

RICARDO BAEZA-YATES,
 DIRECTOR DE INVESTIGACIÓN DEL INSTITUTO
 DE IA EXPERIENCIAL DE LA UNIVERSIDAD
 NORTHEASTERN (EEUU):

“La irrupción de DeepSeek no es un momento Sputnik, sino un pequeño terremoto tecnológico y comercial”



■ Desde Silicon Valley, el ingeniero y Ph.D en Computación, abordó las aristas tecnológicas, comerciales y geopolíticas de la aparición del chatbot chino que sacudió a la industria y a los mercados.

POR RENATO OLMOS

Esta semana, la startup china DeepSeek sacudió los mercados luego que se conociera que su modelo de inteligencia artificial (IA) generativa R1 igualó el rendimiento de sistemas como los de OpenAI y Meta, pero con menores costos y capacidad de cómputo.

Quien ha seguido de cerca este fenómeno es el ingeniero eléctrico y doctor en computación Ricardo Baeza-Yates. Radicado hace años en California, Estados Unidos, es director de investigación del Instituto de IA Experiencial de la Universidad Northeastern en el campus de Silicon Valley.

En entrevista con DF desde California, el también cofundador de la startup Theodora AI, analizó la irrupción del modelo chino desde las perspectivas tecnológica, económica, geopolítica y comercial. Destacó cómo la firma china logró

avances sin acceso a *hardware* de última generación, poniendo en duda el paradigma de que más datos y más recursos computacionales son la única forma de mejorar los modelos de IA y, generando con ello, un terremoto tecnológico y comercial.

- **¿Cómo evalúa la llegada de R1?**
 - Era algo que iba a pasar, pero que no se sabía de dónde iba a venir. La sorpresa es que sea de China y no de Medio Oriente o Europa. El resto del mundo estaba pensando en cómo poder hacer nuevos modelos con menos recursos y más baratos. Y eso fue lo que sucedió.

De hecho, esto pone en duda toda la iniciativa de (Donald) Trump de hacer el proyecto Stargate y los nuevos centros de datos, porque si no necesitamos tantos recursos, ¿para qué vamos a gastar US\$ 500 mil millones? Tal vez lo que deberíamos hacer es tener más talento, que se puede importar de cualquier

parte del mundo.

- **El entrenamiento de R1 habría demandado US\$ 5,6 millones y GPT-4 de OpenAI unos US\$ 100 millones. ¿Son cifras comparables?**

- Es difícil comparar porque esos US\$ 5,6 millones fueron solamente para el entrenamiento final.

Lo que sí es claro es que supieron hacerlo más barato porque estuvieron obligados a innovar, ya que tenían estas restricciones geopolíticas. Eso es muy interesante porque es un excelente ejemplo de lo que siempre menciono: la innovación no tiene nada que ver con la regulación o no. Hoy, parte del problema es que hay mucha innovación mala o que no es ética, legítima y que no respeta el medio ambiente, que gasta recursos sin pensar en las consecuencias.

Lo de China sirve para pensar primero y no tratar de usar la fuerza bruta.

- **¿Es factible crear un modelo**

de igual eficiencia, pero a un costo muy inferior? Se ha puesto en duda la cantidad de tarjetas de Nvidia que usó DeepSeek...

- Habría que replicarlo. Uno no puede llegar y decir que no se puede hacer. Por lo que veo hicieron innovaciones muy interesantes en la parte de ingeniería. El hecho de que el modelo sea en gran parte abierto permite que la gente verifique eso, así que sería muy estúpido que estuvieran mintiendo porque entonces fácilmente los pillarían.

- **En esa línea y de cara a los usuarios, ¿es mejor este modelo de IA generativa que otros en términos de desempeño?**

- Tengo datos anecdóticos. Le pregunté qué sabía de mí y tuvo dos errores, que estaba trabajando en Barcelona, eso pasó hace 10 años, y que había ganado un premio famoso en Europa, que no creo que vaya a ganar nunca, pero alucinaba menos que ChatGPT.

Pareciera que, en cierto sentido, es más sensato y eso se debe a la ingeniería que hicieron al usar el entrenamiento o aprendizaje reforzado que se utiliza para verificar si lo que uno está diciendo es más correcto o no. Estos sistemas no tienen conocimiento de verdad, hacen predicciones, pero sí pueden tratar de asegurarse de que la predicción sea mejor.

Hay ejemplos que he visto que son impresionantes y mejores que el modelo o1 de ChatGPT.

“No es un momento Sputnik”

- **¿Cómo vio la recepción de las grandes tecnológicas y de los usuarios en Silicon Valley?**

- El impacto comercial es el más interesante y tiene dos dimensiones. Uno es el hecho de que ya no se necesita tanta capacidad de cálculo para hacer estas cosas y eso fue lo que más ha afectado a Nvidia que perdió cerca del 20% de su valor de mercado.

Lo otro, que sea abierto, aunque no completamente, sí tenemos la arquitectura y rompe un poco el



CONGRESO FUTURO

“DeepSeek no tiene la capacidad de escalar porque no tiene tantos recursos para tener, por ejemplo, mil millones de usuarios o más, como tiene OpenAI y eso significa que no todas las compañías lo van a poder implementar”.

“No es descabellado pensar que surgirán nuevas startups, pero con un modelo de negocio distinto, no como servicio (SaaS), sino tal vez más concreto de aplicaciones específicas para la economía chilena”.

modelo de negocio de OpenAI y de otras compañías que querían ganar dinero con esto. Conozco mucha gente que dejó de pagar las suscripciones a estos modelos.

Esto no es un momento Sputnik (en referencia a la comparación del inversionista Marc Andreessen con la forma en que la Unión Soviética sorprendió a EEUU al poner el primer satélite en órbita), sino más bien un pequeño terremoto tecnológico y comercial, porque están diciendo que pueden mejorar la tecnología, pero, además, que la regalan. Y todo en un momento en que el nuevo Gobierno de Estados Unidos está haciendo planes de cómo seguir dominando, ahora la pregunta es quién es el dominante.

– OpenAI y Microsoft acusan a DeepSeek de robar datos para entrenar a su modelo. ¿Es factible o solo una estrategia de contraataque?

– Es una mala excusa por dos razones. Una es que OpenAI y otras compañías han usado mucho contenido que tenía derechos de autor sin pedir permiso. Así que, primero, no tienen la ética o la moral para

reclamar que otros hayan robado sus resultados, porque ellos lo han hecho en una escala mayor.

Segundo, si lo hubieran hecho habría sido carísimo, porque tendrían que estar usando ChatGPT y eso es caro. Tendría que agregarse a los US\$ 5,6 millones y podría ser una cantidad incluso más grande porque es algo a gran escala.

– ¿Esto puede dar pie a que compañías como OpenAI y otras de código cerrado sigan en esa lógica más hermética y no abran sus códigos?

– Podría ser, pero en ese caso sería bueno que sean más transparentes y se llamen ClosedAI (IA cerrada) y dejen claro que tienen fines de lucro.

– ¿Qué tan eficiente es DeepSeek respecto de LLaMA, la IA generativa de código abierto de Meta?

– Según Yann LeCun (científico jefe de IA de Meta) usaron muchas cosas que vienen de LLaMA. Son las mismas ideas puestas de otra manera, pero lo que hicieron fue mejorarlas 10 veces.

Por lo menos están siendo transparentes en el sentido de que si usaron código abierto, también

están entregando parte de vuelta y eso le va a servir también a Meta. Estoy seguro que ya tienen todo esto instalado y viendo cómo lo mejoran.

De hecho, están muy preocupados, porque tal como en OpenAI pensaban que eran los líderes, Meta pensaba que era el líder en el tema de código abierto. Ahora tampoco es cierto y son dos carreras distintas en las que fueron sobrepasados.

“DeepSeek no tiene la capacidad de escalar”

– ¿Qué implica para la industria que este modelo sea de bajo costo y abierto?

– Varias cosas. Muchas empresas que tienen la capacidad para usar esto lo van a implementar en sus propios servidores, e incluso hay una versión que funciona en un navegador y celulares. Van a haber muchas aplicaciones que entrarán en operación pronto gracias a este desarrollo. Sin embargo, estoy seguro que DeepSeek no tiene la capacidad de escalar porque no tiene tantos recursos para tener, por ejemplo, mil millones de usuarios o más como tiene OpenAI y eso significa

que no todas las compañías lo van a poder implementar.

– ¿Pero tiene el potencial para cambiar la industria de la inteligencia artificial a nivel global?

– El cambio importante es que le da la razón a Meta, que se puede hacer código abierto y se puede hacer mejor que los cerrados. Destruye el modelo de negocio tradicional.

– Aterrizando la discusión a nivel local, ¿qué impacto podría tener en la región o en Chile?

– Si uno usara las mismas técnicas de DeepSeek para el proyecto que lidera el Centro Nacional de IA para desarrollar un modelo para Latinoamérica, resulta mucho más barato de lo que estaban pensando. Entonces, es mucho más factible y rápido, y tal vez podríamos ver resultados este año con una mayor calidad.

No es descabellado pensar que surgirán nuevas startups, pero con un modelo de negocio distinto, no como servicio (SaaS), sino tal vez más concreto de aplicaciones específicas para la economía chilena.

– ¿Ve algún riesgo en usar R1, entendiendo que puede tener

sesgos o limitaciones de información, dado el control que ejerce el Gobierno chino en el desarrollo tecnológico?

– La mayor parte de la regulación política de China es interna, es decir, no nos afecta mucho, aunque por supuesto no podremos conseguir cosas negativas de China.

Pero lo que sí veo es un problema, y que lo he visto en los otros modelos estadounidenses, es la colonización cultural que uno obtiene sin darse cuenta.

Arista geopolítica

– ¿La aparición de DeepSeek podría cambiar la correlación de fuerzas entre EEUU y China en la hegemonía del desarrollo de la IA?

– Lo que puede cambiar es que, como ya se demostró que se puede hacer, entren más actores como Medio Oriente, que tienen muchos recursos y que han importado mucho talento para sus universidades y centros de investigación.

Va a aumentar el número de actores, reforzar a otros y le da esperanza a otras partes del mundo de que no es necesario tratar de hacer esto por fuerza bruta ni máquinas o más dinero, sino de poner más cerebro.

– ¿Entonces no habrá una carrera por tener la hegemonía?

– Será una carrera permanente, un poco como la Guerra Fría.

– Mencionó que con DeepSeek el proyecto Stargate se pone en duda, ¿por qué?

– Pongo en duda que la solución sea gastar US\$ 500 mil millones en centros de datos cuando eso tiene un impacto ambiental gigantesco. (...) Lo repensaría y ves que hay tantas alternativas buenas para gastar ese dinero, como en educación o salud. Ahora esto demuestra que se puede hacer lo mismo, pero gastando, quizás, US\$ 50 mil millones. De hecho, ahora desplegar el proyecto podría ser más rápido y fácil.

– ¿En qué pie queda Europa frente a esto?

– Creo que se exagera un poco. Es cierto que no se hace tanto, pero tienen cosas importantes como Mistral (startup francesa de IA generativa). Lo que tiene que hacer Europa es seguir el ejemplo de respetar los derechos humanos, tener soluciones más éticas y ser un faro moral del mundo.

A mí no me molesta que no esté en la carrera tecnológica, sino que al revés, yo creo que es bueno que alguien no tenga los incentivos incorrectos, porque el incentivo financiero es muy fuerte.