



La irrupción de DeepSeek es un terremoto del que solo hemos empezado a ver los primeros temblores y que rediseñará el futuro del sector.

# ¡Que viene DeepSeek!



José García Montalvo

Este lunes, en los despachos corporativos y reuniones de ejecutivos y consejeros se repetía la misma frase: “¿Ya te has bajado la app de DeepSeek?”. No se notaba tanta excitación entre las compañías de IA desde que ChatGPT irrumpió en el imaginario público en noviembre de 2022. Y, mientras tanto, las grandes tecnológicas se desplomaban en los mercados. Donald Trump anunció la semana pasada que Oracle, Softbank y OpenAI habían comprometido una inversión de 500.000 millones de dólares para acelerar la IA, pero el lunes, al poco de la apertura del mercado, las tecnológicas norteamericanas habían perdido 675.000 millones de valor de mercado. Oracle caía un 13% y Soft-Bank un 8%. La concentración de los índices en los siete magníficos hacía que el Nasdaq cayera un 3,4%. Las acciones de Apple, criticada en el pasado reciente por su retraso en entrar en la guerra de la IA, subían un 3%. Y justo por eso. Paradojas del destino. Otra paradoja es que la retrasada Europa también ha sufrido parte del impacto. La holandesa ASML, que produce las máquinas litografía de luz ultravioleta extrema, fundamentales para producir chips de última generación de 3 ó menos nanómetros, caía también un 7,6%.

¿Qué ha pasado? El viernes, en su último número, *The Economist* había avanzado que dos modelos de inteligencia artificial hechos en China (QWQ de Alibaba y DeepSeek del fondo *quant High Flyers*) estaban haciendo significativos avances en

grandes modelos de IA tipo LLM (*Large Language Models*, como ChatGPT) a pesar de los crecientes obstáculos que Estados Unidos estaba imponiendo a China para avanzar en la utilización y producción de chips avanzados. Estaban a meses o incluso semanas de distancia de los modelos más avanzados producidos en Estados Unidos. Entre los obstáculos recordemos que Nvidia tiene prohibido exportar a China sus procesadores más avanzados y debe hacer versiones “degradadas” de sus chips para tener permiso para exportarlos. Por su parte, ASML también ha recibido amenazas de Estados Unidos para que no exporte máquinas litográficas avanzadas para impresión de circuitos integrados, impidiendo que las tecnológicas chinas puedan producir chips usando tecnologías Deep Ultra Violet (DUV) o Extreme UV (EUV).

## Unos competidores con costes mucho menores que el resto baja la expectativa de ingresos de todos

unos conocimientos tecnológicos muy avanzados para poder construir chips tan avanzados como los de TSMC o Nvidia. Mientras TSMC en Taiwan conseguirá este año hacer procesadores de 2 nanómetros, en China todavía están peleando con los 7 nanómetros. La idea era, por expresar lo gráfico, que con chips de nevera era inviable desarrollar un modelo de IA complejo y que aprendiera a gran velocidad. Sin embargo, DeepSeek ha puntuado más alto en los test de *benchmarking* que la mayoría de los otros modelos de IA, comparable solo con o1 de OpenAI y los modelos propios de Google.

¿Cómo lo han conseguido? ¿Cómo se puede hacer una IA avanzada con cafeteras? Parece que el sistema con-



siste en hacer más eficiente el uso de los procesadores jugando con la precisión de los cálculos (menor precisión para acelerar y mayor precisión en las fases más avanzadas del razonamiento), la conexión más eficiente entre procesadores evitando tiempos muertos y simultaneando tareas sin necesidad de esperar a que finalicen las mismas, y con un aprendizaje reforzado utilizando la versión del mismo modelo que usa cadenas de razonamiento para el procesamiento de preguntas difíciles, y que está también en la base de o1 de OpenAI.

## Menores costes

Pero, ¿por qué este impacto tan significativo sobre los mercados? Si fuera otra IA al uso no habríamos visto mucha acción. Desde ChatGPT de Open AI hemos asistido a una letanía de nuevos LLMs (Gemini de Google, Llama de Meta, Copilot de Microsoft, etc) que habían tenido externalidades positivas en la valoración de las empresas que los habían creado. En este caso hay varias diferencias significativas. En primer lugar, DeepSeek utiliza unos 2.000 procesadores no muy avanzados en lugar

de, por ejemplo, los 16.000 de la última versión de Llama. Por tanto, el coste fijo en infraestructura es mucho más bajo. El modelo de DeepSeek es mucho más grande: 685.000 millones de parámetros frente a los 405.000 millones de Llama 3.1. En tercer lugar, el coste de entrenar a ChatGPT 4.0 fue de unos 100 millones y se hablaba que para ChatGPT 5 podría superar los 1.000 millones de dólares. Entrenar a DeepSeek parece que ha costado tan solo 6 millones de dólares (algo más que los 4,5 millones que costó entrenar a ChatGPT 3.0). En cuarto lugar, se trata de un código con una licencia en abierto: cualquier programador puede bajarse los 20 gigas de código y roarlo en sus propios sistemas. En quinto lugar, y a diferencia de otros modelos de IA, los programadores de la compañía son libres de publicar los secretos del funcionamiento de DeepSeek en revistas científicas, lo que hace que el conocimiento se comparta con rapidez y pueda dejar obsoletos sistemas aparentemente más sofisticados. Finalmente, pero no menos importante, hacía ya meses que se hablaba de una burbuja en las

empresas de IA, y este acontecimiento ha coordinado a los agentes del mercado en un cambio claro de expectativas.

Las grandes diferencias respecto a los modelos de IA anteriores hacen que su impacto sea mucho mayor. En un momento de evidente guerra de precios entre las distintas empresas que ofrecen servicios de grandes modelos de IA, la emergencia de competidores con costes mucho menores que el resto reduce las expectativas de ingresos de todos los integrantes del sector, afectando negativamente a sus cotizaciones. Además, la menor necesidad de infraestructura hace que la demanda futura de chips no vaya a ser tan grande como se esperaba. Recordemos que hasta hace unos días las grandes tecnológicas competían por hacer centros de datos con el mayor número de chips (Meta se planteaba alcanzar los 350.000 procesadores) y los chips de Nvidia empezaban a utilizarse incluso como colateral de operaciones de financiación, con un mercado de deuda que superaba los 11.000 millones. Esto afecta mucho a la cotización de productores de GPUs, los chips más eficientes para realizar las operaciones de productos matriciales necesarias para el funcionamiento de los LLM, como Nvidia (-17% en Bolsa el lunes) o AMD (-6,5%). El hecho de que no sea necesario utilizar procesadores tan sofisticados para conseguir buenos resultados hace que las expectativas de la demanda de las máquinas de litografía ultraprecisa de AMLs sean menores, haciendo caer también su cotización.

Otro sector también muy afectado es el de centros de datos, que estaba en una expansión exponencial, pues no harán falta infraestructuras tan grandes. Y, por supuesto, el sector de la energía (eléctrica o nuclear) también sufrirá las consecuencias del *shock*, puesto que la expectativa era que sería necesaria una cantidad ingente de energía para hacer funcionar los modelos. Así se ha visto en la cotización de Cameco o de Siemens Energy. En resumen, un terremoto del que solo hemos empezado a ver los primeros temblores y que rediseñará el futuro del sector.

Catedrático de Economía de la UP