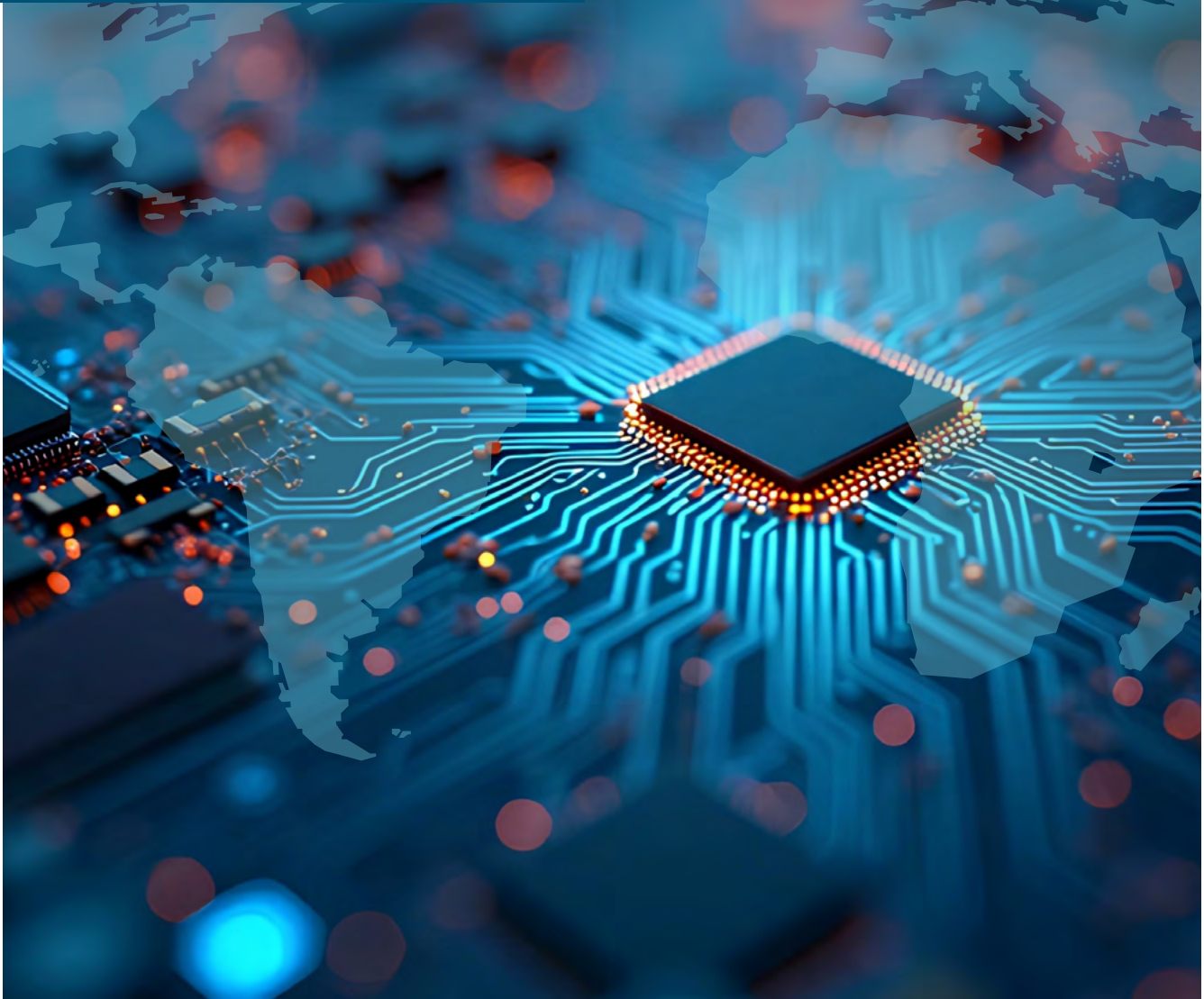




La base material de la
inteligencia artificial:

Chips, datacenters y geopolítica





Felipe Bravo-Márquez

PhD por la Universidad de Waikato, Nueva Zelanda. Profesor Asociado del Departamento de Ciencias de la Computación de la Universidad de Chile y Director de la Iniciativa de Datos e Inteligencia Artificial. Investigador Principal del Centro Nacional de Inteligencia Artificial e Investigador del Instituto Milenio Fundamentos de los Datos. Sus intereses incluyen NLP, aprendizaje automático, inteligencia artificial y recuperación de información.

✉ fbravo@dcc.uchile.cl



Claudio Gutiérrez

Ph.D. en Computer Science por la Wesleyan University, Estados Unidos, y Licenciado en Matemáticas por la Universidad de Chile. Profesor Titular del Departamento de Ciencias de la Computación de la Universidad de Chile e Investigador Senior del Instituto Milenio Fundamentos de los Datos (IMFD). Sus líneas de investigación son los fundamentos de la computación, la lógica aplicada a la computación, las bases de datos, la semántica de la Web y las máquinas sociales.

✉ cgutierrez@dcc.uchile.cl



Luis Mateu

Doctor en Informática por la Université de Paris VI, Francia. Profesor de Jornada Parcial del Departamento Ciencias de la Computación de la Universidad de Chile. Sus áreas de interés son la arquitectura de computadores y los lenguajes de programación.

✉ lmateu@uchile.cl

Resumen / El reciente auge de la inteligencia artificial ha devuelto al hardware su papel central en la industria tecnológica. Entrenar y ejecutar modelos avanzados requiere enormes capacidades de computación, por lo que las GPUs y la cadena global de semiconductores se han convertido en un ámbito de competencia estratégica. Un reducido número de empresas, como NVIDIA, TSMC y ASML, concentra las capacidades clave para fabricar y operar chips de última generación.

Al mismo tiempo, la inteligencia artificial depende de una infraestructura material que requiere grandes cantidades de minerales, energía, agua y centros de datos. En este contexto, distintas regiones del mundo (incluida América Latina) comienzan a integrarse gradualmente en esta nueva geografía del cómputo, lo que abre oportunidades, pero también plantea interrogantes sobre cómo participar en la economía global de la inteligencia artificial.

Introducción

Durante gran parte de las últimas tres décadas, la industria tecnológica digital estuvo dominada fundamentalmente por la idea de que el valor real residía en las capas lógicas: el software y los datos. Este cambio se hizo patente con el desplazamiento de IBM (el icono clásico del hardware) por parte de Microsoft (el software) hace casi medio siglo. El poder pasó de quienes controlaban la infraestructura física a quienes controlaban el software y, posteriormente, las plataformas (Google). El hardware se convirtió en un bien de consumo intercambiable cuya producción podía delegarse, y quien controlaba las plataformas y las experiencias de usuario tenía el poder.

Sin embargo, la irrupción de la Inteligencia Artificial (IA) generativa y los Grandes Modelos de Lenguaje (LLMs por su sigla inglesa) ha roto esa tendencia. La IA es, por definición, una tecnología “hambrienta” de cómputo donde entrenar un modelo del estado del arte es una operación industrial que requiere billones de multiplicaciones de matrices que las CPUs tradicionales no pueden realizar con eficiencia [1]. Este fenómeno ha situado a las Unidades de Procesamiento



Nodo de fabricación	Aplicación principal	Estado estratégico
28 nm y superiores	Electrónica de consumo masivo (IoT, automotriz básica, electrodomésticos), chips “commodities” de bajo costo	Madura / Sin restricciones. Son los procesos más económicos en términos de costo fijo y suficientemente eficientes para aplicaciones que no requieren alta densidad ni velocidad.
14 nm – 10 nm	Principalmente memorias y algunos chips heredados o de reemplazo	Estable / Bajo escrutinio. La reducción de escala ofrece beneficios limitados en ciertas aplicaciones, por lo que estos nodos se mantienen por razones de costo y madurez tecnológica.
7 nm – 5 nm	GPUs de alto rendimiento (incluyendo IA), CPUs de notebooks y móviles de gama media alta	Vanguardia / Altamente restringida. Permiten alta densidad y eficiencia, aunque con costos crecientes. Se utilizan en chips grandes y en procesadores modernos que equilibran costo y desempeño.
3 nm – 2 nm	CPUs y sistemas integrados en un solo chip (SoC), que combinan CPU, GPU y otros componentes en un único circuito, típicos de notebooks y dispositivos móviles de última generación. De forma incipiente, comienzan a explorarse también en GPUs de nueva generación.	Frontera tecnológica / Reservados para diseños donde la eficiencia energética y la densidad justifican el alto costo. Su uso en chips muy grandes, como GPUs, sigue siendo limitado por desafíos de rendimiento de fabricación, aunque fabricantes como NVIDIA están comenzando a migrar progresivamente hacia estos nodos en sus arquitecturas más recientes. [6]

Tabla 1 / Estado de los nodos de fabricación de semiconductores (2025-2026).

to Gráfico (GPU), un tipo especial de hardware paralelo, en el centro de la competencia industrial y geopolítica. Empresas como NVIDIA han alcanzado valoraciones de mercado históricas, superando a gigantes del software como Microsoft, simbolizando la revalorización de la infraestructura física como una frontera estratégica en el siglo XXI.

Esto ha traído impensadas tensiones geopolíticas. El software estandarizado había comenzado a hacer transparentes las fronteras entre países. Pero hoy, para los Estados la capacidad de ejecutar modelos de IA en territorio propio ha pasado a ser una garantía, tanto de soberanía y seguridad nacional como de poderío económico e industrial. Dependier exclusivamente de proveedores de nube externos (*hyperscalers*) ubicados en otras jurisdicciones expone a los países a riesgos de costos incontrolables, restricciones de acceso por cambios en políticas de exportación y la pérdida de control sobre información estratégica [2].

En lo que sigue, abordamos estos temas, comenzando por explicar los detalles técnicos que sostienen este nuevo revival del hardware y sus relaciones con la geopolítica mundial.

Hardware de la IA y geopolítica

El escalamiento nanométrico.

Las GPU son procesadores paralelos que permiten acelerar uno de los principales cuellos de botella de la IA: el cómputo masivo en paralelo. Su potencia reside, en gran medida, en la densidad de transistores integrados en el silicio. Tradicionalmente, este avance se ha descrito mediante los llamados “nodos” de fabricación, expresados en nanómetros (nm).¹ Un nanómetro equivale a una millonésima parte de un milímetro (para referencia, un cabello humano tiene aproximadamente 70.000 nm de grosor).

¹ Sin embargo, estas cifras ya no corresponden a una dimensión física específica del transistor, sino que funcionan como etiquetas que sintetizan un conjunto de mejoras tecnológicas. En este sentido, los nombres comerciales de los procesos de fabricación, como N3E de TSMC [3] o 18A de Intel [4], no describen una escala en nanómetros, sino una generación tecnológica completa que abarca el tipo de transistor, los materiales, las reglas de diseño y las técnicas de litografía empleadas por cada fabricante.

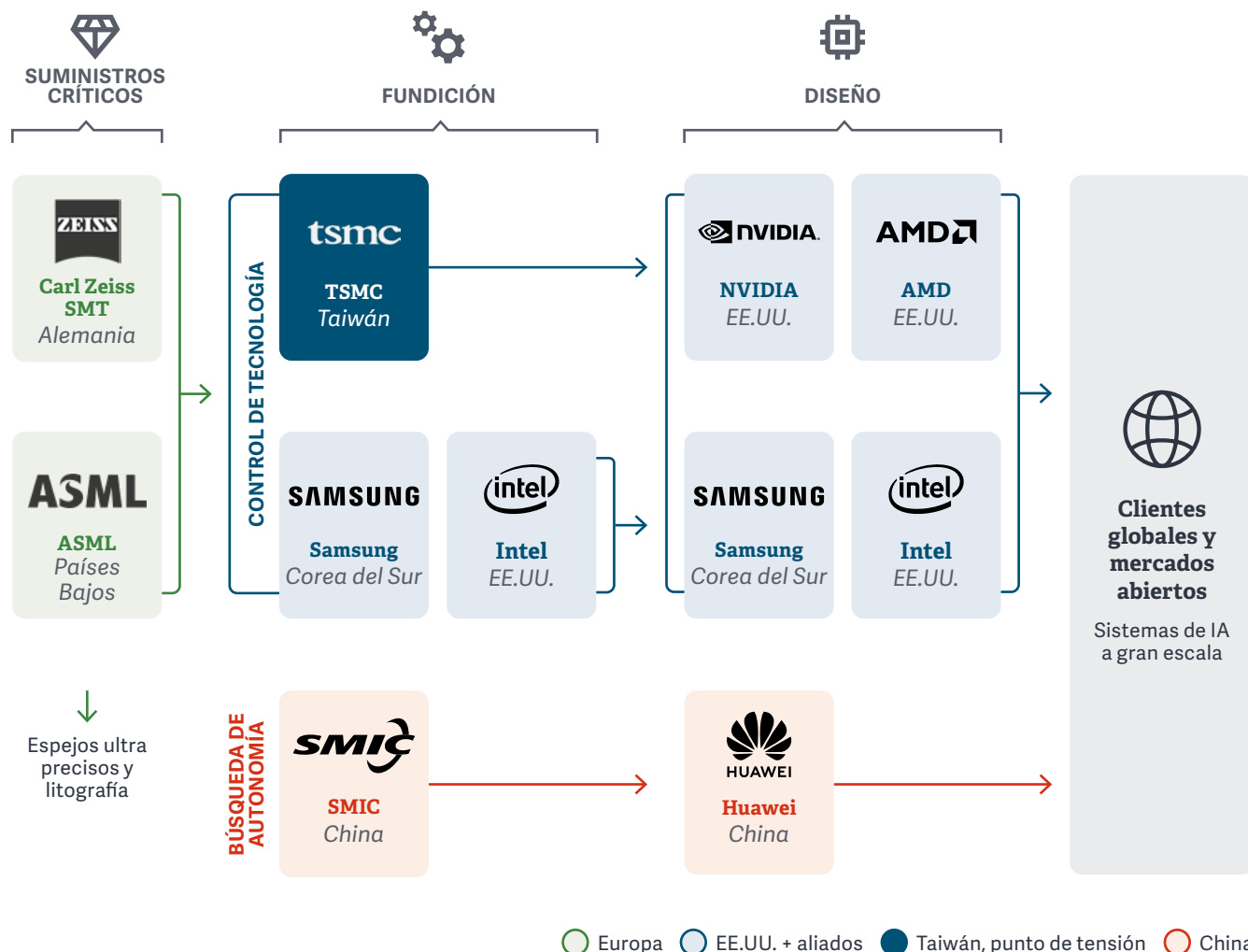


Figura 1 / La delicada cadena de suministro global de GPUs.

A partir de escalas cercanas a los 14 nm, las dimensiones críticas de los dispositivos comienzan a aproximarse a límites físicos y tecnológicos cada vez más exigentes. Este fenómeno se intensifica en generaciones posteriores y, hacia los 7 nm, se observa un punto de inflexión: el costo por transistor deja de disminuir de forma sistemática, rompiendo así una de las tendencias históricas más relevantes de la industria.²

Hoy en día, fabricar chips de última generación (como los de 3 nm) requiere un nivel de precisión extremadamente alto: algunas de sus partes son tan pequeñas que se acercan al tamaño de unos pocos átomos. Esto hace que di-

señarlos y producirlos sea cada vez más caro, porque las máquinas necesarias son muy complejas (como las herramientas litográficas —ver más adelante) y los diseños son mucho más difíciles de desarrollar. Como resultado, entrar en esta industria se ha vuelto muy difícil y costoso, lo que hace que sólo unas pocas empresas en el mundo —como TSMC o Samsung— puedan competir en este nivel tan avanzado [5].

En la Tabla 1 resumimos los distintos nodos de fabricación, sus aplicaciones principales y su relevancia estratégica. Asimismo, en la Figura 1 se presenta una visión esquemática de la cadena de suministro de GPUs.

² La famosa Ley de Moore. Hay que decir que la reducción de escala continúa siendo rentable gracias a mejoras en rendimiento y eficiencia energética por transistor.



Esta adopción basada exclusivamente en NVIDIA ha generado un “vendor lock-in” profundo: migrar a otro proveedor implicaría reescribir o adaptar años de código optimizado para CUDA y sus bibliotecas.

Litografía

El proceso de fabricación de un microchip comienza con la litografía, que es básicamente una técnica de “dibujo con luz”. Los chips electrónicos son discos muy delgados y perfectamente pulidos de silicio que sirven como base para fabricar chips electrónicos. Por eso les llaman “oleas de silicio”. Los patrones microscópicos que definen dónde irán los componentes del chip se graban (“dibujan”) en la oblea mediante la proyección de rayos ultravioleta a través de un patrón de molde. Cuando hablamos de nodos más avanzados (como los de 7, 5 o 3 nm), nos referimos a la capacidad de hacer esos dibujos cada vez más compactos. Cuanto más pequeño es el “nodo”, más transistores caben en el mismo espacio, lo que hace que el procesador sea más potente y consuma menos energía.

En la frontera actual de la tecnología, alcanzar estos niveles de miniaturización depende críticamente de la litografía ultravioleta extrema, conocida como EUV (siglas en inglés). Esta capacidad está altamente concentrada en una sola empresa, la holandesa ASML, cuyos equipos son, en la práctica, los únicos capaces de imprimir los patrones necesarios para fabricar chips de última generación.

Cada una de estas máquinas es una obra maestra de ingeniería que condensa décadas de inversión y desarrollo. Los modelos más recientes, conocidos como High NA EUV y desplegados progresivamente entre 2025 y 2026, tienen un costo cercano a los 370 millones de dólares por unidad [7]. Estas nuevas generaciones de equipos están pensadas para habilitar futuros nodos tecnológicos, como los procesos 14A en desarrollo por Intel, mientras que los nodos actualmente en despliegue industrial, como 18A, utilizan variantes más maduras de EUV.

Esta dependencia de un único proveedor ha creado un cuello de botella que trasciende lo comercial para volverse un asunto de soberanía nacional. El acceso a las máquinas de

ASML se ha transformado en una herramienta geopolítica: Estados Unidos y sus aliados limitan su venta a países como China para frenar su avance en inteligencia artificial y sistemas militares. Sin embargo, el monopolio de ASML encierra una paradoja: ellos mismos dependen de un tercero. Los espejos ultraprecisos necesarios para reflejar la luz ultravioleta dentro de estas máquinas son fabricados exclusivamente por la división SMT de Carl Zeiss (alemana) [8]. Actualmente, no existe otra entidad en el mundo capaz de fabricar esos espejos con la precisión requerida. Así, el eslabón más importante de la tecnología moderna depende, irónicamente, de un proveedor de espejos...

Automatización del diseño electrónico

La dependencia de HW-SW es muy estrecha. Antes de que un chip llegue a una fundición debe ser diseñado mediante software de Automatización de Diseño Electrónico (EDA), un mercado altamente concentrado dominado hoy por tres empresas: Synopsys, Cadence Design Systems y Siemens EDA.

Estas plataformas permiten diseñar circuitos con miles de millones de transistores y optimizar variables críticas como consumo energético, área y rendimiento. Sin ellas resulta prácticamente imposible desarrollar chips competitivos. Por lo mismo, el software EDA se ha transformado también en una herramienta de poder y disputa geopolítica.

Estados Unidos ha utilizado este control como instrumento de política tecnológica, restringiendo en distintos momentos la exportación de estas herramientas para ciertos países o para nodos tecnológicos avanzados (por ejemplo, inferiores a 14 nanómetros). Sin embargo, estas restricciones no han sido estáticas: se endurecen o se relajan según el contexto geopolítico [9].

Un dato estratégico interesante para nuestra región es que Synopsys mantiene oficinas en Santiago, donde ingenieros locales participan en el desarrollo de estas herramientas críticas. A diferencia de la fabricación física de chips, que requiere inversiones de decenas de miles de millones de dólares, el software EDA depende principalmente de capital humano altamente calificado. Esto abre una oportunidad para países que no poseen capacidad de fabricación avanzada, ya que les permite integrarse en una de las partes más intensivas en conocimiento de la cadena de valor de los semiconductores.

Fundiciones

Finalmente, una vez completado el diseño, este se envía a una fundición (“foundry”) para su materialización. El panorama de las fundiciones está marcado por una concentración

Fundición	País	Cuota de mercado en chips avanzados	Estado estratégico
TSMC	Taiwán	~90%	Líder absoluto; fabrica GPUs de NVIDIA, AMD y Apple
Samsung	Corea del Sur	Competidor clave	Fuerte en servicios de fundición y memorias HBM
Intel	Estados Unidos	Intentando recuperar terreno	Foco en servicios de fundición para el bloque occidental
SMIC	China	Limitada a tecnologías antiguas	Restringida por falta de máquinas EUV

Tabla 2 / Principales fundiciones de semiconductores avanzados.

extrema en el sudeste asiático (véase Tabla 2). Mientras TSMC domina ampliamente la producción de chips avanzados y Samsung Electronics se mantiene como su principal competidor, Intel intenta recuperar terreno con su estrategia de servicios de fundición. Por su parte, la fundición china SMIC enfrenta fuertes limitaciones tecnológicas debido a las restricciones de acceso a máquinas EUV, lo que la mantiene confinada principalmente a nodos de fabricación más antiguos [10]. Sin embargo, reportes recientes sugieren que la empresa ha logrado avances parciales utilizando técnicas de multipatrón con litografía DUV³, lo que evidencia cierta capacidad de adaptación tecnológica pese a las restricciones de exportación [11, 12].

La hegemonía de TSMC no se explica sólo por la disponibilidad de máquinas EUV, sino también por sus excepcionales niveles de rendimiento, es decir, el porcentaje de chips utilizables producidos en cada oblea de silicio. En nodos de 3 nm, TSMC ha logrado yields superiores al 70%, mientras que sus competidores enfrentan tasas significativamente menores. Esto hace que la producción de chips avanzados fuera de Taiwán sea, por ahora, mucho más costosa y arriesgada [13].

Esta concentración productiva convierte a Taiwán en un punto crítico para la economía global. A ello se suma un factor geopolítico ineludible: China reclama históricamente la soberanía sobre la isla de Taiwán y ha planteado en diversas ocasiones la posibilidad de una reunificación forzada. Un escenario de conflicto o interrupción significativa de las operaciones de TSMC tendría consecuencias sistémicas,

afectando no sólo la producción de GPUs y servidores de inteligencia artificial, sino también amplias cadenas industriales a nivel global. En términos de impacto económico, diversos analistas lo comparan con disrupciones mayores en mercados energéticos estratégicos, dada la centralidad de los semiconductores en la infraestructura tecnológica contemporánea. [14]

El ecosistema de hardware y software

Decíamos que la dependencia HW-SW es compleja y muy entrelazada. La dominancia de NVIDIA no se explica únicamente por sus chips, sino por su ecosistema de software propietario: CUDA (Compute Unified Device Architecture). Lanzado en 2007, se convirtió en el estándar de facto para la computación científica y el aprendizaje profundo. Durante casi dos décadas, universidades, empresas y proyectos open source han construido sus capas de software sobre CUDA. Esta adopción basada exclusivamente en NVIDIA ha generado un “*vendor lock-in*” (atadura al proveedor) profundo: migrar a otro proveedor implicaría reescribir o adaptar años de código optimizado para CUDA y sus bibliotecas.

Sin embargo, comienzan a aparecer alternativas. AMD impulsa ROCm como un stack abierto compatible con frameworks de IA modernos, mientras que Apple apuesta por su propio ecosistema basado en Metal para ejecutar modelos en los chips Apple Silicon [15]. En paralelo, Huawei desarrolla aceleradores y software propios para reducir la dependencia de CUDA en sus infraestructuras de IA.

3 DUV (Deep Ultraviolet) es una tecnología de litografía más antigua que utiliza luz ultravioleta profunda, típicamente con longitudes de onda de 193 nm, para definir patrones sobre obleas de silicio. Fue el estándar de la industria durante décadas previas a la introducción de la litografía EUV (Extreme Ultraviolet), que emplea longitudes de onda mucho más cortas y permite alcanzar nodos de fabricación más avanzados. Debido a sus limitaciones de resolución, la DUV requiere técnicas adicionales como multipatrón para aproximarse a escalas más pequeñas, lo que incrementa la complejidad y el costo del proceso.



[Latinoamérica] se consolida como un destino privilegiado para la instalación de data centers debido a la disponibilidad de recursos naturales, espacio físico y energía a costos competitivos.

Aunque la “brecha de CUDA” persiste, se aprecian intentos por diversificar el hardware que puede ejecutar modelos de aprendizaje profundo.

Un factor clave en este cambio es el rol creciente de frameworks como PyTorch y JAX. En lugar de programar directamente sobre el hardware, muchos desarrolladores trabajan con estas bibliotecas de alto nivel, que delegan la ejecución a distintos backends. Bajo este paradigma, estos frameworks actúan como una capa intermedia entre los modelos y los diversos tipos de aceleradores. Al mismo tiempo, muchos chips modernos incorporan NPUs (Neural Processing Units), unidades especializadas para ejecutar redes neuronales de forma eficiente. A diferencia de las GPUs, que están diseñadas para una amplia gama de operaciones paralelas, las NPUs están optimizadas para cargas de trabajo específicas de aprendizaje profundo, lo que en muchos casos permite una mayor eficiencia energética y una mejor relación costo–rendimiento en tareas de inferencia, especialmente en dispositivos personales. Por ejemplo, los procesadores de la familia Apple M-series, utilizados en dispositivos como el Mac mini, integran en un mismo chip CPU, GPU y una NPU denominada Neural Engine, lo que permite ejecutar modelos de IA directamente en el dispositivo. Gracias a estas arquitecturas integradas, comunidades de desarrolladores ya experimentan con la ejecución de LLMs en hardware personal, como Mac minis o PCs con GPUs, NPUs o aceleradores dedicados. En conjunto, estas tendencias apuntan hacia un ecosistema más diverso, donde la computación de IA se distribuye entre nubes, centros de datos y dispositivos personales.

Comunicación entre GPUs

Las cargas de trabajo de IA moderna no se ejecutan en una sola GPU, sino en clústeres de miles de ellas que deben intercambiar datos de forma casi instantánea. Las tecnologías de interconexión son, por tanto, el sistema circulatorio de los centros de datos de IA. NVIDIA controla estándares propietarios como NVLink e InfiniBand, optimizados para su propio hardware, lo que refuerza la dependencia de su ecosistema

completo. Frente a esto, la industria está impulsando estándares abiertos como Ethernet y nuevas iniciativas como Ultra Accelerator Link (UALink) para permitir que chips de diferentes proveedores trabajen juntos con baja latencia. La decisión de un Estado sobre qué arquitectura de red adoptar para sus supercomputadores nacionales definirá si queda atado a un proveedor específico o si mantiene la flexibilidad para integrar tecnologías diversas en el futuro.

La inserción estratégica de América Latina en la geopolítica de la inteligencia artificial

Brecha estructural, asimetría del “stack” y funciones

América Latina presenta en el terreno de la IA una brecha que no es meramente tecnológica sino *estructural* [17]. Esta asimetría se manifiesta en la ausencia de capacidades para el diseño de semiconductores (circuitos integrados) y un control nulo sobre las capas profundas del *stack* tecnológico (desde el silicio hasta los modelos fundacionales).

En este sentido, América Latina no es percibida por las grandes *big tech* como un polo de innovación autónomo, sino como un espacio de recursos que cumple tres funciones críticas para el sistema-mundo digital.

Primero, un *territorio de despliegue y extractivismo energético*: La región se consolida como un destino privilegiado para la instalación de *hyperscale data centers* debido a la disponibilidad de recursos naturales, espacio físico y energía a costos competitivos. Países como Chile, Brasil y México funcionan como “enclaves tecnológicos” donde la infraestructura física se beneficia de recursos naturales y marcos regulatorios favorables, similar a la dinámica histórica de la industria minera [18].

Segundo, como nodo en la *red global de cómputo*: La infraestructura local actúa como un punto de presencia (PoP) esencial para reducir la latencia en la red global de las *big tech*. Esta distribución geográfica permite la expansión de servicios de IA en tiempo real, optimizando el rendimiento de modelos generativos y servicios en la nube a escala continental.

Tercero, como espacio de *disputa geopolítica “blanda”*: En el marco de la “Guerra Fría Tecnológica” entre Estados Unidos y China, la región representa un terreno de disputa por la hegemonía de estándares y acceso a recursos. El caso del “cable chino” en Chile es un muy buen ejemplo de esta disputa [19]. La elección de proveedores de infraestructura



(e.g., la adopción de tecnología 5G de Huawei versus alternativas occidentales) define no sólo la dependencia tecnológica a largo plazo, sino también la alineación de los marcos regulatorios y de gobernanza de datos.

Geopolíticamente, la competencia regional no se centra necesariamente en el liderazgo de la frontera tecnológica, sino en las condiciones de su integración subordinada. Es importante entender que la presencia de infraestructura física en territorio nacional no equivale directamente a soberanía tecnológica.

El rol asignado por los actores globales

La estrategia de las empresas líderes hacia la región varía según su posición en la cadena de valor. Los dos principales grupos son los proveedores de HW y los de plataformas (*hyperscalers*).

Para los proveedores de hardware (NVIDIA, AMD, Intel), esto es, las firmas de semiconductores, América Latina es estrictamente un mercado de consumo masivo y un nodo de distribución de potencia de cómputo (vía *Cloud*). No existe una intención manifiesta de descentralizar el diseño o la manufactura hacia la región. La relación es de proveedor-cliente: se busca capturar el crecimiento de la demanda de GPUs para centros de datos locales sin transferir las capacidades de innovación que definen el valor agregado del sector.

Para las plataformas e *hyperscalers* (Google, Microsoft, AWS), la región es estratégica y multidimensional. Por un lado, la instalación de “Regiones de Cloud” busca la cobertura total del

mercado latinoamericano. Por otro, al integrar a gobiernos, *startups* y grandes empresas en sus ecosistemas de nube, aseguran que el desarrollo de IA local se construya exclusivamente sobre sus *stacks* propietarios (enclaustramiento tecnológico o *vendor lock-in*). Finalmente, la cercanía física / geográfica permite incidir en las políticas públicas locales, optimizar costos operativos y asegurar que la regulación nacional sea compatible con sus modelos de negocio globales.

En resumen, América Latina se encuentra en una encrucijada donde la abundancia de infraestructura física podría hacer perder de vista una creciente dependencia cognitiva y técnica. La región compite por ser el soporte material de la IA global, mientras el control intelectual y económico de la misma permanece concentrado fuera de sus fronteras.

El caso de Chile

En el escenario regional, Chile ha adoptado una postura relativamente proactiva en el desarrollo de infraestructura y capacidades para inteligencia artificial, articulada en torno a tres tendencias: habilitación regulatoria, fortalecimiento del supercómputo y descentralización incipiente [20, 21].

Por una parte, el país ha avanzado en generar condiciones para la instalación de centros de datos, apoyándose en su matriz energética con alta penetración de renovables, lo que lo posiciona como un nodo atractivo para cómputo intensivo a nivel regional [22].

Por otra, se observa un fortalecimiento de capacidades nacionales de supercómputo a través de iniciativas académicas y



público-privadas, como el NLHPC en la Universidad de Chile o desarrollos impulsados por la Pontificia Universidad Católica de Valparaíso en conjunto con actores tecnológicos. Estas iniciativas combinan capacidades para entrenamiento de modelos y provisión de servicios de inferencia, con foco en habilitar tanto investigación como adopción productiva [23].

Finalmente, comienzan a emerger esfuerzos de descentralización, como el proyecto de supercómputo de la Universidad de Tarapacá en el norte del país, que apunta a desconcentrar capacidades y fortalecer la resiliencia territorial de la infraestructura digital [24].

El desafío hacia adelante no es sólo expandir esta base material, sino articularla con el ecosistema productivo y científico. En particular, será clave asegurar transferencia tecnológica efectiva y desarrollo de capacidades locales, evitando una dinámica de “extractivismo digital” donde la infraestructura se instala localmente, pero el valor es capturado externamente [25].

Conclusiones

A partir del análisis presentado, es posible extraer algunas conclusiones estratégicas sobre la base material de la inteligencia artificial y su impacto en el tablero geopolítico global, con especial énfasis en América Latina y Chile.

En primer lugar, vemos un retorno a la materialidad y los temas de soberanía tecnológica. El boom de las aplicaciones de IA ha generado la ilusión de que su valor reside exclusivamente en el software. Hoy, como hemos visto, la soberanía nacional y el poderío económico dependen de la capacidad física de cómputo y del acceso a una cadena de suministro de semiconductores extremadamente concentrada. Esto levanta amenazas como la dependencia de proveedores externos de nube (*hyperscalers*) y de hardware específico, lo que genera riesgos de seguridad, costos incontrolables y vulnerabilidad ante cambios en políticas de exportación de potencias extranjeras. En este sentido, los Estados que invierten en infraestructura propia, como supercomputadores y data centers nacionales, infraestructura de redes y comunicaciones, pueden garantizar una mayor autonomía en la ejecución de modelos estratégicos o de alto poder.

Una segunda conclusión general es la vulnerabilidad del monopolio global —asunto que es buena noticia para balancear concentración única de poder. Hemos visto que el ecosistema de la IA descansa sobre “cuellos de botella” críticos como empresas como ASML (litografía), TSMC (fundición) y NVIDIA (diseño y software CUDA). Sin embargo, permanecen riesgos como la concentración de la producción avanzada (nodos de

3nm a 2 nm en Taiwán representan un peligro para la estabilidad global ante posibles conflictos geopolíticos en la zona). También hay riesgo de dependencia de sistemas, en particular hoy el ecosistema CUDA de NVIDIA que ha creado un *vendor lock-in* que obliga a la industria a seguir utilizando su hardware, dificultando la migración a alternativas más abiertas.

¿Cómo se posiciona América Latina en este escenario? La región se encuentra en una encrucijada donde corre el riesgo de ser meramente un soporte material para la IA de otros países y de las Big Tech. Actualmente América Latina actúa principalmente como territorio de despliegue para *data centers* (aprovechando energía y agua) y como nodo para reducir la latencia de servicios extranjeros. Se genera entonces un riesgo de “extractivismo digital”: el peligro de que los países financien la infraestructura física y energética, mientras que el valor intelectual y económico es capturado por plataformas globales (algo que la industria tradicional latinoamericana conoce bien...).

Para finalizar este análisis, sugerimos algunos caminos para la región y para Chile. Para evitar una “integración subordinada”, queremos proponer para la discusión las siguientes líneas de acción:

1. Fomentar el capital humano en capas lógicas, esto es, dado que fabricar chips avanzados es prohibitivamente caro, países como Chile pueden integrarse al apoyo en el diseño (por ej. mediante software EDA como las oficinas de Synopsys en Santiago), que requiere alta calificación técnica pero menor inversión física.
2. Es crucial para países o economías de menor volumen el impulso a la adopción de estándares abiertos. Esto significa impulsar el uso de arquitecturas como Ethernet o UALink y *frameworks* como PyTorch o JAX para romper monopolios como el de CUDA, y así permitir una infraestructura de hardware más diversa, flexible y accesible.
3. En el área de soberanía de datos y energía, es muy importante condicionar la instalación de grandes centros de datos a la transferencia tecnológica efectiva, en particular a la participación de personal local no sólo en tareas de servicios, y al cumplimiento de estrictos criterios de sostenibilidad ambiental, pues el medio ambiente es el recurso por el cual las grandes big tech se interesan en nuestros países hoy.
4. Finalmente, dada nuestras geografías, es muy relevante la descentralización, lo que significa llevar la capacidad de cómputo fuera de las capitales y las grandes ciudades, para aumentar la resiliencia digital y el desarrollo de los talentos y las capacidades locales. **B**

Referencias

- [1] Miller, C. (2022). *Chip war: The fight for the world's most critical technology*. Simon and Schuster.
- [2] Le Monde Diplomatique. (2021). *Geopolítica de los chips*. <https://www.lemondediplomatique.cl/2021/08/geopolitica-de-los-chips.html>.
- [3] Taiwan Semiconductor Manufacturing Company. *3nm technology*. https://www.tsmc.com/english/dedicatedFoundry/technology/logic/l_3nm.
- [4] Intel Corporation. (2023). *Semiconductor manufacturing process technologies*. <https://www.intel.la/content/www/xl/es/foundry/process.html>.
- [5] PatentPC. (2025). *The future of Moore's Law: Are we nearing the limit?* <https://patentpc.com/blog/the-future-of-moores-law-are-we-nearing-the-limit-latest-semiconductor-trends>.
- [6] Garreffa, A. (2025, November 9). *NVIDIA's unstoppable AI chip demand has TSMC boosting 3nm production by a huge 50%*. Tweak-Town. <https://www.tweaktown.com/news/108769/nvidias-unstoppable-ai-chip-demand-has-tsmc-boosting-3nm-production-by-a-huge-50-percent/index.html>.
- [7] IO+. (2026). *China's all-or-nothing race for chip independence*. <https://ioplus.nl/en/posts/chinas-all-or-nothing-race-for-chip-independence>.
- [8] Reuters. (2016, November 3). *ASML buys 24.9 percent stake in Carl Zeiss subsidiary SMT for 1 billion euros*. <https://www.reuters.com/article/world/europe/asml-buys-249-percent-stake-in-carl-zeiss-subsiary-smt-for-1-billion-euros-idUSKBN12YOKA/>.
- [9] Freifeld, K., & Misra, S. (2025, July 3). *As trade war truce with China holds, US lifts curbs for chip design software and ethane*. Reuters. <https://www.reuters.com/world/china/siemens-says-us-has-lifted-chip-software-curbs-china-bloomberg-news-reports-2025-07-03/>.
- [10] Techzine. (2025). *Calls for a 'Chinese ASML' are growing louder, but it is still a long way off*. <https://www.techzine.eu/news/infrastructure/139303/calls-for-a-chinese-asml-are-growing-louder-but-it-is-still-a-long-way-off/>.
- [11] TechInsights. (2025, December 11). *SMIC N+3 confirmed: Kirin 9030 analysis reveals how close SMIC is to 5nm*. <https://www.techinsights.com/blog/smic-n3-confirmed-kirin-9030-analysis-reveals-how-close-smic-5nm#register>.
- [12] Kr-Asia. (2026). *China aims for fivefold increase in advanced chip output to meet AI demand*. <https://kr-asia.com/china-aims-for-fivefold-increase-in-advanced-chip-output-to-meet-ai-demand>.
- [13] Tom's Hardware. (2026). *China's top chip execs admit fragmentation is undermining the country's ASML alternative*. <https://www.tomshardware.com/tech-industry/semiconductors/chinas-top-chip-execs-admit-fragmentation-is-undermining-the-countrys-asml-alternative>.
- [14] Center for Strategic and International Studies. (2022, October 11). *Why Taiwan matters: From an economic perspective*. <https://www.csis.org/analysis/why-taiwan-matters-economic-perspective>.
- [15] StrategyLens. (2026). *NVIDIA's CUDA vs AMD's ROCm: software ecosystems powering generative AI*. <https://strategytransforml.co/category/gpu-and-high-performance-compute-chips/article/nvidia-cuda-vs-amd-rocm-software-ecosystems-generative-ai>.
- [16] Jung, J., & Katz, R. L. (2026). *Impacto económico de la inteligencia artificial en América Latina: transformación tecnológica y reza-go en materia de inversión y capacidades laborales*. Comisión Económica para América Latina y el Caribe (CEPAL). <https://www.cepal.org/es/publicaciones/81909-impacto-economico-la-inteligencia-artificial-america-latina-transformacion>.
- [17] Lehdonvirta, V. (2022). *Cloud Empires: How Digital Platforms Are Overtaking the State and How We Can Regain Control*. MIT Press.
- [18] BBC News Mundo. (2026, marzo 4). *El cable de fibra óptica chino que rompió la transición de gobierno entre Boric y Kast en Chile (y por el que EE.UU. sancionó a funcionarios del país)*. <https://www.bbc.com/mundo/articles/c8d554v943lo>.
- [19] Barros, A. (2025, 3 de agosto). *Plan de Data Centers, aciertos y temas pendientes*. AlejandroBarros.com. <https://www.alejandrobarrros.com/plan-de-data-centers-aciertos-y-temas-pendientes/>.
- [20] Gobierno de Chile. (2024). *Plan Nacional de Data Centers 2024–2030*. Ministerio de Ciencia. https://minciencia.gob.cl/uploads/filer_public/95/6b/956b8c9f-d937-4b4d-8f6c-a871495a52ff/plan_nacional_de_data_centers_pdata.pdf.
- [21] DLA Piper. (2025). *Chile implements its national data center plan (2024–2030)*. <https://www.dlapiper.com/es-cl/insights/publications/2025/05/chile-implements-its-national-data-center-plan>.
- [22] Inria Chile. (2025, 26 de junio). *Inria Chile refuerza capacidad nacional de supercómputo para IA*. <https://inria.cl/es/inria-chile-refuerza-capacidad-nacional-de-supercomputo-para-ia-aliado-clave-de-los-dos-centros>.
- [23] El País. (2026, enero 19). *La Universidad de Tarapacá proyecta en Arica el mayor centro de supercómputo de Chile*. <https://elpais.com/chile/branded/2026-01-19/la-universidad-de-tarapaca-proyecta-en-arica-el-mayor-centro-de-supercomputo-de-chile.html>.
- [24] Effoduh, J. O. (2025). *Digital colonialism and the role of local intermediaries*. Cambridge University Press.