DeepSeek abre la puerta a la IA física

20 de octubre de 2026, 钛媒体 (TMTPost). Selección: Miguel Acevedo

ia-fisica, modelos-fundacionales, inversion, china, deepseek, limx-dynamics, winverse-ai, tencent, meituan-longzhu, unitree

Benchmarks	DeepSeek V4.1-Flash-1017	DeepSeek V4.1-Flash	Kimi K3	GLM 5.3	GPT 5.6-Sol	Claude Opus 5
HLE (w/o tools)	42.6/44.0*	36.8/39.1*	43.5	42.0*	44.5	**56.3**
Terminal-Bench 4.0	42.3	31.2	12.6	37.9	39.9	**51.8**
DeepSWE v1.1	**77.0**	74.2	67.5	66.9	73.0	74.0
ProgramBench	29.1	20.3	17.5	19.0	23.0	**37.0**
NL2Repo-Bench	68.2	65.4	58.0	58.0	56.8	**75.3**
Automation-Bench V1.0.6	**60.5**	54.8	46.7	48.8	45.8	50.3
Agents' Last Exam	**35.4**	31.8	27.6	28.5	26.7	28.6
Apex Bench v1.1	**72.1**	64.6	47.8	56.6	54.3	65.8
Chartography (w/tools)	83.0	78.9	68.1	-	79.9	**84.0**
BabyVision (w/tools)	92.5	89.6	85.7	-	88.9	**94.1**
ZeroBench-main (w/tools)	51.0	49.0	41.0	-	**53.0**	52.0

*代表仅测试 HLE 的文本子集；我们在对 Terminal-Bench 4.0 进行的测试中使用 Claude Code 框架，DeepSWE 使用 Mini-SWE-V2 框架；ZeroBench-main 使用 Pass@5 评测。

Después de un corte de varias horas, en el círculo de la IA muchos apostaban a que DeepSeek (深度求索) preparaba un movimiento grande. El fin de semana lo confirmó: publicó DeepSeek-V4.1-Flash-1017 y lanzó la consigna de que programar por deseos se cumple con un clic. De escribir código a armar un proyecto, la barrera volvió a bajar.

Si el salto es tan grande, ¿por qué no se llama 4.2? Porque la arquitectura del modelo base no cambió. La actualización insiste en iterar capacidades del modelo que ya existía. Entre las señales oficiales y las pruebas de los usuarios, V4.1 da para más que un número de versión nuevo.

En inversión, DeepSeek (深度求索) tampoco estuvo quieta. En las últimas semanas circularon varias versiones de apuestas en IA física. Por un lado, quiere que la IA termine

mejor las tareas del mundo digital. Por el otro, explora cómo actúa un robot en el mundo real. El mapa se está agrandando.

性价比拉满的“许愿式编程”

Esta actualización casi no dejó aviso previo.

Antes del anuncio formal, algunos usuarios vieron por un momento el nombre “DeepSeek V4.1 Flash-1017” en la lista de modelos. El “1017” coincide con el 17 de octubre, el modo habitual de marcar una instantánea con la fecha.

¿Por qué no 4.2? Porque V4.1 Flash-1017 es una iteración de capacidades sobre la versión actual. La empresa lo resume como una mejora de Pareto entre rendimiento y costo: el rendimiento sube y el costo baja.

Hay que separar bajar el costo de bajar el precio. La forma de llamar a la API y la tarifa no cambian. El desarrollador sigue entrando al modelo nuevo con deepseek-flash. El mecanismo real es otro: la tarea corre menos tiempo y la salida es más corta. Como el modelo grande se cobra por token, una respuesta más ajustada y una llamada más breve reducen el costo combinado.

En DeepSWE v1.1, el Pass@1 pasó de 74,2 a 77,0, por encima de Claude Opus 5 (74,0) y de GPT-5.6 Sol (73,0). Terminal-Bench 4.0 saltó de 31,2 a 42,3. Las dos pruebas miden lo mismo: terminar el trabajo solo.

No es una ventaja en todo. En HLE, el banco de problemas de nivel experto, sacó 42,6, bastante por debajo del 56,3 de Claude Opus 5. La dirección es clara: empujar a fondo la capacidad de ingeniería y no gastar en amplitud pura de conocimiento.

El cambio más importante es que el modelo insiste en construir un proyecto completo desde cero. Según la empresa, la versión nueva se entrenó con aprendizaje por refuerzo basado en autorretroalimentación: escribe el código, lo ejecuta, lo revisa y lo corrige.

Mostró cuatro casos: un modelo 3D de la Ciudad Prohibida, un juego de carreras en la web, una recreación de Portal 2 con el motor Godot y una escena 3D llamada Valley of Ages. Cubren armado de escenas tridimensionales, lógica de juego, uso de un motor e interacción con un entorno. La empresa dice que bastaron pocos prompts, poco material de referencia y rondas cortas de interacción.



Ahí está el sentido de programar por deseos: con muy pocas palabras, ayudar a generar un resultado cercano a lo que el usuario tenía en la cabeza.

Un desarrollador pidió convertir un recorte de papel en un pueblo costero. El modelo tenía que construir la escena web en código, hacer que las casas se levantaran con el desplazamiento de la página, y que el faro, el muelle y las olas de papel aparecieran en orden, con la cámara cambiando de ángulo. En la prueba, después de levantarse las casas se notan el grosor del papel, los pliegues y la sombra. Al mover el control de tiempo, el cielo, el mar y las luces de las ventanas cambian entre el día y la noche.

La fluidez es lo que más se elogia. En V2EX, un desarrollador dijo que la versión vieja "parecía haber perdido inteligencia" y que con 1017 "volvió a la normalidad". En el foro de Trae, alguien describió la versión anterior como una conversación entre un pollo y un pato, y dijo que la nueva daba "la ilusión de que hasta el propio coeficiente intelectual había subido". Un usuario de Zhihu armó Mario Kart con las dos versiones: en la vieja la pista salía torcida y el auto no se movía; en la nueva la imagen es más cuidada y el control, más fluido.

También hay quien enfría el entusiasmo. En Reddit, un empleado de Nebius que trabaja en agentes de programación y en evaluación probó la versión nueva con 10 tareas de programación de su trabajo diario. Repitió cada tarea 3 veces y revisó el resultado con pruebas ocultas.



En esa prueba personal, V4.1-Flash-1017 con el marco de programación Pi y una intensidad alta de razonamiento aprobó las 30 corridas. El costo medio por tarea fue de unos 4,5 centavos de dólar. En otra prueba, con la intensidad máxima de razonamiento, la V4.1-Flash anterior aprobó 24 de 30 corridas.

El mismo evaluador informó que la versión nueva consume unos 550.000 tokens por tarea. Aunque una llamada salga barata, ese volumen recuerda algo: que el modelo responda más rápido no significa que todo el flujo de trabajo gaste menos recursos.

Otro usuario intentó generar un cortometraje de 15 segundos con aspecto de cine. V4.1-Flash-1017 costó 2,5 dólares y tardó 4 horas. Claude Opus costó 54,2 dólares y tardó 7 horas. Lo que esta actualización acorta es la distancia entre “suficientemente fuerte” y “suficientemente barato”.

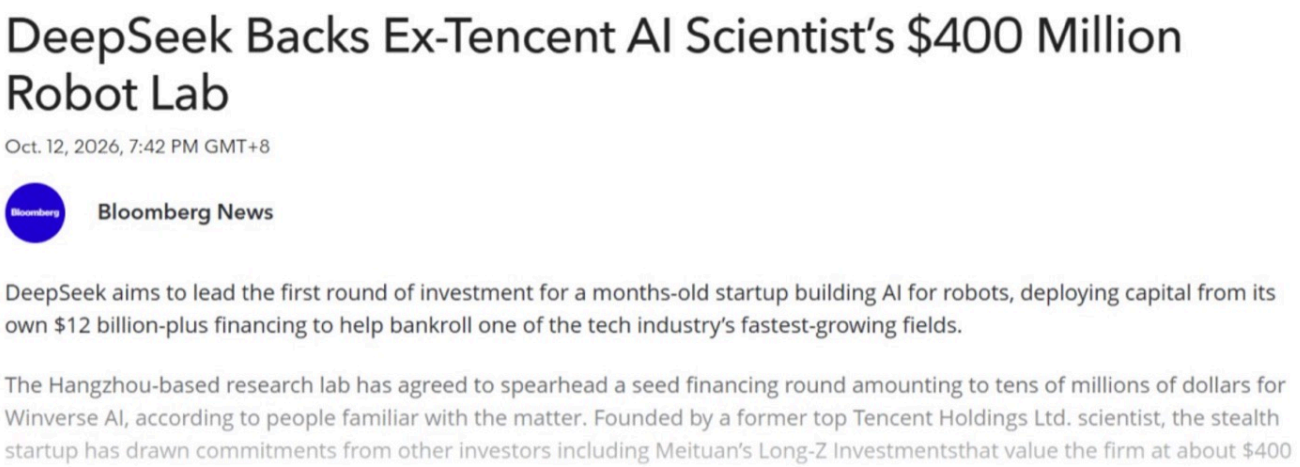


梁文锋的AGI雄心

La actualización del modelo fue una fiesta en el círculo de la IA, pero no conviene perder de vista el movimiento de capital. En corto: DeepSeek (深度求索) está comprando, para el modelo fundacional que viene, una entrada al mundo físico.

El 18 de octubre, versiones del sector dijeron que DeepSeek (深度求索) planea invertir en la startup de Yu Tao, exresponsable del modelo fundacional de humanoides en LimX Dynamics (逐际动力). El rumbo es modelos de lenguaje más IA física. Las dos rondas se habrían valuado en unos 70 millones y 200 millones de dólares.

El 12 de octubre, según los reportes, DeepSeek (深度求索) planea liderar la inversión en Winverse AI, fundada por Han Hu, excientífico de Tencent (腾讯). La valuación posterior a la inversión rondaría los 400 millones de dólares. Meituan Longzhu (美团龙珠) también figura entre los participantes.



DeepSeek Backs Ex-Tencent AI Scientist's $400 Million Robot Lab

Oct. 12, 2026, 7:42 PM GMT+8

Bloomberg News

DeepSeek aims to lead the first round of investment for a months-old startup building AI for robots, deploying capital from its own $12 billion-plus financing to help bankroll one of the tech industry's fastest-growing fields.

The Hangzhou-based research lab has agreed to spearhead a seed financing round amounting to tens of millions of dollars for Winverse AI, according to people familiar with the matter. Founded by a former top Tencent Holdings Ltd. scientist, the stealth startup has drawn commitments from other investors including Meituan's Long-Z Investmentsthat value the firm at about $400

Antes, en agosto, DeepSeek (深度求索) aportó unos 141 millones de yuanes a la colocación estratégica de la salida a bolsa de Unitree (宇树科技), con un período de bloqueo de 36 meses, y firmó un memorando de desarrollo conjunto.

Desde el lado de las empresas de IA física, las tres apuestas compran tres capas distintas de la pila técnica.

Unitree (宇树科技) pone el cuerpo: tiene muchos robots reales, y esa es una entrada escasa a datos físicos de verdad.

Han Hu representa el cerebro y los ojos: es autor principal de Swin Transformer y ganador del premio Marr del ICCV 2021. Su investigación fue de los modelos fundacionales de visión a los modelos de visión y lenguaje (VLM) y a los modelos fundacionales de IA física. El problema que ataca es cómo un robot entiende lo que ve.

Yu Tao está más cerca del cerebelo y de los nervios motores: en LimX Dynamics (逐际动力) armó desde cero el sistema de imitación de movimiento y de seguimiento de cuerpo completo System 0. Han Hu camina del cerebro hacia el cuerpo. Yu Tao camina del cuerpo hacia el cerebro. Las dos rutas se encuentran.

Liang Wenfeng ya había dejado la pista. Su hoja de ruta pública hacia la AGI tiene seis escalones: modelo de lenguaje, cadena de pensamiento, agente, aprendizaje continuo, autoiteración e IA física. El paso hacia la IA física se está acelerando.

Durante un tiempo, el círculo de los modelos grandes dio por sentado que un VLA era difícil de hacer general. En septiembre, la aparición de GPT-6 Astra mostró que se puede construir un modelo de IA física capaz de generalizar sin ser especialista en IA física. Al mismo tiempo, la capacidad de programación ya se ve cerca del techo. Las empresas de modelos grandes necesitan otro frente donde poner el próximo golpe, y entrar a la IA física se volvió difícil de evitar.

El problema es que la IA física es un trabajo duro, de software acoplado a hardware. Una empresa de modelos grandes tiene escenas y datos limitados. Por eso DeepSeek (深度求索) invierte primero en cerebro, cerebelo y cuerpo, en lugar de fabricar ella misma el robot completo.

La pista está al rojo vivo. Según recuentos de medios, solo en el primer semestre el financiamiento interno superó los 90 mil millones de yuanes, cinco veces el mismo período de 2025. En el segundo semestre el viento cambió: cayeron con fuerza los proyectos que las firmas llevan a la revisión de salida a bolsa, y a nivel estatal se advirtió a la industria robótica que evite el seguidismo ciego y las estampidas.

En ese punto, DeepSeek (深度求索) aparece como nuevo comprador e integrador, y acelera una división entre pocos cerebros y muchos cuerpos. Se ve con claridad que está extendiendo el mapa del mundo digital al mundo físico.

Las vacantes ya lo insinuaban. La descripción del puesto de investigador en comprensión multimodal dice que la AGI no se va a quedar en el texto, y que un eslabón clave es dar al modelo la capacidad de entender el mundo físico.

La competencia de hoy sigue en las tablas de puntaje y en la guerra de precios de las API. La siguiente va a estar en la línea de producción, en el dormitorio y en la calle.

Fuente: 钛媒体 (TMTPost)

roboticaasialatam.com/2026-10-20-deepseek-83be4b/