Unitree UnifoLM-WLA-1.0: un modelo de 6B para 64 tareas humanoides

11 de septiembre de 2026, Robots Daily. Selección: Miguel Acevedo

unitree-unifolm, china, modelo-fundacional-de-humanoides, modelo-fundacional-unitree, ia-fisica, unitree



Unitree (宇树科技) publicó nuevos detalles sobre **UnifoLM-WLA-1.0**, su más reciente **modelo fundacional para robots humanoides** de propósito general. Según la compañía, el modelo reúne la manipulación de precisión sobre mesa y la **manipulación móvil de cuerpo completo** en un solo sistema, usando un **espacio de acción unificado** para manejar distintos efectores finales. El objetivo de Unitree (宇树科技) es usar un solo modelo para toda la cadena, desde la percepción y las decisiones de interacción hasta la generación de acciones.

El modelo tiene **6,000 millones de parámetros**. Unitree (宇树科技) dice que fue entrenado con unas 2,500 horas de datos de alta calidad de robots reales. Las pruebas con robots reales de la compañía cubren 64 tareas: 10 tareas de manipulación de cuerpo completo y 54 tareas sobre mesa. Es compatible con pinzas paralelas y dos tipos de manos diestras.

Para los desarrolladores de robots, la cifra de 64 tareas quizá no sea lo más importante. La pregunta más grande es si el modelo puede seguir funcionando en distintos **cuerpos de robot** y efectores finales, y qué tan bien generaliza a tareas para las que no fue entrenado. Todavía no hay validación externa. Unitree (宇树科技) lanzó una página de proyecto y dice que publicará el modelo, el código y los conjuntos de datos. Pero las páginas de Code, Models

y Datasets todavía dicen “Próximamente”. Los pesos y los conjuntos de datos no están disponibles para descargar.

64 tareas, pero lo difícil es la coordinación

Unitree (宇树科技) enumera tareas como sacar la basura, meter ropa en la lavadora, organizar zapatos, limpiar baños y cocinas, tender camas, recoger verduras, ordenar sofás, doblar toallas y ropa, guardar platos y cargar baterías. Diez son de manipulación de cuerpo completo; 54 son sobre mesa.

La diferencia no es solo el tamaño del espacio de trabajo. Las tareas sobre mesa a menudo pueden hacerse desde una posición relativamente fija, con el robot controlando principalmente sus brazos, manos y efectores finales. En una cocina, un baño o una recámara, el robot primero tiene que cambiar su posición y postura, y luego llevar ambos brazos al espacio de trabajo correcto. Tareas como tender una cama u ordenar un sofá requieren que las manos y el cuerpo se muevan juntos.

El enfoque de Unitree (宇树科技) es que un solo modelo maneje ambos tipos de acción en lugar de construir modelos separados para la manipulación sobre mesa y la de cuerpo completo. El modelo tampoco está atado a una sola pinza. Unitree (宇树科技) dice que UnifoLM-WLA-1.0 funciona con pinzas paralelas y dos tipos de manos diestras.

Un espacio de acción para manos, brazos y parte inferior del cuerpo

Para controlar todo el cuerpo de un humanoide con un solo modelo, el primer problema es cómo representar las acciones. Unitree (宇树科技) divide las acciones en tres partes: pose del efector final, articulaciones del efector final y articulaciones de la parte inferior del cuerpo. Luego usa **cuantización vectorial residual, o RVQ**, para convertir acciones continuas en tokens discretos. Los tokens de distintas partes del cuerpo se alinean en pasos de tiempo compartidos y se alimentan juntos al modelo, para que pueda aprender cómo trabajan juntas las manos, los efectores finales y la parte inferior del cuerpo.

El resultado es una **representación de acción discreta** que el modelo puede aprender y predecir de manera unificada. El punto no es una acción individual, sino la coordinación entre acciones. Caminar hacia una lavadora, por ejemplo, no es simplemente una acción de “moverse” seguida de una acción de “agarrar”. El cuerpo tiene que llegar a la posición correcta, los brazos tienen que ajustarse y las manos tienen que operar según el objeto objetivo. Poner estas acciones en la misma línea de tiempo permite al modelo aprender cómo se relacionan entre sí.

Antes de actuar, el robot tiene que saber qué va a cambiar

Otra decisión de diseño en UnifoLM-WLA-1.0 es que no se detiene en entender la imagen actual. Si el robot está a punto de tomar un plato, necesita más que "hay un plato aquí". También necesita saber cómo cambiará la escena después de retirar el plato, cómo se deformará una toalla al doblarla y qué partes de la escena cambiarán realmente cuando se empuje un objeto. Unitree (宇树科技) llama a estas áreas **Regiones Dinámicas**.

El método usa flujo óptico para extraer regiones cambiantes de pares de fotogramas de video, y luego usa un autoencoder variacional con cuantización vectorial (VQ-VAE) para comprimir cambios continuos en tokens discretos de longitud fija. El modelo puede entonces predecir futuras regiones de cambio con base en la imagen actual, la descripción de la tarea o las condiciones de la acción. Unitree (宇树科技) llama a esto **modelado de mundo centrado en la interacción**. La idea es predecir cómo cambiará la escena cuando el robot interactúe con ella, no solo describir el fotograma actual.

UnifoLM-ER-Flow une visión y acción

En este marco, Unitree (宇树科技) coloca tokens visuales, de lenguaje, de regiones dinámicas y de acción en un modelo multimodal unificado llamado **UnifoLM-ER-Flow**. UnifoLM-WLA-1.0 agrega encima un módulo **MMDiT Action Expert** para generar acciones continuas del robot. Unitree (宇树科技) dice que el sistema combina razonamiento de IA física, predicción de regiones dinámicas futuras y aprendizaje de acciones discretas en un solo marco multimodal.

El pipeline tiene tres partes conectadas: entender el espacio y la tarea, predecir cómo la interacción puede cambiar las cosas y luego generar acciones del robot.

2,500 horas de datos de robots reales

UnifoLM-WLA-1.0 fue entrenado con unas 2,500 horas de datos de robots reales. Unitree (宇树科技) dice que los datos provienen de sus propios conjuntos de datos de código abierto y de recursos públicos como HIW-500, y cubren múltiples cuerpos de robot y entornos de operación. El modelo también usa un espacio de acción unificado y priors de transferencia entre cuerpos.

Esa distinción importa al llamar "**modelo fundacional**" a un modelo de robot. Un modelo que solo completa un conjunto de tareas en un robot fijo, con un efector final fijo y una escena fija, está más cerca de un modelo de control especializado. Un modelo fundacional debería conservar cierta capacidad de transferencia cuando cambia el hardware o la tarea. Unitree (宇树科技) incluyó distintos cuerpos de robot, dos tipos de manos diestras y pinzas paralelas en su marco de entrenamiento y pruebas. Pero la información pública no alcanza para juzgar qué tan bien transfiere a robots de terceros. Será más fácil evaluar qué tan bien transfiere a hardware de terceros una vez que la publicación esté disponible.

5 millones de muestras de razonamiento de IA física de UnifoLM-ER-1-4B

Detrás del modelo de acción hay un modelo de razonamiento de IA física separado. **UnifoLM-ER-1-4B** de Unitree (宇树科技) está basado en **Qwen3-VL-4B** y usa más de 5 millones de **muestras de razonamiento de IA física**. Los datos cubren predicción de puntos en imágenes, detección de objetos, razonamiento multiimagen, predicción de trayectorias 2D, detección de objetos 3D y respuesta a preguntas espaciales multiimagen, y se entrenan conjuntamente con datos generales de imagen y texto.

Este modelo trata principalmente sobre la **comprensión espacial** antes de la acción. El robot necesita saber dónde están los objetos, cómo se relacionan entre sí, dónde está la ubicación objetivo y qué partes de la escena pueden verse afectadas por sus acciones. Esa información le da al modelo de acción el contexto espacial que necesita.

Embodied Reasoning Benchmark Results

MODEL	OPEN SOURCE	ROBOVQA	EGO-PLAN2	REF SPATIAL BENCH	WHERE2 PLACE	PIXMO-POINT	SPATIAL UNDERSTANDING: BLINK	SPATIAL UNDERSTANDING: CV-BENCH	SPATIAL UNDERSTANDING: EMB SPATIAL
UnifoLM-ER-1-4B	Yes	62.4	55.1	**61.7**	**82.0**	**73.8**	**93.4**†	88.6	**88.9**
RoboBrain2.0-7B*	Yes	30.0	32.23	42.2	63.6	54.7	83.9	88.7	76.3
Robix-7B*	No	**63.6**	/	/	41.9	29.5	87.6	86.5	77.4
Pelican-7B*	Yes	31.8	33.7	22.3	57.3	20.4	/	79.4	73.2
Cosmos-R1-7B*	Yes	38.8	26.0	5.6	2.9	8.2	/	76.7	68.9
Cosmos3-Super-64B*	Yes	/	/	57.0	71.0	/	90.3′	88.0	/
Qwen3-VL-4B*	Yes	47.7	40.7	46.6	63.0	48.3	85.0′	85.1	79.6
Qwen3-VL-8B*	Yes	43.3	49.7	54.2	63.9	51.0	73.8′	86.2	78.5
Embodied-R1-3B*	Yes	51.8	26.5	39.7	69.5	49.4	78.5′	82.7	67.4
Embodied-R1.5-8B*	Yes	61.0	53.8	54.2	74.0	64.8	83.0′	86.9	78.1
Molmo2-ER-4B*	Yes	/	/	52.5	54.0	/	85.7′	87.8	78.8
Hy-Embodied-VLM-1.0* 30B-A3B	Yes	/	49.6	53.4	65.0	64.6	87.3′	**89.7**	82.7
MiMo-Emb-7B*	Yes	62.0	43.0	48.0	63.6	42.35	81.3	88.2	76.2
Thinker-4B*	Yes	62.7	**63.7**	61.0	72.0	57.4	84.6	86.3	80.2
Wall-OSS-0.5-3B*	Yes	/	/	/	15.0	/	/	/	/
Lumo-1-Stage1-7B*	No	/	/	51.0	69.1	/	82.4	86.4	75.6
Gemini-ER 2†	No	/	/	35.4	/	/	90.6′	**90.4**	81.4
Gemini-ER 1.5†	No	/	/	41.8	48.0	/	/	83.6	73.4
Gemini 2.5 Pro†	No	/	/	33.6	37.0	/	88.6′	85.9	78.0
Gemini 2.5 Flash†	No	/	/	41.2	48.0	/	80.3′	85.5	76.2
Gemini 3.1 Pro†	No	/	/	70.0	61.0	/	86.1′	88.6	/
GPT-5.6-sol‡	No	/	/	58.3	51.1	/	85.6′	85.2	80.7
GPT-6-Astra†	No	/	/	**79.6**	69.0	/	90.4′	87.3	83.3

* Results are sourced from the models' official technical reports or publicly available papers.
‡ Results were obtained through tests using the models' official APIs.
† BLINK scores are averaged over the Relative Depth and Spatial Relation subtasks only; all reported results were obtained in our own testing.

Unitree (宇树科技) reporta las mejores puntuaciones en siete benchmarks

Unitree (宇树科技) publicó resultados en 16 **benchmarks de percepción y comprensión multimodal**. UnifoLM-ER-1-4B obtuvo la puntuación más alta entre los

modelos de código abierto evaluados en siete: RefSpatial-Bench, Where2Place, PixMo-Point, BLINK, EmbSpatial, RoboSpatial y VSR.

Las puntuaciones publicadas incluyen:

- **Where2Place:** 82.0
- **PixMo-Point:** 73.8
- **EmbSpatial:** 88.9
- **RoboSpatial:** 73.1

Unitree (宇树科技) también compartió comparaciones con algunos modelos de código cerrado. En pruebas que ambos sistemas realizaron, UnifoLM-ER-1-4B obtuvo 82.0 en Where2Place, 13 puntos más que **GPT-6 Astra**; 93.4 en BLINK, 3 puntos más; 88.6 en CV-Bench, 1.3 puntos más; y 88.9 en EmbSpatial, 5.6 puntos más.

Esas cifras necesitan contexto. Son resultados de benchmarks publicados por Unitree (宇树科技), no clasificaciones de una organización de pruebas independiente. El material disponible no incluye reproducciones de terceros. Los resultados de benchmarks no muestran por sí solos que UnifoLM-WLA-1.0 supere a otros modelos en tareas de robots reales. La prueba más sólida serán las tasas de éxito en robots reales, los tipos de fallo y si el modelo puede transferir a otros robots sin reentrenamiento dirigido.

Plan de código abierto anunciado, pero los pesos aún no llegan

La página del proyecto UnifoLM-WLA-1.0 está en línea y Unitree (宇树科技) dice que publicará el modelo, el código y los conjuntos de datos. Por ahora, las secciones de Code, Models y Datasets del sitio siguen marcadas como “Próximamente”. El código fuente, los pesos del modelo y los conjuntos de datos no se han publicado oficialmente para descargar.

Eso significa que los desarrolladores pueden leer sobre la arquitectura, la cobertura de tareas y los resultados de benchmarks, pero todavía no pueden descargar el paquete completo de código abierto ni reproducir el trabajo de manera independiente. La verdadera prueba llegará cuando investigadores externos puedan correr el modelo en tareas y plataformas de robots más allá de la evaluación publicada por Unitree (宇树科技).

Página del proyecto: https://unigen-x.github.io/unifolm-wla.github.io/

Fuente: Robots Daily

roboticaasialatam.com/2026-09-11-unitree-unifolm-wla-1-0-6b-model-for-64-humanoid-robot-tasks-f9f6cc/