investors to advance its semiconductor value chain at SEMICON SEA 2025. https://www.mida.gov.my/media-release/malaysia-seeks-global-investors-to-advance-its-semiconductor-value-chain-at-semicon-sea-2025

Malaysian Investment Development Authority. (s. f.). National Semiconductor Strategy to guide industry up the value chain. https://www.mida.gov.my/mida-news/national-semiconductor-strategy-to-guide-industry-up-value-chain

Malaysian Investment Development Authority. (s. f.). NSS framework passed to develop 110 local high-tech companies. https://www.mida.gov.my/mida-news/nss-framework-passed-to-develop-110-local-high-tech-companies-miti/

Malaysian Investment Development Authority. (s. f.). Setting the stage: GDS summit showcases data centre supply chain ecosystem, further strengthening Malaysia's digital infrastructure. https://www.mida.gov.my/media-release/setting-the-stage-gds-summit-showcases-data-centre-supply-chain-ecosystem-further-strengthening-malaysias-digital-infrastructure/

Mandaro, L. (2026, 12 de enero). "OpenAI agrees to buy AI healthcare App for about US$ 100 million." The Information.

Manufacturing USA. (s. f.). About Manufacturing USA. https://www.manufacturingusa.com/about-us

Markets y Markets. (2025). AI Agents Market size, share and growth and latest trends (April).

Martínez, C. (2024, 16 de mayo). Hot Sale: Mercado Libre se prepara con inteligencia artificial para la edición 2024. The Logistics World. https://thelogisticsworld.com/tecnologia/hot-sale-mercado-libre-se-prepara-con-inteligencia-artificial-para-la-edicion-2024/

Mazanti Pulse. (2026, 28 de enero). VC market update: Status of 2025 and the outlook for 2026. https://www.mazantipulse.com/insights/vc-market-update-status-of-2025-and-the-outlook-for-2026/

Mazza, M., Sachdeva, P., Aruntyunyan, S. y Alatovik, T. (2025, 19 de noviembre). The evolution of neoclouds and their next moves. McKinsey & Co.

MBN. (2025). "Foxconn invests US$ 168 million to expand serve production". Mexico Business. Disponible. https://mexicobusiness.news/cloudanddata/news/foxconn-invests-us168-million-expand-ai-server-production

McKinsey y Co. (2025, 19 de diciembre). Accelerating Europe's AI adoption: The role of sovereign AI. https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/accelerating-europes-ai-adoption-the-role-of-sovereign-ai

Mexico Daily. (2025, 27 de agosto). "Foxconn invests US$ 168 million to expand AI server production." Mexico News Daily. https://mexiconewsdaily.com/business/foxconn-investing-in-mexico/

Microsoft. (2024, 26 de septiembre). Microsoft announces 14.7 billion Reais investment over three years in Cloud and AI infrastructure and provide AI training at scale to upskill 5 million people in Brazil. https://news.microsoft.com/source/latam/company-news-pt-br/microsoft-announces-14-7-billion-reais-investment-over-three-years-in-cloud-and-ai-infrastructure-and-provide-ai-training-at-scale-to-upskill-5-million-people-in-brazil/

Miller, L. (2026, 6 de enero). "Why an NVIDIA chip could supercharge TSMC's next rally". Market Beat. https://www.marketbeat.com/originals/why-an-nvidia-chip-could-supercharge-tsmcs-next-rally/

Ministerio de Desarrollo Productivo. (s. f.). Ley de economía del conocimiento: Se establecen las actividades promovidas. https://www.argentina.gob.ar/noticias/ley-de-economia-del-conocimiento-se-establecen-las-actividades-promovidas

Ministerio de Industria, Energía y Turismo. (s. f.). Niveles de madurez de la tecnología. https://www.mintur.gob.es/publicaciones/publicacionesperiodicas/economiaindustrial/revistaeconomiaindustrial/393/notas.pdf

Ministerio Federal de Asuntos Económicos. (s. f.). EXIST Forschungstransfer. https://exist.de/en/programm/exist-forschungstransfer

Ministère de l'Économie, des Finances et de la Souveraineté industrielle et numérique. (s. f.). France and Germany join forces with Mistral AI and SAP SE to launch a sovereign AI for public administration. https://presse.economie.gouv.fr/france-and-germany-join-forces-with-mistral-ai-and-sap-se-to-launch-a-sovereign-ai-for-public-administration

Ministère de l'Économie, des Finances et de la Souveraineté industrielle et numérique. (s. f.). France and Germany join forces with Mistral AI and SAP SE to launch a sovereign AI for public administration. https://presse.economie.gouv.fr/france-and-germany-join-forces-with-mistral-ai-and-sap-se-to-launch-a-sovereign-ai-for-public-administration

Ministério da Ciência, Tecnologia e Inovação. (s. f.). Programa Nacional de Software para Exportação (SOFTEX). https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/lei-de-tics/2022-programa-nacional-de-software-para-exportacao-2013-softex

Ministério da Ciência, Tecnologia e Inovação. (s. f.). Programa Nacional de Software para Exportação (SOFTEX). https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/lei-de-tics/2022-programa-nacional-de-software-para-exportacao-2013-softex

Ministério do Planejamento e Orçamento. (s. f.). Lei de informática. https://www.gov.br/planejamento/pt-br/acesso-a-informacao/participacao-social/conselhos-e-orgaos-colegiados/cmap/politicas/area/comunicacao-ciencia-e-tecnologia/subsidios/lei-informatica

Ministério do Planejamento e Orçamento. (s. f.). Lei de informática. https://www.gov.br/planejamento/pt-br/acesso-a-informacao/participacao-social/conselhos-e-orgaos-colegiados/cmap/politicas/area/comunicacao-ciencia-e-tecnologia/subsidios/lei-informatica

Mistral AI Models 2025: European AI Excellence Guide for Developers & Researchers. https://local-ai-zone.github.io/brands/mistral-ai-european-excellence-guide-2025.html

Mistral AI nears $14B valuation in latest funding round, positioning it as Europe's AI frontrunner. https://techfundingnews.com/mistral-ai-raises-2-billion-valuation-europe-ai-frontrunner/

Mistral AI: The OpenAI competitor you need to know about", Just Think (May 24). https://www.justthink.ai/blog/mistral-ai-the-openai-competitor-you-need-to-know-about

Moloney, E. (2024, 9 de enero). "Accenture to spread GenAI across Asia Pacific and Latin America" ERP Today. https://erp.today/accenture-to-spread-genai-across-asia-pacific-and-latin-america/

Mordor Intelligence. (2025). China Data Center Server Market Size & Share Analysis - Industry Research Report. https://www.mordorintelligence.com/industry-reports/china-data-center-server-market

Mordor Intelligence. (2025). China Data Center Server Market Size & Share Analysis - Industry Research Report. https://www.mordorintelligence.com/industry-reports/china-data-center-server-market

Morgan, T.P. "The only thing worse than selling AI servers is not selling them", The Next Platform (February 28). https://www.nextplatform.com/2025/02/28/the-only-thing-worse-than-selling-ai-servers-is-not-selling-them/

Muppidi, S. (2026, 4 de enero). "OpenAI's International Conundrum", The Information.

Muppidi, S. y Palazzolo, S. (2025, 24 de diciembre). "OpenAI's ads push starts taking shape", The Information.

Mutt Data. (s. f.). Muttdata: Modern Data Platform & AI Consulting. https://www.muttdata.ai/

Nadlall V. (2025, 30 de octubre). Telecom GPU-as-a-Service: Beyond the hype, RCR Wireless News. https://www.rcrwireless.com/20251030/analyst-angle/telecom-gpu-as-a-service

Narechania, T. y Sitaraman, G. (2024). "An antimonopoly approach to governing artificial intelligence". Yale Law & Policy Review 43, Issue (Fall). https://yalelawandpolicy.org/antimonopoly-approach-governing-artificial-intelligence

Nathan, A.; Grimberg, J., Rhodes, A. (2024, 25 de junio). GenAI: too much spend, too little benefit? Goldman Sachs Research Top of Mind Issue 129. https://www.goldmansachs.com/images/migrated/insights/pages/gs-research/gen-ai--too-much-spend,-too-little-benefit-/TOM_AI%202.0_ForRedaction.pdf

National Research Council Canada. (s. f.). Financial support for technology innovation. https://nrc.canada.ca/en/support-technology-innovation/financial-support-technology-innovation

National Research Council Canada. (s. f.). NRC IRAP advisory services. https://nrc.canada.ca/en/support-technology-innovation/nrc-irap-advisory-services

Neusoft. (2025, 28 de octubre). Neusoft Leads China's Healthcare Data Intelligence Solutions Industry with Dual Market Share Leadership. https://www.neusoft.com/neusoft-leads-chinas-healthcare-data-intelligence-solutions-industry-with-dual-market-share-leadership/

NEWS. (2016, 26 de enero). Chinese AI Models Reportedly Hit ~15% Global Share in November 2025, Fueled by DeepSeek Open-Source Push. https://www.trendforce.com/news/2026/01/26/news-chinese-ai-models-reportedly-hit-15-global-share-in-nov-2025-fueled-by-deepseek-open-source-push/

Nhat, P. (2025, 25 de octubre). "Intel expands Vietnam operations as it relocates chip assembly from Costa Rica". Hanoi Times. https://hanoitimes.vn/intel-expands-vietnam-operations-as-it-relocates-chip-assembly-from-costa-rica.887172.html

NIDHI – Technology Business Incubators. (s. f.). NIDHI–Technology Business Incubators. https://nidhi.dst.gov.in/nidhitbi

NotCo. (s. f.). Conoce a Giuseppe. https://notco.com/cl/sobre/giuseppe

Novaes Magalhaes, L., y Fucuchima, L. (2025, 8 de enero). Brookfield seeks partner for Ascenty amid Brazil data center boom, sources say. Reuters. https://www.reuters.com/business/finance/brookfield-seeks-partner-ascenty-amid-brazil-data-center-boom-sources-say-2025-01-08/

NVIDIA. (2025, 26 de febrero). NVIDIA announces financial results for fourth quarter and fiscal 2025. https://nvidianews.nvidia.com/news/nvidia-announces-financial-results-for-fourth-quarter-and-fiscal-2025

NVIDIA. (2025, 28 de octubre). NVIDIA and partners build America's AI infrastructure and create blueprint to power the next industrial revolution. https://nvidianews.nvidia.com/news/nvidia-partners-ai-infrastructure-america

NVIDIA. (2026). NVIDIA Corporation Annual Review. Form 10-K.

OMDIA. (2022). AI maturity study. https://omdia.tech.informa.com/om024511/ai-market-maturity-survey-2022-ai-taking-root-with-csps

OMDIA. (2026). "Mainland China's cloud infrastructure market accelerates to..." (febrero 13). https://omdia.tech.informa.com/pr/2026/feb/mainland-chinas-cloud-infrastructure-market-accelerates-to-24percent-growth-in-q3-2025

OpenAI. (2025, 13 de octubre). OpenAI and Broadcom announce strategic collaboration to deploy 10 gigawatts of OpenAI-designed AI accelerators. https://openai.com/index/openai-and-broadcom-announce-strategic-collaboration/

Or, B. (2025, 6 de junio). "Open vs. Closed LLMs in 2025: Strategic Tradeoffs for enterprise AI". Data Science Collective. https://medium.com/data-science-collective/open-vs-closed-llms-in-2025-strategic-tradeoffs-for-enterprise-ai-668af30bffa0

Organisation for Economic Co-operation y Development. (s. f.). Promoting the development of the semiconductor ecosystem in Mexico: Understanding the policy and regulatory landscape. https://www.oecd.org/en/publications/promoting-the-development-of-the-semiconductor-ecosystem-in-mexico_02c81dec-en/full-report/understanding-the-policy-and-regulatory-landscape_01546ba3.html

Orosz, G. (2025, 12 de agosto). "What are Forward Deployed Engineers, and why are they so in demand".

Parque Científico e Tecnológico da Unicamp. (2025, 16 de abril). EPTV | Empresas vinculadas ao Parque da Unicamp aumentam 33%, e faturamento chega a R$ 129 milhões. https://parque.inova.unicamp.br/eptv-empresas-vinculadas-a-unicamp-aumentam-33-e-faturamento-chega-a-r-129-milhoes/

Pau, V. (2026, 16 de enero). "Nine tech IPO contenders to watch this year." The Information.

Peers, M. y Osawa, J. (2026, 12 de enero). "Anthropic unveils New healthcare features for Claude", The Information.

Piovesana, M. (2026, 25 de febrero). Modelo de IA de Nubank impulsa expansión del ingreso neto por intereses. Bloomberg Línea. https://www.bloomberglinea.com/negocios/modelo-de-ia-de-nubank-impulsa-expansion-del-ingreso-neto-por-intereses/

PitchBook. (s. f.). Rocketmat company profile. https://pitchbook.com/profiles/company/520489-99#research

Portal Fusões y Aquisições. (2025). Jusbrasil investe na Maritaca, de olho em modelo de linguagem brasileiro (24 de junho). https://fusoesaquisicoes.com/acontece-no-setor/jusbrasil-investe-na-maritaca/

Pradhan, M. (2025, 30 de diciembre). NVIDIA vs. TSMC: What's the better buy? The Motley Fool. https://www.fool.com/investing/2025/12/30/nvda-vs-tsm-whats-the-better-buy/

Presario Hub: Entrevista a VP de Telecomunicaciones de Oracle América Latina. https://www.prensario hub.com/oracle-abre-nueva-region-cloud-en-mexico/

Presidencia de la República de México. (s. f.). Presidenta Claudia Sheinbaum anuncia creación del Centro Nacional de Diseño de Semiconductores Kutsari. https://www.gob.mx/presidencia/prensa/presidenta-claudia-sheinbaum-anuncia-creacion-del-centro-nacional-de-diseno-de-semiconductores-kutsari

Press Information Bureau. (s. f.). Cabinet approves IndiaAI Mission. https://www.pib.gov.in/PressReleaseIframePage.aspx?PRID=2030838&reg=3&lang=2

Press Information Bureau. (s. f.). Semicon India Programme attracts global and domestic investments. https://www.pib.gov.in/PressNoteDetails.aspx?NoteId=157237&ModuleId=3&reg=3&lang=1

PROCOMER. (2023, 30 de agosto). Intel reafirma su compromiso con Costa Rica como un actor clave en la creciente demanda global de semiconductores. https://procomer.com/intel-reafirma-su-compromiso-con-costa-rica-como-un-actor-clave-en-la-creciente-demanda-global-de-semiconductores/

PYMNTS.com (2025, 3 de septiembre). Mistral Targets $14B Valuation in New Round. https://www.pymnts.com/artificial-intelligence-2/2025/mistral-ai-nears-close-of-funding-round-lifting-valuation-to-14-billion/

Qanta Says AI server plans on track. Taipei Times (August 13). https://www.taipeitimes.com/News/biz/archives/2025/08/13/2003841930

Razzaq, A. (2025, 25 de agosto). "Europe's Top AI Models of 2025: Multilingual, Open, and Enterprise-Ready", MarkTechPost. https://www.marktechpost.com/2025/08/15/europes-top-ai-models-of-2025-multilingual-open-and-enterprise-ready/

Redacción. (2024, 8 de octubre). Foxconn construirá en México la fábrica de superchips de Nvidia más grande del mundo. Mexico Industry. https://mexicoindustry.com/noticia/foxconn-construira-en-mexico-la-fabrica-de-superchips-de-nvidia-mas-grande-del-mundo

Redacción. (2026, 14 de febrero). LatamGPT, el modelo regional que busca reducir sesgos. Diario Río Negro. https://www.rionegro.com.ar/tecnologia/LatamGPT-el-modelo-regional-que-busca-reducir-sesgos/

Redação Nubank. (2024, 7 de mayo). How does Nubank use Artificial Intelligence and other technologies to improve customers' lives? Blog Nubank. https://blog.nubank.com.br/nubank-artificial-intelligence-technology/

RedUSERS. (2024, 8 de octubre). Foxconn instalará su planta más grande en México. https://www.redusers.com/noticias/foxconn-instalara-planta-mas-grande-mexico/

Rennison, J. (2025, 26 de diciembre). "As AI companies borrow billions, debt investors grow wary" New York Times.

Reuters. (2024). Nubank ultrapassa 100 milhões de clientes na América Latina. Forbes Brasil (8 de maio). https://forbes.com.br/forbes-money/2024/05/nubank-ultrapassa-100-milhoes-de-clientes-na-america-latina/

Romero, M. (2025, 1 de septiembre). Recomendaciones Inteligentes para CPG: Cómo anticiparse al tendero y vender más. Yalo. https://www.yalo.ai/blog/recomendaciones-inteligentes-para-cpg

Rua, C. (2025). "El futuro de la innovación en Europa: la respuesta al informe Draghi" PierNext, (febrero 21). https://piernext.portdebarcelona.cat/gobernanza/el-futuro-de-la-innovacion-en-europa-la-respuesta-al-informe-draghi/

S&P Global Ratings. (2023). Research Update: Sixsigma Networks Mexico S.A. de C.V. https://www.spglobal.com/ratings/en/regulatory/article/-/view/sourceId/12652684

S&P Global. (2015, 30 de mayo). "Generative AI experiences rapid adoption, but with mixed outcomes – Highlights from Vote: AI & machine learning". https://www.spglobal.com/market-intelligence/en/news-insights/research/ai-experiences-rapid-adoption-but-with-mixed-outcomes-highlights-from-vote-ai-machine-learning

Sacra. (2025). ScaleAI. https://sacra.com/c/scale-ai/

Said, S. y Guido, M. (2024). How Vodafone is using gen AI to enhance network life cycle. Google Clkoud. https://cloud.google.com/blog/topics/telecommunications/vodafone-gen-ai-enhances-network-lifecycle

Salanki, P. (2025). CoreWeave Ranks as #1 AI Cloud, Backed by SemiAnalysis's Platinum ClusterMAX™ Rating. https://www.coreweave.com/blog/coreweave-ranks-as-1-ai-cloud-backed-by-semianalysiss-platinum-clustermax-tm-rating

Scala Data Centers. (2025). Scala Data Centers closes USD 328 million international deal, securing funds for three data centers and a power substation in Chile (July). https://scaladatacenters.com/scala_na_midia/scala-data-centers-closes-usd-328m-international-deal-securing-funds-for-three-data-centers-and-a-power-substation-in-chile/

Shi, M. (2025, 21 de noviembre). "Data center pile in shows little concern about exits" PitchBook.

Shimabukuro, J. (2025, 27 de agosto). "AI data labelling and processing", Education Technology and Change. https://etcjournal.com/2025/08/27/ai-data-labeling-and-processing-update-august-2025/

Shrivastava, A., Gaffey, B., Grundin, G. Cubela, S. y Lajous, T. (2025). AI infrastructure: a new growth avenue for telco operators. Mckinsey and Co. https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/ai-infrastructure-a-new-growth-avenue-for-telco-operators

Si, M. (2026, 30 de enero). "Chinese AI chips gain huge traction in market," China Daily. https://www.chinadaily.com.cn/a/202601/30/WS697ccbf7a310d6866eb36b18.html

Si, M. (2026, 30 de enero). "Chinese AI chips gain huge traction in market," China Daily. https://www.chinadaily.com.cn/a/202601/30/WS697ccbf7a310d6866eb36b18.html

Siegel, J. (2023, 15 de noviembre). With a systems approach to chips, Microsoft aims to tailor everything "from silicon to service" to meet AI demand. https://news.microsoft.com/source/features/ai/in-house-chips-silicon-to-service-to-meet-ai-demand/

Sifted AI 100: Europe's rising AI startups in 2025. https://sifted.eu/rankings/ai-100-2025

SK telecom. (2023, 29 de septiembre). SK Telecom announces AI Pyramid strategy to become a global AI company. https://www.sktelecom.com/en/press/press_detail.do?idx=1582

Small Business Innovation Research. (s. f.). About SBIR and STTR. https://www.sbir.gov/about

Small Business Innovation Research. (s. f.). Performance benchmarks. https://www.sbir.gov/performance-benchmarks

SO Development. (2025, 10 de octubre). Top 10 Chinese Data-Collection companies. https://so-development.org/top-10-chinese-data-collection-companies-2025/

Software Technology Parks of India. (s. f.). About STPI. https://www.stpi.in/about-stpi

Software Technology Parks of India. (s. f.). Directorate statutory services. https://bengaluru.stpi.in/hi/directorate-statutory-services

Software Technology Parks of India. (s. f.). STP Scheme. https://stpi.in/index.php/en/stp-scheme

Software Technology Parks of India. (s. f.). STPI objectives. https://stpi.in/en/stpi-objectives

Stefanini Group. (2024, 31 de enero). Leadership and Innovation in the Era of Generative AI. https://stefanini.com/en/insights/news/stefanini-group-leadership-and-innovation-in-the-era-of-generative-ai

Stefanini Group. (2025, 12 de marzo). SophieX: Elevating the digital workplace with a superior employee experience. https://stefanini.com/en/insights/articles/sophiex-elevating-the-digital-workplace-with-a-superior-employee-experience

Stigler, G. (1951). "The division of labor is limited by the extent of the market", The Journal of Political Economy, Vol. 59, No. 3 (June), pp. 185-193.

Strategic Technologies for Europe Platform. (s. f.). OpenEuroLLM. https://strategic-technologies.europa.eu/step-results/step-stories/openeurollm_en

Struta, I. (2025, 17 de diciembre). Record-breaking data center investments, M&A in 2025 amid AI demand. S&P Global.

Swinhoe, D. (2015, 13 de octubre). "OpenAI plans 500 MW data center in Argentina". Data Center Dynamics. https://www.datacenterdynamics.com/en/news/openai-plans-500mw-data-center-in-argentina/

Tan, J., Jackson, B., Berjon, R. y Coyle, D. (2025). Airbus for AI: a global strategy for value creation. Bennett School for Public Policy, University of Cambridge. https://publicai.co/airbus-for-ai.pdf

Tao, Y., Viberg, O., Baker, R. y Kizilcec, R. (2024). "Cultural bias and cultural alignment of large language models". PNAS Nexus, 3, pp. 346.

Taw Y. (2025, 5 de febrero). "The Golden age of customized chips". Neuberger Berman. https://www.nb.com/en/global/insights/the-golden-age-of-customized-ai-chips

TBRC Business Research. (2026, 7 de enero). Artificial Intelligence (AI) In Chip Design Market Competition Analysis 2025: How Players Are Shaping Growth - EIN Presswire. https://www.einpresswire.com/article/880783563/artificial-intelligence-ai-in-chip-design-market-competition-analysis-2025-how-players-are-shaping-growth

TechConstant. (2026). The Efficiency Trap: China Won the Cost Curve, The US Owns The Cash Register. https://www.techconstant.com/the-efficiency-trap-china-won-the-cost-curve-the-us-owns-the-cash-register/

Techla Staff. (2024, 18 de noviembre). Blip recauda US$60 millones en ronda liderada por SoftBank y Microsoft. https://techla.pro/2024/11/18/blip-recauda-us60-millones-en-ronda-liderada-por-softbank-y-microsoft/

Technasia. (2025) Alibaba outpaces ByteDance, Tencent in China's AI cloud: report (September 10). https://www.techinasia.com/news/alibaba-outpaces-bytedance-tencent-in-chinas-ai-cloud-report

Tenembaum, M. (2025, 10 de febrero). IA y logística: reconectando el mundo. Cohen Perspectivas. https://perspectivas.cohen.com.ar/articulos/tematicas-mundo-moderno-ia-y-logistica-reconectando-el-mundo

The Associated Press. (2026, 28 de enero). "ASML made record $11.5 billion profit in 2025 thanks to AI-driven demand, plans to cut 1,700 jobs". Newsday. https://www.newsday.com/business/asml-semiconductor-profit-netherlands-jobs-artificial-intelligence-y27701

The Information Staff. (2025, 17 de septiembre). "Deep Research Report: AI 2026 Predictions". The Information.

The State Council. (2025, 4 de noviembre). China to unveil plans to integrate AI with manufacturing. https://english.www.gov.cn/news/202511/04/content_WS6909f081c6d00ca5f9a07504.html

The Tech Buzz. (2025, 1 de septiembre). Latin America builds rival to ChatGPT wih 50 billion parameter model. https://www.techbuzz.ai/articles/latin-america-builds-rival-to-chatgpt-with-50b-parameter-model

Thompson, D., Lenoir, T. y Lu, P. (2025, 2 de diciembre). Geopolitics of data centers: an AI showdown that will reshape the world. S&P Global. https://www.spglobal.com/en/research-insights/special-reports/look-forward/data-center-frontiers/geopolitics-data-sovereignty-data-center-security?utm_source=newsletter&utm_medium=link&utm_campaign=corporates&utm_content=NAMM

Tiger Brokers. (2026, 10 de febrero). Mainland China's Cloud Infrastructure Services Market Hits $13.4 Billion in Q3 2025 with 24% Yearly Growth. https://www.itiger.com/news/1121904572

Tirupattur, V. (2025, 6 de agosto). "How credit markets could finance AI's trillion-dollar gap". Morgan Stanley Thoughts on the Market. https://www.morganstanley.com/insights/podcasts/thoughts-on-the-market/credit-markets-ai-financing-gap-vishy-tirupattur-vishwas-patkar

Tobin, M. (2026, 14 de febrero). "How China built a chip industry, and why it's still not enough", The New York Times.

Tracxn. (2026). Rocketmat - 2026 Company Profile & Team. https://tracxn.com/d/companies/rocketmat/__90CDhJBbFYEuZHUPFPK44UwqtgICX0RX2zGInYwxpmo

Trendforce News. (2023, 24 de octubre). Quanta, Wiwynn, and major manufacturers scale up to meet rising demand for AI servers. https://www.trendforce.com/news/2023/10/24/insights-quanta-wiwynn-and-major-manufacturers-scale-up-to-meet-rising-demand-for-ai-servers/

U.S. Geological Survey. (2026). Lithium. In Mineral commodity summaries 2026. https://pubs.usgs.gov/periodicals/mcs2026/mcs2026-lithium.pdf

Valente, G., y Carmody, J. (2025, 12 de agosto). DigitalBridge awaits better value for data centers sale. ION Analytics / Infralogic. https://ionanalytics.com/insights/infralogic/digitalbridge-awaits-better-valuation-for-scala-data-centers-sale/

Vanian, J. y Leswing, K. (2023, 13 de marzo). "ChatGPT and generative AI are booming, but the costs can be extraordinary" CNBC. https://www.cnbc.com/2023/03/13/chatgpt-and-generative-ai-are-booming-but-at-a-very-expensive-price.html

Vaughan, Ann. (2025). "AI Infrastructure", The Information.

Vieira, S. (2026). Oracle está com cheque pronto para investir em IA no Brasil, mas há uma barreira no caminho. NeoFeed (6 de março). https://neofeed.com.br/negocios/oracle-esta-com-cheque-pronto-para-investir-em-ia-no-brasil-mas-ha-uma-barreira-no-caminho/

Wali, K. (2024). Hardware-software Co-design for power efficient Edge-AI Systems", Artificial Intelligence Machine Learning Data Science 2024 2(4), 2754-2760.

Warburg Pincus. (2022, 7 de junio). Take Blip Announces a US$70 million Series B Financing Round. https://warburgpincus.com/2022/06/07/take-blip-announces-a-us70-million-series-b-financing-round/

Waters, J. (2025, 3 de noviembre). Microsoft, OpenAI rewrite partnership rules ahead of AGI race", Pure AI. https://pureai.com/articles/2025/11/04/microsoft-openai-rewrite-partnership-rules-ahead-of-agi-race.aspx

Weltevreden, J. (2025). Uruguay AI Ecosystem mapping 2025 - Methodology and key findings.

Wheeler, K. (2025). "What's behind OpenAI's first custom chip design with TSMC," AI Magazine (February 10. https://aimagazine.com/articles/whats-behind-openais-first-custom-chip-design-with-tsmc

Wheeler, K. (2025, 8 de octubre). "Top 10: GPU platforms for Deep learning". AI Magazine. https://aimagazine.com/top10/top-10-gpu-platforms-for-deep-learning

Williamson, O. (1985). The Economic Institutions of Capitalism. New York: The Free Press, pp. 85-130.

World Bank. (s. f.). DE4LAC: Digital Economy for Latin America and the Caribbean. https://www.worldbank.org/en/programs/lac-digital/de4lac

Wu, D., L. Z. (2025) Foundation Models and AI Innovation: Evidence from the Hugging Face Platform Face Platform. ResearchGate (October). https://www.researchgate.net/publication/396561257_Foundation_Models_and_AI_Innovation_Evidence_from_the_Hugging_Face_Platform_Face_Platform

Wu, J. (2025, 9 de diciembre). China market news. https://nai500.com/blog/2025/12/9-chinese-ai-stocks-riding-the-open-source-wave/

Yates, S. (2025, 24 de septiembre). Magnum Ice Cream partners with NotCo to develop products using AI. FoodBev Media. https://www.foodbev.com/news/magnum-ice-cream-partners-with-notco-to-innovate-product-development-using-ai

Zi, Ch. (2026, 11 de febrero). Baidu integrates AI into search as rivals pursue superapp ambitions. https://amp.kr-asia.com/baidu-integrates-ai-into-search-as-rivals-pursue-superapp-ambitions

Zou, M; Zhang, Lu. (2025). Navigating China's regulatory approach to generative artificial intelligence and large language models. https://www.cambridge.org/core/journals/cambridge-forum-on-ai-law-and-governance/article/navigating-chinas-regulatory-approach-to-generative-artificial-intelligence-and-large-language-models/969B2055997BF42DE693B7A1A1B4E8BA

Anexos

Anexo A1
Ingresos de la cadena de valor de Inteligencia Artificial en Estados Unidos

Cuadro A1.1
Equipamiento: ingresos, 2025
(En millones de dólares)

Segmento	Empresa	Ingresos
Diseño de GPUs	NVIDIA	130 497[a]
	AMD	34 639[b]
Fabricación de GPUs	TSMC	122 424[c]
Ensambladores	Dell Technologies	95 567[d]
	Supermicro	21 972[e]
	Foxconn (ODM)	215 000 (Año Fiscal 2024)[f]

Fuente: Recopilación de los autores de Memorias y Balances de Empresas.
[a] 2025 NVIDIA Corporation Annual Review (Form 10-K).
[b] AMD Form 10-K (2025).
[c] Taiwan Semiconductor Manufacturing Company Limited (TSMC), *Consolidated Condensed Balance Sheet and Statements of Comprehensive Income (Audited)*, 31 de diciembre de 2025.
[d] Reporte 10-K de Dell Technologies (SEC, 2025).
[e] Super Micro Computer, Inc. Form 10-K para el año fiscal finalizado el 30 de junio de 2025 (SEC).
[f] Hon Hai Precision Industry Co., Ltd. (Foxconn), Annual Report 2024, sección "Analysis of financial data and profitability".

Cuadro A1.2
Infraestructura: ingresos, 2025
(En millones de dólares)

Empresa	Ingresos	Cuota de mercado mundial de cloud (en porcentajes)[a]
AWS	128.700[b]	10,01
Microsoft Azure	106,265[c]	8,26
Google Cloud	58,705[d]	4,57
CoreWeave (Neo-cloud)	5 131[e]	0,40
Oracle	49 230[f]	3,83

Fuente: Recopilación de los autores de memorias y balances de empresas.
[a] La cuota de mercado es calculada en relación a los ingresos mundiales de los proveedores de servicios de nube que es de US$ 1,285,788 millones (Telecom Advisory Services LLC (2025), "Economic Impact of Artificial Intelligence and cloud computing development in Asia and the Pacific").
[b] Amazon.com, Inc. 2025 Annual Report (Form 10-K). Los ingresos netos de US$ 128,725 millones para el segmento AWS representaron un crecimiento del 20% respecto al año anterior.
[c] Microsoft Corporation 2025 Annual Report (Form 10-K). Los ingresos de US$ 106,265 millones corresponden al segmento Intelligent Cloud para el año fiscal finalizado el 30 de junio de 2025.
[d] Cifra basada en el segmento reportado de Google Cloud, el cual alcanzó un ingreso operativo de US$ 13,910 millones en 2025.
[e] CoreWeave, Inc. (2026). Form 10-K: Annual Report for the fiscal year ended December 31, 2025. Los ingresos de la compañía crecieron un 168% interanual, alcanzando los US$ 5,131 millones.
[f] Oracle Corporation (2025). Form 10-K: Annual Report for the fiscal year ended May 31, 2025. $ 57,399 totales. $ 49,230 corresponden a servicios e infraestructura de cloud.

Cuadro A1.3
Preparación y gestión de datos: ingresos, 2025
(En millones de dólares)

Empresa	Ingresos	Ingresos estimados 2025
Scale AI	La compañía proyectó alcanzar los $ 2,000[a]	$ 2,000
Gretel.ai	$ 14.6[b]	$ 14.6
Pinecone (Vector DB)	$ 26.6 en el 2024. Crecimiento del 66.6% con relación al 2023[c]	$ 26.6

Fuente: Recopilación de los autores de fuentes públicas.
[a] Fuente: https://aifundingtracker.com/scale-ai-sales-led-growth-strategy/
[b] Fuente: https://growjo.com/company/Gretel.ai
[c] Fuente: https://www.bitget.com/news/detail/12560604938970

Cuadro A1.4
Modelos fundacionales: ingresos, 2025
(En millones de dólares)

Empresa	Ingresos	Ingresos estimados 2025	Cuota de mercado[a]
OpenAI	OpenAI alcanzó una tasa de ingresos anualizados de $ 25,000 a fines de febrero de 2026. Previamente, la empresa había superado los $ 20,000 en ingresos durante el año 2025[b]	20 000	4,65%
Anthropic	$ 9,000[c] A fines de febrero, la empresa reporto ingresos anualizados de $ 19,000[d]	9 000	2,09%
Google (DeepMind)	No Disponible / Sin discriminación por segmento	-	Sin Información
xAI	$107 en el tercer trimestre de 2025[e] Ingresos anualizados de $428	428	0,10%
Meta	Sin ingresos directos por su modelo Llama. Pero con beneficios indirectos en la mejora de rendimientos de los anuncios	-	0,00%

Fuente: Recopilación de los autores de fuentes públicas.
[a] La cuota de mercado es calculada en relación a los ingresos mundiales generados por la IA de $429,830 millones (Fuente: IDC)
[b] Fuente: https://seekingalpha.com/news/4560518-anthropic-approaches-20b-revenue-run-rate-amid-pentagon-clash-over-ai-use
[c] Fuente: https://winbuzzer.com/2026/03/06/openai-hits-25-billion-revenue-anthropic-closes-gap-xcxwbn/ y https://www.tradingkey.com/analysis/stocks/us-stocks/261490879-openai-revenue-advertisement-tradingkey
[d] Fuente: https://seekingalpha.com/news/4560518-anthropic-approaches-20b-revenue-run-rate-amid-pentagon-clash-over-ai-use. Muppiddi S. and Pau, V. (2026). "Anthropic CEO credits "mission" for staff retention after startup hits $ 19 billion in annualized revenue". *The Information* (March 3). Esos valores surgen de anualizar los ingresos extraordinarios de Febrero 2026, lo cual no necesariamente es extrapolable a nivel anual.
[e] https://www.techdogs.com/tech-news/td-newsdesk/elon-musks-xai-reports-146-billion-loss-as-startup-rapidly-expands-operations

Cuadro A1.5
Plataformas de software y herramientas: ingresos, 2025
(En millones de dólares)

Empresa	Ingresos
Hugging Face	130[a]
Databricks	5 400[b]
LangChain	12 a 16[c]
Suno	300[d]

Fuente: Recopilación de los autores de fuentes públicas.
[a] Fuente: https://sacra.com/c/hugging-face/ y https://dev.to/alextroitskyai/the-11-largest-ai-companies-in-the-world-december-2025-edition-3de4 . Tras alcanzar una tasa de ingresos de US$ 70 millones a finales de 2023, las proyecciones para 2025 sitúan su facturación en torno a los US$ 130 millones gracias a sus contratos de servicios gestionados con gigantes como NVIDIA y Amazon.
[b] Fuente: https://www.databricks.com/company/newsroom/press-releases/databricks-grows-65-yoy-surpasses-5-4-billion-revenue-run-rate
[c] Fuente: https://www.techshotsapp.com/business/langchain-nears-unicorn-valuation-with-1b-ivpled-funding-round
[d] Fuente: https://www.musicinafrica.net/magazine/suno-reports-two-million-paid-subscribers-300m-annual-revenue

Cuadro A1.6
Desarrollo e implementación de aplicaciones: ingresos, 2025
(En millones de dólares)

Empresa	Ingresos
Harvey	190[a]
Palantir	4 475[b]
Abridge	100[c]
Accenture	5 900[d]

Fuente: Recopilación de los autores de fuentes públicas
[a] El CEO Winston Weinberg confirmó que la empresa alcanzó una tasa de ingresos anualizados de US$ 190 millones al cierre de 2025 (https://www.forbes.com/sites/iainmartin/2026/02/09/legal-ai-startup-harvey-in-talks-to-raise-200-million-at-11-billion-valuation/).
[b] Fuente: https://investors.palantir.com/news-details/2026/Palantir-Reports-Q4-2025-U-S--Comm-Revenue-Growth-of-137-YY-and-Revenue-Growth-of-70-YY-Issues-FY-2026-Revenue-Guidance-of-61-YY-and-U-S--Comm-Revenue-Guidance-of-115-YY-Crushing-Consensus-Expectations/
[c] Fuente: https://sacra.com/c/abridge/ .
[d] Accenture Reports Fourth-Quarter and Full-Year Fiscal 2025 Results. $ 69,673 (total de la empresa). $5,900 de modo directo por IA.

Anexo A2
Costos operativos de la cadena de valor de Inteligencia Artificial en Estados Unidos

Cuadro A2.1
Equipamiento: costos operativos, 2025
(En millones de dólares s)

Segmento	Empresa	Estructura de costos	OPEX	OPEX/Ingresos (en porcentajes)
Diseño de GPUs	NVIDIA	Costos directos atados a la subcontratación de la manufactura avanzada. Al delegar la producción física, su estructura operativa se concentra masivamente en I&D de nuevas arquitecturas de frontera, compensación en acciones para retener talento especializado y gastos de ventas y administración	16 405[a]	12,57
	AMD	Costos generados por la subcontratación de la manufactura avanzada. Opera con un modelo *fabless* análogo al de NVIDIA, enfocando sus gastos operativos recurrentes en I&D para su línea Instinct (GPUs) y EPYC (CPUs), además de agresivos esfuerzos de ventas para capturar cuota de mercado	13 458[b]	38,85
Fabricación de GPUs	TSMC	Costos operativos intensivos en el consumo masivo de materiales, investigación de ciencia de materiales para bajar de los 3nm, y el mantenimiento continuo de las instalaciones de manufactura de frontera	11 109[c]	9,07
Ensambladores	Dell Technologies	Adquisición de GPUs de terceros, gastos de logística global, desarrollo de soluciones de optimización térmica y ensamblaje físico continuo de los racks	15 013[d]	15,71
	Supermicro	Adquisición directa de GPUs y componentes de terceros, gastos de logística, optimización térmica a nivel de rack (Direct Liquid Cooling) y ensamblaje	1 177[e]	5,36
	Foxconn (ODM)	Operaciones logísticas de ensamblaje a gran escala y pago intensivo de mano de obra manufacturera (633,167 empleados). La estructura se enfoca en la automatización de procesos ("Lights-out factories") para mitigar el alza de costos	7 043 (Año Fiscal 2024)[f]	3,28

Fuente: Recopilación de los autores de fuentes públicas

[a] Fuente: 2025 NVIDIA Corporation Annual Review (Form 10-K). El 78.7% de sus gastos operativos totales (US$ 16,405 millones) se destina puramente a I&D (US$ 12,914 millones).

[b] Fuente: AMD Form 10-K (2025). Corresponde al total de Operating Expenses, compuesto principalmente por Investigación y Desarrollo (US$ 6,384 M), Gastos de Ventas, Generales y Administrativos (US$ 3,380 M) y la amortización de activos intangibles (US$ 3.694 M).

[c] Taiwan Semiconductor Manufacturing Co., Ltd. (TSMC), *4Q25 Management Report - Investors*, 15 de enero de 2026. La cifra de gastos operativos consolida las partidas de Investigación y Desarrollo, Gastos Generales y Administrativos y Marketing, calculada sobre el tipo de cambio promedio del ejercicio fiscal reportado por la compañía. Así también la empresa tiene costo de ventas por US$ 49,104 millones.

[d] Fuente: *Consolidated Statements of Income* del Formulario 10-K de Dell Technologies para el Año Fiscal 2025 (SEC). Los gastos operativos totales de US$ 15,013 millones se componen de US$ 11,952 millones en rubros de comercialización y administración, y US$ 3,061 millones destinados a Investigación y Desarrollo. Así también la empresa tiene costos de ventas por US$ 74,317 millones.

[e] Cifra obtenida de la línea Total operating expenses del Formulario 10-K (2025). Estos US$ 1,177 M se componen de US$ 637 M en Investigación y Desarrollo y US$ 540 M en Gastos de Ventas, Generales y Administrativos. Así también la empresa tiene costo de ventas por US$ 19,542 millones.

[F] Hon Hai Precision Industry Co., Ltd. (Foxconn), Annual Report 2024, sección "Comparative analysis of financial performance", pág. 135. Los gastos operativos totales para 2024 fueron de NT$ 228,338,691 (en miles de NT$). Dentro de esta cifra, los gastos de I+D reportados en la pág. 105 fueron de NT$ 115,771,718 (en miles de NT$).

Cuadro A2.2
Infraestructura: costos operativos, 2025
(En millones de dólares)

Empresa	Estructura de costos	OPEX	OPEX/Ingresos
AWS	Consumo eléctrico masivo, sistemas de enfriamiento avanzado, mantenimiento de infraestructura de IA y costos de personal técnico especializado.	83 119[a]	64,58%
Microsoft Azure	Consumo eléctrico masivo, enfriamiento, sumado a los gastos de licenciamiento, I&D en ingeniería de IA y computación en la nube para integrar OpenAI en sus sistemas.	61 676[b]	58,04%
Google Cloud	Incluye la compensación de empleados, así como costos asignados de infraestructura técnica (energía, equipos y capacidad de red) y servicios de terceros.	44 795[c]	76,31%
CoreWeave (Neo-cloud)	Al ser una infraestructura pura de IA, sus costos operativos recaen principalmente en gastos energéticos intensivos para enfriamiento líquido.	3 724[d]	72,58%
Oracle	Los costos operativos están impulsados por mayores gastos en infraestructura física, consumo de energía y personal técnico especializado.	39 721[e]	80,68%

Fuente: Recopilación de los autores de fuentes públicas.
[a] Amazon.com, Inc. 2025 Annual Report (Form 10-K). La cifra de US$ 83,119 millones corresponde a los Gastos Operativos reportados para el segmento AWS en la Nota 10 — Segment Information.
[b] Microsoft Corporation 2025 Annual Report (Form 10-K). La cifra de US$ 21,505 millones corresponde al segmento de Intelligent Cloud de OPEX. Así también para esta empresa se incluyó un Costo de Ventas por US$ 40,171 millones.
[c] Alphabet, Annual Report 2025 (Form 10-K). La cifra de US$ 44,795 millones refleja exclusivamente los gastos operativos atribuidos al segmento de Google Cloud, que incluye tanto costos directos como la porción de infraestructura técnica compartida que se le asigna por uso.
[d] CoreWeave, Inc. (2026). Form 10-K: Annual Report for the fiscal year ended December 31, 2025. La estructura de OPEX de US$ 3,724 millones se desglosa principalmente en Tecnología e Infraestructura (US$ 2,929 millones), Ventas y Marketing (US$ 144 millones) y Gastos Generales y Administrativos (US$ 651 millones). Así también existen Costo por ventas de US$ 1,453 millones no contabilizados en el OPEX.
[e] Oracle Corporation (2025). Form 10-K: Annual Report for the fiscal year ended May 31, 2025. La compañía reportó un OPEX total de US$ 39,721 millones. Dentro de este monto, los costos directos de "Servicios en la nube y soporte de licencias" ascendieron a US$ 11,569 millones, seguido por investigación y desarrollo (US$ 9,860 millones) y ventas y marketing (US$ 8,651 millones). La empresa incluye dentro del OPEX amortizaciones de intangibles por US$ 2,307 millones.

Cuadro A2.3
Preparación y gestión de datos: costos operativos, 2025
(En millones de dólares)

Empresa	Estructura de costos	OPEX
Scale AI	Fuerza de trabajo elástica: Su mayor costo operativo no es el cómputo, sino el pago a su red de más de 700,000 colaboradores y expertos. Se reporta que han pagado "cientos de millones de dólares" a estos especialistas en el último año.[a] Contratos de Defensa: Parte de su OPEX se destina a mantener infraestructuras de seguridad de nivel militar para contratos como el proyecto *Thunderforge* del Departamento de Defensa.[b]	Sin Información
Gretel.ai	Se concentra principalmente en talento humano (plantilla de aproximadamente 80 empleados), infraestructura de nube e investigación y desarrollo para modelos de datos sintéticos.[c]	Absorbida por NVIDIA a principios de 2025 Sin Información específica sobre el OPEX
Pinecone (Vector DB)	Se basa en una nómina de talento especializado (aproximadamente 129 empleados a febrero de 2026) y el arrendamiento de infraestructura en la nube para sostener su arquitectura serverless.[d]	Sin Información

Fuente: Recopilación de los autores de fuentes públicas.
[a] Fuente: https://outlier.ai/experts/philosophy
[b] Fuente: https://scale.com/blog/thunderforge-ai-for-american-defense
[c] Fuente: https://www.crn.com/news/ai/2025/nvidia-buys-synthetic-data-startup-gretel-to-boost-ai-llms-report
[d] Fuente: https://tracxn.com/d/companies/pineconesystems/__oUwanTwdq5VEQ_i9ABozYwVOolyHuunHPhQ2BgGYPG4#about-the-company

Cuadro A2.4
Modelos fundacionales: costos operativos, 2025

Empresa	Estructura de costos	OPEX	OPEX/Ingresos
OpenAI	Alquiler masivo de capacidad de cómputo en la nube para la fase de entrenamiento y, fundamentalmente, para sostener la inferencia diaria de ChatGPT y sus APIs.	La compañía ha elevado drásticamente sus previsiones, proyectando US$ 115.000 millones entre 2025 y 2029[a] Sin valor específico para el 2025, se estima en US$ 20,000 millones[b]	100,00%
Anthropic	Al igual que su principal rival, su OPEX está abrumado por los inmensos pagos recurrentes destinados al alquiler de poder de cómputo masivo para el entrenamiento y ejecución de su familia de modelos Claude, sumado a los paquetes de compensación necesarios para atraer a investigadores.	2025: C*ash Burn* de US$ 5,200 millones[c] Sin información específica sobre el OPEX, pero lo estimamos en US$ 14,200 millones[d.] 2026: Gasto proyectado de US$ 7,000 millones únicamente para la ejecución (*running*) de modelos para usuarios finales[e]	157,78%
Google (DeepMind)	Investigación y desarrollo avanzado de IA centralizado a nivel de Alphabet, enfocado en modelos de frontera. Incluye compensación de empleados y costos de uso de infraestructura técnica para el desarrollo de modelos generales de IA.	US$ 16,760 millones (Reportado como "Alphabet-level activities", Año Fiscal 2025)[f]	Negativo Sin ingresos específicos de Deep Mind
xAI	Inversión masiva en chips y electricidad para entrenamiento, gestión de centros de datos (Colossus) y reclutamiento de talento de investigación de IA.	Sin discriminación del OPEX	Sin datos para realizar una estimación

Fuente: Recopilación de los autores de fuentes públicas.
[a] Fuente: https://mlq.ai/news/openai-raises-projected-cash-burn-to-115-billion-through-2029-far-exceeding-prior-estimates/
[b] Estimación de los autores considerando ingresos por US$20,000 millones, CAPEX por US$8,000 millones y cash burn de US$8,000.
[c] Fuente: https://www.theinformation.com/articles/anthropic-lowers-profit-margin-projection-revenue-skyrockets
[d] Estimación de los autores sumando los ingresos de US$ 9,000 millones a la quema de efectivo de US$ 5,200 millones para obtener exclusivamente el flujo de gasto operativo puro.
[e] Fuente: https://es.tradingview.com/news/gurufocus:892a33de7094b:0-anthropic-quietly-raises-the-ceiling/
[f] Alphabet, Annual Report 2025 (Form 10-K). Google DeepMind y el desarrollo de modelos de frontera se reportan contablemente bajo "Alphabet-level activities". Los US$ 16,760 millones representan la pérdida operativa de estas actividades, lo cual constituye el costo neto de investigación y desarrollo de IA que Alphabet no asigna a sus segmentos comerciales (Services o Cloud) para reflejar su naturaleza de investigación compartida.

Cuadro A2.5
Plataformas de software y herramientas: costos operativos, 2025

Empresa	Estructura de costos	OPEX	OPEX/Ingresos
Hugging Face	Dominada por costos de infraestructura en la nube para mantener el repositorio de modelos, subsidio de la capa gratuita y nómina de ingenieros especializados.	US$ 180 millones - US$ 260 millones[a]	169,23%
Databricks	Pago por infraestructura a hiperescaladores, una fuerza de ventas corporativa global masiva e inversión en I+D para optimizar motores de bases de datos y asistentes como *Genie*.	US$ 4,200 millones – US$ 4,800 millones[b]	83,33%
LangChain	Enfocada estrictamente en la captación y retención de talento de ingeniería de primer nivel para desarrollar productos como LangSmith y en marketing para evangelizar la plataforma B2B.	US$ 35 millones - US$ 55 millones[c]	321,43%
Suno	Arrendamiento intensivo de granjas de GPUs para entrenamiento y altísima demanda de inferencia (7M de canciones diarias), sumado a reservas para gastos legales por litigios de derechos de autor.	US$ 280 millones - US$ 380 millones[d]	110,00%

Fuente: Recopilación de los autores de fuentes públicas.
[a] Fuente: https://observer.com/2026/02/hugging-face-monetization-chief-jeff-boudier-business-model/ . Basado en un ingreso estimado superior a los US$ 100M y el hecho de que mantienen "dinero en el banco" de su ronda de US$ 235M, lo que sugiere un gasto operativo controlado para su escala.
[b] Fuente: https://www.databricks.com/company/newsroom/press-releases/databricks-surpasses-4b-revenue-run-rate-exceeding-1b-ai-revenue. La empresa reportó un ingreso de US$ 5,400 millones a principios de 2026 y alcanzó un flujo de caja libre positivo, lo que sitúa su OPEX anual ligeramente por debajo de sus ingresos por suscripción.
[c] Fuente: https://blog.langchain.com/series-b/. Con un ingreso de aproximadamente US$ 16M a mediados de 2025 y una inyección de US$ 125M en su Serie B, su "burn rate" operativo se estima para escalar su infraestructura de software sin activos físicos.
[c] Fuente: https://suno.com/blog/series-c-announcement. Su OPEX es excepcionalmente alto para su tamaño debido a que la generación de audio es más costosa en cómputo que el texto; sus ingresos de US$ 300M se ven presionados por los costos de inferencia y litigios.

Cuadro A2.6
Desarrollo e implementación de aplicaciones: costos operativos, 2025

Empresa	Estructura de costos	OPEX
Harvey	Pago constante de tokens a modelos fundacionales, reclutamiento de ingenieros de IA de élite y abogados expertos para validación (RLHF legal).	Estimación de entre US$ 230 millones a US$ 290 millones[a]
Palantir	Salarios de ingenieros de despliegue avanzado, actividades de ventas y marketing, y desarrollo de la plataforma.	US$ 2,272 millones[b]
Abridge	Procesamiento de audio médico en tiempo real, talento clínico para validación de datos e inversiones en ciberseguridad y cumplimiento normativo.	Entre US$ 180 millones a US$ 240 millones[c]
Accenture	Pago de nómina de cientos de miles de consultores, inversiones masivas en capacitación interna (re-skilling).	Sin información específica para el segmento de IA[d]

Fuente: Recopilación de los autores de fuentes públicas.

[a] Estimación basada en ingresos por US$ 190 millones (https://www.forbes.com/sites/iainmartin/2026/02/09/legal-ai-startup-harvey-in-talks-to-raise-200-million-at-11-billion-valuation/) y en que la empresa tuvo levantamiento de capital por US$ 300 millones en junio de 2025 (Serie E) para expansión intensiva.

[b] Cifra extraída del Informe Anual (Form 10-K) para el año fiscal finalizado el 31 de diciembre de 2025. El total de gastos operativos se desglosa en: Ventas y marketing (US$ 1,056.9 millones), Investigación y desarrollo (US$ 557.7 millones) y Generales y administrativos (US$ 657.7 millones). Así también tuvo costo de ventas por 789 millones.

[c] Estimación basada en haber alcanzado una tasa de ingresos anualizados (ARR) de US$ 100 millones a mediados de 2025 (https://sacra.com/c/abridge/) y en su fase de expansión intensiva tras captar US$ 300 millones en su Serie E de junio de 2025 para integrarse en grandes sistemas de salud.

[d] Dato obtenido del Informe Anual 2025 (año fiscal finalizado el 31 de agosto). Representa el total de gastos operativos (US$ 59.4B), que incluye el costo de los servicios (US$ 47,438 millones), gastos de ventas, generales y administrativos (US$ 11,394 millones) y gastos de optimización de costos (US$ 615 millones).

Anexo A3
Rentabilidad de la cadena de valor de Inteligencia Artificial en Estados Unidos

Cuadro A3.1
Infraestructura: rentabilidad, 2025

Empresa	Margen EBITDA
AWS	52,09%[a]
Microsoft Azure	41,96%[b]
Google Cloud	23,69%[c] (Margen Operativo, incluye amortizaciones)
CoreWeave (Neo-cloud)	46,93%[d]
Oracle	41,55%,[e] Considera la empresa en su conjunto

Fuente: Recopilación de los autores de fuentes públicas.
Nota: La rentabilidad de estas empresas es más baja que la reflejada en el Margen EBITDA, ya que tienen un costo importante de amortización de hardware que se deprecia en un corto plazo.

[a] Amazon.com, Inc. 2025 Annual Report (Form 10-K). Los ingresos netos de US$ 128,725 millones para el segmento AWS representaron un crecimiento del 20% respecto al año anterior. El Margen EBITDA (52.09%) se calculó sumando el Ingreso Operativo del segmento (US$ 45,606 millones) y los cargos por Depreciación y Amortización específicos de AWS (US$ 21,450 millones), según se desglosa en la Nota 10 — Segment Information.

[b] Microsoft Corporation 2025 Annual Report (Form 10-K). Los ingresos de US$ 106,265 millones corresponden al segmento Intelligent Cloud para el año fiscal finalizado el 30 de junio de 2025. El Margen Operativo (41.96%) se calculó dividiendo el Ingreso Operativo del segmento (US$ 44,589 millones) por sus ingresos netos. Es importante señalar que Microsoft no asigna activos, depreciación ni amortización de forma individualizada a sus segmentos en el reporte 10-K, por lo que se reporta el margen operativo como la métrica de rentabilidad segmentada más precisa disponible.

[c] Cifra basada en el segmento reportado de Google Cloud, el cual alcanzó un ingreso operativo de US$ 13,910 millones en 2025. El reporte 10-K no desglosa la depreciación específica por segmento para calcular un margen EBITDA exacto, por lo que se utiliza el margen operativo oficial del segmento (23,69%) como métrica de rentabilidad.

[d] CoreWeave, Inc. (2026). Form 10-K: Annual Report for the fiscal year ended December 31, 2025. Los ingresos de la compañía crecieron un 168% interanual, alcanzando los US$ 5,131 millones. El margen EBITDA del 46.93% se calculó a partir de una pérdida operativa de US$ 46 millones, ajustada por US$ 2,454 millones en gastos de depreciación y amortización reportados en el estado de flujos de efectivo.

[e] Oracle Corporation (2025). Form 10-K: Annual Report for the fiscal year ended May 31, 2025. Los ingresos y la rentabilidad reflejan los resultados consolidados de la compañía para el ejercicio fiscal 2025. El margen EBITDA del 41.55% se calculó a partir de una utilidad operativa de US$ 17,678 millones, ajustada por US$ 3,867 millones en gastos de depreciación y US$ 2,307 millones en amortización de activos intangibles. Esta métrica global se utiliza debido a que la estructura de costos operativos y de soporte se publica de manera centralizada a nivel corporativo.

Cuadro A3.2
Preparación y gestión de datos: rentabilidad, 2025

Empresa	Margen EBITDA
Scale AI	50% (2024)[a]
Gretel.ai	Sin Información
Pinecone (Vector DB)	Sin Información

Fuente: Recopilación de los autores de fuentes públicas.
[a] Fuente: https://thedigipalms.medium.com/scale-ai-story-from-garage-startup-to-14b-empire-1e84aee1e911

Cuadro A3.3
Modelos fundacionales: rentabilidad, 2025

Empresa	Margen EBITDA
OpenAI	La empresa sigue operando con déficit. La compañía no espera alcanzar el punto de equilibrio y la rentabilidad hasta el año 2030[a]
Anthropic	Anthropic no es rentable en la actualidad, ya que los inmensos costos de la infraestructura de frontera superan las ganancias generadas por los ingresos. Para el 2027 la empresa espera acercarse al equilibrio con una quema de efectivo del 9% de los ingresos[b] y ser positiva al 2028[c]
Google (DeepMind)	Pérdida de US$ 16.760 millones[d]
xAI	Pérdida de US$ 7,800 millones en los primeros 9 meses del 2025[e]
Meta	Negativo. No es un modelo que genere monetización directa

Fuente: Recopilación de los autores de fuentes públicas.
[a] https://winbuzzer.com/2026/03/06/openai-hits-25-billion-revenue-anthropic-closes-gap-xcxwbn/ y https://www.tradingkey.com/analysis /stocks/us-stocks/261490879-openai-revenue-advertisement-tradingkey
[b] Fuente: https://www.investing.com/news/company-news/anthropic-set-to-become-profitable-before-openai-amid-ai-race-4349418
[c] Fuente: https://es.tradingview.com/news/gurufocus:892a33de7094b:0-anthropic-quietly-raises-the-ceiling/
[d] Google DeepMind y las actividades de desarrollo de modelos de frontera (como Gemini) se consolidan contablemente en "Alphabet-level activities". Al ser un laboratorio que provee infraestructura cognitiva a toda la corporación, no genera ingresos externos directos de modo contable (aunque indirectamente se generan ingresos por los planes de Gemini y las ventas de API). La cifra de US$ 16,760 millones representa el costo operativo neto de la investigación avanzada en IA y gastos corporativos compartidos no asignados a los segmentos comerciales.
[e] Fuente: https://www.techdogs.com/tech-news/td-newsdesk/elon-musks-xai-reports-146-billion-loss-as-startup-rapidly-expands-operations

Cuadro A3.4
Plataformas de software y herramientas: rentabilidad, 2025

Empresa	Margen EBITDA
Hugging Face	Breakeven[a]
Databricks	Flujo de caja positivo[b]
LangChain	Negativo[c]
Suno	Bajo o negativo por altos costos operativos[d]

Fuente: Recopilación de los autores de fuentes públicas.
[a] Fuente: https://observer.com/2026/02/hugging-face-monetization-chief-jeff-boudier-business-model/
[b] Fuente: https://www.revefi.com/blog/databricks-growth-surges-hits-record-revenue-2026
[c] Fuente: https://dataconomy.com/2025/10/21/langchain-hits-1-25-billion-valuation-in-series-b-led-by-ivp/
[d] Fuente: https://www.musicinafrica.net/magazine/suno-reports-two-million-paid-subscribers-300m-annual-revenue

Cuadro A3.5
Desarrollo e implementación de aplicaciones: rentabilidad, 2025

Empresa	Margen EBITDA
Harvey	Negativo (Fase de escalamiento)[a]
Palantir	50,95%[b]
Abridge	Negativo (Fase de expansión)[c]
Accenture	Sin información específica para el segmento de IA[d]

Fuente: Recopilación de los autores de fuentes públicas.
[a] El CEO Winston Weinberg confirmó que la empresa alcanzó una tasa de ingresos anualizados de US$ 190 millones al cierre de 2025 (https://www.forbes.com/sites/iainmartin/2026/02/09/legal-ai-startup-harvey-in-talks-to-raise-200-million-at-11-billion-valuation/).
[b] El capital se quema activamente en la adquisición de usuarios y el pago de infraestructura de modelos de frontera.
[c] https://investors.palantir.com/news-details/2026/Palantir-Reports-Q4-2025-U-S--Comm-Revenue-Growth-of-137-YY-and-Revenue-Growth-of-70-YY-Issues-FY-2026-Revenue-Guidance-of-61-YY-and-U-S--Comm-Revenue-Guidance-of-115-YY-Crushing-Consensus-Expectations/
[d] Fuente: https://sacra.com/c/abridge/ . Sigue operando con margen negativo debido a la reinversión total de sus ingresos y capital (Serie E de US$ 300 millones) para integrar su tecnología en registros médicos electrónicos rígidos y escalar su infraestructura de ciberseguridad.

Anexo A4
Inversión en la cadena de valor de Inteligencia Artificial en Estados Unidos

Cuadro A4.1
Equipamiento: intensidad de capital, 2025
(En millones de dólares)

Segmento	Empresa	Capex	Ingresos	Capex/Ingresos
Diseño de GPUs	NVIDIA	2 369[a]	130 497	1,82%
	AMD	1 000[b]	34 639	2,89%
Fabricación de GPUs	TSMC	40 895[c]	122 424	33,40%
Ensambladores	Dell Technologies	2 652[d]	95 567	2,78%
	Supermicro	127[e]	21 972	0,58%
	Foxconn	2 306[f]	215 000 (Año Fiscal 2024)	1,07%

Fuente: Análisis de los autores.

[a] Fuente: 2025 NVIDIA Corporation Annual Review (Form 10-K). Estimado como la diferencia en el rubro "Property and equipment, net" entre el 2024 (US$ 3,914 M) y el 2025 (US$ 6,283 M). De ese valor sólo US$ 1,031 M corresponden a Estados Unidos (variación de US$ 2,595 M en 2024 a US$ 3,626 M en 2025). Aunque NVIDIA es el motor principal de la IA, este CAPEX físico incluye operaciones para todas sus divisiones. No obstante, el segmento de Data Center representó el 88.3% de sus ingresos en este período (US$ 115,186 M sobre US$ 130,497 M).

[b] Fuente: AMD Form 10-K (2025). La cifra corresponde estrictamente a la salida de efectivo por Purchases of property and equipment. Se excluyen los movimientos de tesorería en inversiones a corto plazo (US$ 5,500 M) y adquisiciones de empresas (US$ 1,800 M) para reflejar únicamente la inversión orgánica en infraestructura técnica y productiva.

[c] Fuente: Taiwan Semiconductor Manufacturing Co., Ltd. (TSMC), 4Q25 Management Report: Capital Allocation and Consolidated Statements of Cash Flows, 15 de enero de 2026. La cifra representa el gasto real ejecutado en adquisición de propiedad, planta y equipo durante el ejercicio fiscal 2025.

[d] Fuente: Reporte 10-K de Dell Technologies ante la SEC para el Año Fiscal 2025. La cifra representa el CAPEX físico destinado primordialmente a la optimización de capacidades de logística global, diseño de soluciones térmicas como refrigeración líquida y la adecuación de líneas de ensamblaje continuo para servidores de IA.

[e] Fuente: Formulario 10-K de Super Micro Computer, Inc. (Año Fiscal 2025). La inversión de US$ 127.2 millones se destinó principalmente a la expansión de servidores, centros de datos e infraestructura de red propia para dar soporte a sus operaciones globales de ensamblaje de alta densidad.

[f] Foxconn, con sede central en Taiwán, se incorpora en este análisis del mercado de Estados Unidos debido a su rol estructural como el músculo laboral de los hiperescaladores con una base masiva de empleados.

Cuadro A4.2
Infraestructura: intensidad de capital, 2025
(En millones de dólares)

Empresa	Capex	Capex/Ingresos
AWS	96 496[a]	74,98%
Microsoft Azure	64 550[b]	60,70%
Google Cloud	75 000[c]	127,76%[d]
CoreWeave (Neo-cloud)	10 309[e]	200,92%
Oracle	21 215[f]	43,09%

Fuente: Recopilación de los autores de fuentes públicas.

[a] Amazon.com, Inc. 2025 Annual Report (Form 10-K). La cifra corresponde a las "Total net additions to property and equipment" específicamente asignadas al segmento AWS durante el año fiscal 2025. Estas inversiones reflejan principalmente el despliegue de infraestructura tecnológica (servidores y equipo de red) para soportar el crecimiento del negocio de la nube y las iniciativas de IA y aprendizaje automático. Amazon incrementó esta inversión en AWS de US$ 53,267 millones en 2024 a US$ 96,496 millones en 2025, un aumento del 81% interanual.

[b] Microsoft Corporation 2025 Annual Report (Form 10-K). La cifra de US$ 64,551 millones corresponde a las "Additions to property and equipment" reportadas en el Estado Consolidado de Flujos de Efectivo para el año fiscal que cerró el 30 de junio de 2025. Esta inversión representa un incremento del 45.1% respecto a los US$ 44,474 millones invertidos en el ejercicio anterior. Según las declaraciones de la CFO, Amy Hood, y el CEO, Satya Nadella, la gran mayoría de este capital se destinó a expandir la capacidad de los centros de datos globales y a la adquisición de servidores (incluyendo CPUs y GPUs) para satisfacer la creciente demanda de servicios de Azure AI y el entrenamiento de modelos de próxima generación. La inversión en IA anunciada para el 2025 fue de US$ 80,000 (Fuente:

https://blogs.microsoft.com/on-the-issues/2025/01/03/the-golden-opportunity-for-american-ai/), entonces asumimos que la totalidad de la inversión reportada en el balance corresponde a IA.

[c] $91,447 fue la cifra que corresponde a los gastos de capital totales reportados por Alphabet Inc. para el año fiscal 2025. La inversión se destinó primordialmente a infraestructura técnica, la cual comprende la adquisición de servidores, equipos de red, terrenos para centros de datos y la construcción o mejora de edificios. Debido a la estructura operativa de la empresa, estos activos se gestionan de forma centralizada a nivel consolidado, por lo que el monto total incluye la infraestructura que da soporte tanto a Google Cloud como a otras divisiones de la compañía (como Google DeepMind). Por tal motivo para el segmento de Cloud se consideró los US$ 75,000 millones anunciados (Fuente: https://www.ciodive.com/news/google-cloud-generative-ai-data-center-capacity-buildouts/739357/)

[d] Google está realizando en 2025 una inversión acelerada para alcanzar el nivel de inversión desplegado previamente por AWS y Microsoft Azure.

[e] CoreWeave, Inc. (2026). Form 10-K: Annual Report for the fiscal year ended December 31, 2025. La compañía reportó desembolsos de efectivo para la compra de propiedad y equipo (incluyendo software de uso interno capitalizado) por US$ 10,309 millones durante el ejercicio. Adicionalmente, el informe detalla adiciones no monetarias por activos de infraestructura financiados (principalmente vía acuerdos con fabricantes u OEMs) por un valor de US$ 11,151 millones

[f] Oracle Corporation (2025). Form 10-K: Annual Report for the fiscal year ended May 31, 2025. La compañía reportó una inversión total en activos fijos de US$ 21,215 millones, lo que representa un incremento del 209% respecto a los US$ 6,866 millones invertidos en el ejercicio anterior. Según el informe, esta expansión masiva de capital se destinó primordialmente a incrementar la capacidad de sus centros de datos y adquirir el equipamiento computacional necesario para soportar el crecimiento de su negocio de nube (Oracle Cloud), respondiendo a la demanda actual y esperada de sus servicios de infraestructura.

Cuadro A4.3
Preparación y gestión de datos: Intensidad de capital, 2025
(US$ millones)

Empresa	Capex
Scale AI	Sin Información
Gretel.ai	Sin Información
Pinecone (Vector DB)	Sin información

Fuente: Recopilación de los autores de fuentes públicas

Cuadro A4.4
Modelos Fundacionales: intensidad de capital, 2025
(En millones de dólares)

Empresa	Información recopilada	Capex 2025
OpenAI	US$ 8 000 millones en el 2025 (destinado a infraestructura y cómputo, sin contar con la discriminación) US$ 600 000 millones anunciado en cómputo e infraestructura hasta el 2030[a] US$ 100 mil millones se destinarán a costos estrictamente vinculados a servidores y centros de datos que la propia empresa controlaría de forma directa	8 000
Anthropic	Plan de US$ 50,000 millones (anunciado en nov. 2025) para la construcción de centros de datos propios en Texas y Nueva York[b.] No se encontró inversión propia en CAPEX en el 2025, sólo a través de terceros (Por ejemplo, Project Rainier, a través de AWS) Inversión para entrenar modelos proyectada al 2026 por US$ 12,000 millones[c]	0
xAI	2025: Inversión de US$ 18,000 millones destinada a la compra de 555,000 GPUs NVIDIA (incluyendo modelos H100, H200 y GB200) para el complejo Colossus[d]. Proyección 2026: Inversión de US$ 20,000 millones para la construcción de un nuevo centro de datos de 2 gigavatios en Southaven, Mississippi[e].	18 000

Fuente: Recopilación de los autores de fuentes públicas.

[a] $91,447 fue la cifra que corresponde a los gastos de capital totales reportados por Alphabet Inc. para el año fiscal 2025. La inversión se destinó primordialmente a infraestructura técnica, la cual comprende la adquisición de servidores, equipos de red, terrenos para centros de datos y la construcción o mejora de edificios. Debido a la estructura operativa de la empresa, estos activos se gestionan de forma centralizada a nivel consolidado, por lo que el monto total incluye la infraestructura que da soporte tanto a Google Cloud como a otras divisiones de la compañía (como Google DeepMind). Por tal motivo para el segmento de Cloud se consideró los US$ 75,000 millones anunciados (Fuente: https://www.ciodive.com/news/google-cloud-generative-ai-data-center-capacity-buildouts/739357/)

[b] Google está realizando en 2025 una inversión acelerada para alcanzar el nivel de inversión desplegado previamente por AWS y Microsoft Azure.

[c] CoreWeave, Inc. (2026). Form 10-K: Annual Report for the fiscal year ended December 31, 2025. La compañía reportó desembolsos de efectivo para la compra de propiedad y equipo (incluyendo software de uso interno capitalizado) por US$ 10,309 millones durante el ejercicio. Adicionalmente, el informe detalla adiciones no monetarias por activos de infraestructura financiados (principalmente vía acuerdos con fabricantes u OEMs) por un valor de US$ 11,151 millones

[d] Oracle Corporation (2025). Form 10-K: Annual Report for the fiscal year ended May 31, 2025. La compañía reportó una inversión total en activos fijos de US$ 21,215 millones, lo que representa un incremento del 209% respecto a los US$ 6,866 millones invertidos en el ejercicio anterior. Según el informe, esta expansión masiva de capital se destinó primordialmente a incrementar la capacidad de sus centros de datos y adquirir el equipamiento computacional necesario para soportar el crecimiento de su negocio de nube (Oracle Cloud), respondiendo a la demanda actual y esperada de sus servicios de infraestructura.

[e] Fuente: https://apnews.com/article/xai-musk-data-center-mississippi-memphis-433691ace945708a04762b4791602f3d

Cuadro A4.5
Plataformas de software y herramientas: intensidad de capital, 2025
(En millones de dólares)

Empresa	Capex
Hugging Face	Sin información Se limita a equipos de oficina y desarrollo menor, ya que no posee data centers propios[a]
Databricks	Sin información Opera como una plataforma de datos e IA que se despliega sobre la infraestructura de terceros (AWS, Azure), encargándose de gestionar los recursos en la cuenta del cliente sin poseer hardware propio[b]
LangChain	Sin información Es un marco de trabajo (*framework*) de software orientado a la orquestación de agentes; su inversión se centra en el desarrollo de herramientas y talento, delegando el gasto en servidores y cómputo a los proveedores de nube[c]
Suno	Sin información Aunque requiere una potencia de cómputo masiva para la generación de audio, utiliza proveedores de servicios externos para el alojamiento, mantenimiento e infraestructura de sus sistemas de información[d]

Fuente: Recopilación de los autores de fuentes públicas.

[a] Fuente: https://moiglobal.com/lattice-work-hyperscalers/

[b] Fuente: https://docs.databricks.com/aws/en/introduction/

[c] Fuente: https://blog.langchain.com/series-b/

[d] Fuente: https://suno.com/privacy-policy

Cuadro A4.6
Desarrollo e implementación de aplicaciones: intensidad de capital, 2025
(En millones de dólares)

Empresa	Capex
Harvey	Sin información[a]
Palantir	33,9 millones[b]
Abridge	Sin información[c]
Accenture	Sin información para el segmento de IA[d]

Fuente: Recopilación de los autores de fuentes públicas.

[a] Al ser una plataforma de software vertical para el sector legal, opera bajo un modelo asset-light sin propiedad de centros de datos.

[b] Según el Informe Anual (Form 10-K) de Palantir Technologies Inc. para el año fiscal finalizado el 31 de diciembre de 2025, el gasto de capital destinado a la compra de propiedad y equipo fue de US$ 33.9 millones. Este nivel de inversión es significativamente bajo en comparación con los proveedores de infraestructura de IA debido a que Palantir opera bajo un modelo de software que depende de la infraestructura de computación en la nube de terceros, para alojar y operar sus plataformas. Al no construir centros de datos propios ni adquirir flotas masivas de GPUs para su propiedad, la empresa traslada el costo de la infraestructura física hacia el gasto operativo por servicios de nube.

[c] Como proveedor de software de transcripción ambiental para salud, delega la infraestructura pesada a hiperescaladores, enfocando su capital en el desarrollo de modelos.

[d] US$ 600 millones reportado en su informe anual 2025 como "adiciones a propiedad y equipo", reflejando la inversión en su red global de oficinas. No es necesariamente gasto relacionado a infraestructura de IA.

Anexo A5
Capitalización de mercado en la cadena de valor de Inteligencia Artificial en Estados Unidos

Cuadro A5.1
Equipamiento: capitalización de mercado, marzo 2026

Segmento	Empresa	Dinámica de valuación	Valuación estimada
Diseño de GPUs	NVIDIA	Valuación bursátil	US$ 4 340 000 millones 33,4 veces sus ingresos anuales
	AMD	Valuación bursátil	US$ 310 000 millones 8,9 veces sus ingresos anuales
Fabricación de GPUs	TSMC	Valuación bursátil	US$ 1 780 000 millones 14,6 veces sus ingresos anuales
Ensambladores	Dell Technologies	Valuación bursátil	$ 97 500 millones 1,0 veces sus ingresos anuales
	Supermicro	Valuación bursátil	US$ 19 500 millones 0,9 veces sus ingresos anuales
	Foxconn (ODM)	Valuación bursátil	US$ 2 550 000 millones 11,9 veces sus ingresos anuales

Fuente: Recopilación de los autores en base a datos de Yahoo Finance.

Cuadro A5.2
Infraestructura: capitalización de mercado, marzo 2026

Empresa	Dinámica de valuación	Valuación estimada
AWS	Valuación bursátil de Amazon, que incluye a AWS, pero también al resto de la empresa	US$ 2,35 billones 3.3 veces los ingresos de la empresa en su conjunto Ponderado por ingresos, el valor del segmento es de US$ 421,865 millones
Microsoft Azure	Valuación bursátil de Microsoft, que incluye a Azure, pero también al resto de la empresa	US$ 3,04 billones 10,8 veces los ingresos de la empresa en su conjunto Ponderado por ingresos, el valor del segmento es de US$ 1,146,674 millones
Google Cloud	Valuación bursátil de Google, que incluye a Cloud, pero también al resto de la empresa	US$ 3,70 billones 9,2 veces los ingresos de la empresa en su conjunto Ponderado por ingresos, el valor del segmento es de US$ 539 198 millones
CoreWeave (Neo-cloud)	Valuación bursátil	US$ 39 279 millones 7,7 veces los ingresos anuales de la empresa
Oracle	Valuación bursátil de Oracle, que incluye a Cloud, pero también al resto de la empresa	US$ 429 662 millones 13,4 veces los ingresos de la empresa en su conjunto Ponderado por ingresos, el valor del segmento es de US$ 368 513 millones

Fuente: Recopilación de los autores en base a Yahoo Finance.

Cuadro A5.3
Preparación y gestión de datos: capitalización de mercado, marzo 2026

Empresa	Dinámica de valuación	Valuación estimada
Scale AI	En base a ronda Serie G de Junio, 2025	US$ 29 000 millones (junio 2025)[a]
Gretel.ai	Precio de adquisición pagado por NVIDIA en marzo de 2025	US$ 320 millones (marzo de 2025)[b]
Pinecone (Vector DB)	Valuación “post-money” de la última ronda de financiamiento Según ronda de financiamiento del 2023 el valor era de US$ 750 millones	US$ 1 680 millones[c]

Fuente: Recopilación de los autores de fuentes públicas.
[a] Fuente: https://www.pminsights.com/companies/scale-ai
[b] Fuente: https://www.crn.com/news/ai/2025/nvidia-buys-synthetic-data-startup-gretel-to-boost-ai-llms-report
[c] Fuente: https://www.pminsights.com/companies/pinecone

Cuadro A5.4
Modelos fundacionales: capitalización de mercado, marzo 2026

Empresa	Dinámica de valuación	Valuación estimada
OpenAI	Valuación "pre-money"	US$ 730 000 millones[a]
Anthropic	Valuación "pre-money	US$ 350,000 millones[b]
xAI	Valuación "post-money" tras Serie E y absorción por SpaceX.	US$ 250,000 millones[c]

Fuente: Recopilación de los autores de fuentes públicas.
Nota: Se excluye a Google por no ser Pure Play y por no reportar ingresos directos del modelo fundacional en su memoria y balance.
[a] Fuente: https://www.cnbc.com/2026/02/20/openai-resets-spend-expectations-targets-around-600-billion-by-2030.html
[b] Fuente: https://www.investing.com/news/stock-market-news/anthropic-plans-employee-share-sale-at-350-bln-valuation-bloomberg-4483510
[c] Fuente: https://techfundingnews.com/spacex-ipo-march-filing-1-75-trillion-valuation/ . SpaceX adquirió xAI en una transacción de acciones que valoró a xAI como una subsidiaria de US$ 250,000 millones dentro de un holding combinado de US$ 1,250,000 millones.

Cuadro A5.5
Plataformas de software y herramientas: capitalización de mercado, marzo 2026

Empresa	Dinámica de valuación	Valuación estimada
Hugging Face	Ronda privada (Serie D) y transacciones activas en el mercado secundario	US$ 4 500 millones[a] 34,6 veces sus ingresos anuales
Databricks	Ronda Serie L (Diciembre, 2025) liderada por Insight Partners; preparación para IPO	US$ 134 000 millones[b] 24,8 veces sus ingresos anuales
LangChain	Ronda Serie B (Octubre, 2025) liderada por IVP para expandir ingeniería de agentes	US$ 1 250 millones[c] 80 veces sus ingresos anuales
Suno	Ronda Serie C (Noviembre, 2025) liderada por Menlo Ventures para un ecosistema de audio	US$ 2 450 millones[d] 8,1 veces sus ingresos anuales

Fuente: Recopilación de los autores de fuentes públicas.
[a] Fuente: https://www.joinprospect.com/explore/hugging-face-stock
[b] Fuente: https://www.prnewswire.com/news-releases/databricks-grows-55-yoy-surpasses-4-8b-revenue-run-rate-and-is-raising-4b-series-l-at-134b-valuation-302643445.html
[c] Fuente: https://dataconomy.com/2025/10/21/langchain-hits-1-25-billion-valuation-in-series-b-led-by-ivp/
[d] Fuente: https://www.goodwinlaw.com/en/news-and-events/news/2025/11/announcements-technology-goodwin-advises-suno-on-250-million-series-c-financing

Cuadro A5.6
Desarrollo e implementación de aplicaciones: capitalización de mercado, marzo 2026

Empresa	Dinámica de valuación	Valuación estimada
Harvey	Ronda de financiamiento	US$ 8 000 millones a 11 000 millones[a] 58 veces sus ingresos anuales
Palantir	Valuación bursátil	US$ 361 932 millones[b] 81 veces sus ingresos anuales
Abridge	Ronda de financiamiento	US$ 5 300 millones[c] 53 veces sus ingresos anuales

Fuente: Recopilación de los autores de fuentes públicas.
Nota: No se incluye a Accenture por incluir en su valuación otros segmentos de negocio.
[a] Fuente: https://www.forbes.com/sites/iainmartin/2026/02/09/legal-ai-startup-harvey-in-talks-to-raise-200-million-at-11-billion-valuation/
[b] Yahoo Finance.
[c] Fuente: https://www.forbes.com/sites/iainmartin/2026/02/09/legal-ai-startup-harvey-in-talks-to-raise-200-million-at-11-billion-valuation/

El análisis de la cadena de valor de la inteligencia artificial ofrece un marco analítico concreto para vincular la transformación digital con la agenda de desarrollo productivo de América Latina y el Caribe. Este enfoque permite examinar no solo la adopción de la tecnología, sino también la posición que ocupa la región en la nueva estructura global de creación y apropiación de valor.

El documento tiene por objetivo analizar, modelar y cuantificar la transformación de la cadena de valor de la industria digital ocasionada por la irrupción de la inteligencia artificial (IA), identificando los mecanismos de captura de valor y sus efectos en la organización industrial. Tomando como punto de partida un análisis de estas tendencias a nivel internacional, que se presenta en la primera parte, el trabajo tiene como propósito ofrecer un marco de análisis que permita determinar la posición de América Latina y el Caribe en dicha cadena y formular políticas públicas para que la región construya una posición competitiva en el escenario mundial.

Versión digital disponible online

https://bit.ly/CEPAL2026-117S

CEPAL

Comisión Económica para América Latina y el Caribe (CEPAL)
Economic Commission for Latin America and the Caribbean (ECLAC)
www.cepal.org