

Alex Fusté

Chief Global Economist Andbank

DeepSeek vs. Occidente. ¿Ventaja Estratégica o Amenaza Temporal?

“DeepSeek ha conseguido superar a ChatGPT en las tiendas móviles de aplicaciones, y también en pruebas de rendimiento y razonamiento con su modelo R1, y a una fracción del coste de entrenamiento”.

Esta frase ha corrido como la pólvora entre los inversores del mundo y, como ya sabrán, ha provocado un auténtico terremoto en los mercados; especialmente en el sector tecnológico occidental. Lo que pone nerviosos a los inversores es la última parte de la afirmación: *“a una fracción del coste de entrenamiento”*. Bien. He escuchado frases como *“se trata de un modelo más eficiente basado en el aprendizaje por refuerzo”*, no sé si en un intento de impresionarme. El aprendizaje por refuerzo (Reinforcement Learning), y específicamente el Aprendizaje por Refuerzo con Retroalimentación Humana (RLHF, por sus siglas en inglés), ya era ampliamente utilizado en los modelos de lenguaje más avanzados en Occidente hace más de un año. OpenAI introdujo el RLHF de forma prominente con GPT-3.5 y GPT4.

Observación 1: Transferencia de costes.

Es necesario saber que la cuestión del menor coste puede tener que ver con algo que no es en absoluto disruptivo. Es posible que DeepSeek, igual que Copilot, no sea un modelo autónomo completo. Copilot (con solo 10 millones de coste de entrenamiento) está entrenado en Codex, un modelo desarrollado con anterioridad por OpenAI y derivado de GPT-3. Significa que cuando Copilot sale a la luz, OpenAI ya había absorbido previamente la mayor parte del coste de entrenamiento del modelo base, dejando a Microsoft el simple trabajo de ajustar el modelo (*fine-tuning*) para tareas específicas relacionadas con el código. ¿Podemos derivar de ello que el modelo de IA Copilot tiene, al final del proceso, un coste significativamente menor? No. En absoluto.

Aunque Deepseek sea de código abierto, esta IA podría estar beneficiándose de una casuística similar a la de Copilot, mediante el uso de un modelo base preexistente. DeepSeek podría estar basado en un modelo preentrenado de código abierto, como LLaMA 2, Falcon o incluso GPT-NeoX. Significa que Deepseek está realizando de facto una transferencia de costes, ya que los costes asociados al modelo base habrían sido absorbidos por la organización que desarrolló dicho modelo original. Entonces, DeepSeek, igual que Microsoft, solo tendría que realizar el *fine-tuning* en un conjunto de datos específicos, que es mucho menos costoso que entrenar un modelo desde cero. Esto es bastante plausible, pues hay startups que suelen recurrir soluciones en la nube para entrenamientos específicos basados en modelos preexistentes, lo que reduce el gasto inicial. ¿Cuál es la diferencia entre Copilot y Deepseek? Que Microsoft pagó en 2019 USD 1000 millones a Open AI, y en 2023 otros 10.000 millones adicionales. Entiendo que en concepto de esa “transferencia de costes”.

Habrà que esperar y ver, pero se me hace muy difícil pensar que, tras décadas de trabajo en materia de redes neuronales, deep learning y, finalmente la IA, los creadores originales de estos modelos permitirán que esta situación de “entrenamiento por transferencia” se materialice dando lugar a modelos mejores, sin incurrir en los costes asociados.

Pero ¿puede realmente una empresa como Deepseek poner en jaque a los desarrolladores occidentales? Depende de la capacidad que estos últimos tengan para defenderse. En este sentido, deben saber que los desarrolladores de los modelos base tienen varias estrategias para proteger su trabajo y evitar que terceros creen modelos mejores mediante el uso de un modelo base preexistente. Estas defensas tienen que ver con:

1. Restricciones de Licencia.
2. Protección a través de Patentes y Propiedad Intelectual (que podrían dar lugar a multas billonarias).
3. Liderar en Innovación Continua con modelos en evolución constante: Los creadores originales suelen mantener una ventaja competitiva al lanzar iteraciones constantes de sus modelos, creando una importante barrera técnica para aquellos que intentan competir mediante "fine-tuning" (o copia). Esto es lo que probablemente ocurriría en este caso. Al fin y al cabo, la IA que tenemos hoy es la peor que tendremos. ¿Qué importa entonces que un competidor iguale los modelos actuales mediante entrenamiento por transferencia, si en poco tiempo (y como es previsible) aparecerán nuevos modelos más potentes?
4. Infraestructura exclusiva: OpenAI y Google tienen acceso a recursos de hardware y software optimizados que les permite seguir liderando la carrera y crear nuevas versiones.
5. Modelos protegidos contra extracción: Implementar técnicas para dificultar la extracción de conocimiento del modelo (modelo robado), como "adversarial training"

Observación 2: Subsidios

Un aspecto relevante a considerar en el análisis del coste reducido de este modelo de inteligencia artificial es el posible efecto de subvenciones gubernamentales, un fenómeno particularmente frecuente en China, especialmente cuando Pekín establece un "Plan Superior" para ciertos sectores, como el de baterías, EVs o paneles solares. No descartaría la posibilidad de la existencia de fuertes subvenciones que ayudarían a enmascarar los costes.

Se ha señalado que el desarrollo de DeepSeek cuenta con el apoyo explícito de un hedge fund denominado High-Flyer. Sin embargo, esta situación resulta inusual en el contexto de Occidente, donde los hedge funds no suelen desempeñar el papel de impulsores principales en el desarrollo de empresas tecnológicas. Este tipo de fondos suelen centrarse en estrategias de inversión que maximizan el retorno financiero a corto o medio plazo, en lugar de asumir los riesgos inherentes al desarrollo tecnológico a largo plazo, que suelen ser costosos, inciertos y de lenta maduración. Los Hedge Funds se dedican al arbitraje, trading algorítmico o estrategias macroeconómicas. También está la propia asimetría de conocimientos, algo que suele mantener a los Hedge Funds bien alejados de proyectos que no entienden y que requieren de una experiencia técnica profunda, así como de equipos especializados. Áreas donde los Hedge Funds suelen carecer de competencia.

Uno podría tener la impresión de que el apoyo de un Hedge Fund como High-Flyer no sea completamente independiente, sino que actúe como un vehículo para canalizar subvenciones

públicas significativas. Este tipo de dinámica, si se confirmara, podría representar un caso evidente de **dumping** y prácticas empresariales cuestionables desde el punto de vista ético y comercial. Dicho escenario no solo erosionaría la equidad en la competencia global, sino que probablemente desencadenaría una respuesta formal por parte de las autoridades regulatorias en Occidente y de organismos internacionales como la Organización Mundial del Comercio (OMC). La implicación de fondos públicos ocultos tras intermediarios financieros pondría de manifiesto una estrategia potencialmente desleal, cuya repercusión en el equilibrio comercial y tecnológico sería motivo de escrutinio internacional.

Observación 3: ¿Estamos comparando adecuadamente?

Un aspecto crucial a considerar en la evaluación comparativa de modelos de inteligencia artificial es la amplitud de sus funciones. Si DeepSeek está diseñado para tareas específicas, como búsqueda avanzada o análisis predictivo, es razonable esperar que sus requisitos computacionales sean significativamente menores en comparación con modelos generalistas como GPT-4, cuya arquitectura está optimizada para abordar una amplia gama de tareas. Este enfoque especializado podría explicar diferencias en eficiencia y coste. Sin embargo, una comparación rigurosa requiere un análisis más detallado de sus capacidades y objetivos, ya que actualmente no es evidente si estamos tratando elementos realmente equivalentes.

Observación 4: ¿Cuál sería el verdadero riesgo para las tecnológicas occidentales?

Una posible explicación para la reducción de costes de entrenamiento podría residir en el uso de hardware de nueva generación, específicamente GPUs o TPUs desarrolladas en China, que ofrecen un rendimiento superior en métricas clave. Estas métricas incluyen una mayor capacidad de cálculo (Floating Point Operations Per Second, FLOPs), una eficiencia energética significativamente mejorada (FLOPs por watt) y menores niveles de latencia (el tiempo requerido para procesar una tarea específica). En otras palabras, si los menores costes de operación del modelo **DeepSeek** derivan de la creación de nuevos superchips que superan a las soluciones actuales de Nvidia, esto representaría un desafío fundamental no solo para Nvidia, sino también para el liderazgo tecnológico de Occidente. Tal avance implicaría un cambio significativo en la competitividad tecnológica global y obligaría a un replanteamiento estratégico en la industria de semiconductores..

Conclusión. ¿Motivo para la alarma?

La materialización de un riesgo real relacionado con la supremacía de China en semiconductores requeriría evidencia concreta de que dispone de chips avanzados superiores a los actuales estándares de la industria. Sin embargo, el consenso predominante entre los expertos en semiconductores indica que, aunque China ha logrado avances notables en datos y algoritmos, sigue dependiendo de la tecnología extranjera en lo que respecta a semiconductores avanzados.

Un análisis destacado en 2023 por Ben Buchanan, de la Universidad de Georgetown, publicado en *Le Grand Continent*, señala que China carece de capacidades significativas en la fabricación de chips lógicos avanzados y sigue siendo dependiente del diseño de software y hardware proveniente de Estados Unidos. Si bien es cierto que los avances tecnológicos en este sector son

dinámicos y rápidos, y que la información de 2023 podría haber quedado desactualizada, un artículo más reciente de *Time* (agosto de 2024) reafirma esta brecha tecnológica. Según esta publicación, una investigación reciente confirma la disparidad considerable en la posesión de chips avanzados de inteligencia artificial, con Estados Unidos dominando el acceso a tecnologías de vanguardia como la **NVIDIA H100**, mientras China carece de estas capacidades.

De ser cierto este panorama, el gran riesgo asociado al desarrollo de "superchips" en China quedaría, al menos por ahora, significativamente mitigado. Este contexto refuerza la percepción de que, aunque China está avanzando en su ecosistema tecnológico, la supremacía en semiconductores avanzados sigue siendo una barrera estratégica que favorece a Estados Unidos.

A continuación, se presenta una tabla que detalla los principales parámetros de eficiencia y coste de los modelos de IA más destacados en el mercado. En relación con el coste de entrenamiento del modelo más reciente de **DeepSeek**, cabe destacar que esta cifra es una estimación, ya que la empresa no ha proporcionado información oficial sobre este aspecto. La proyección se basa en varias premisas: uso de chips comercializados y disponibles en el mercado actual, partiendo de unas supuestas capacidades y eficiencia que lo posicionarían como comparable a modelos avanzados como **GPT** (según reportes preliminares de diversas fuentes). En estas condiciones hipotéticas, el coste total de entrenamiento para DeepSeek se estima en **150 millones de USD**. No obstante, esta estimación deja abiertas dos cuestiones clave: Si DeepSeek emplea hardware innovador y de mayor rendimiento, lo que podría justificar un coste menor y una superioridad permanente. O si su competitividad se debe a una transferencia de costes debido al uso de modelos preentrenados. En tal caso, la presunta superioridad podría ser circunstancial y temporal.

Modelo	Número de Parámetros (B)	Eficiencia (Tokens/s por GPU)	Capacidad de Cálculo (FLOPS)	Coste de Entrenamiento (USD)	Coste Operativo Mensual (USD)
PT-4: OpenAI	175	400	1e17 → 100,000 billones	100M	500k
PaLM 2: Google DeepMind	540	500	1.2e17 → 120,000 billones	200M	700k
Claude 3: Anthropic	100	350	9e16 → 90,000 billones	50M	300k
LLaMA 2: Meta (antes Facebook)	70	450	7e16 → 70,000 billones	30M	200k
Falcon 180B: Technology Innovation Institute	180	420	1.1e17 → 110,000 billones	120M	600k
Copilot	12	320	5e16 → 50,000 billones	10M	100k
DeepSeek*		480	1.3e17 → 130,000 billones	150M*	

Cordiales saludos