Apple corre el humanoide y atrasa el robot de casa

10 de noviembre de 2026, Robots Daily. Selección: Miguel Acevedo

humanoides, investigacion, robot-personal, china, unitree, nvidia, alibaba



En Cupertino nadie había escaneado la habitación. Nadie había caminado el robot por ahí, ni lo había manejado a distancia, ni había pegado una marca en el piso. Un Unitree G1 estaba parado en una alfombra que no tenía en la memoria, con una Mac mini atada a la espalda y un lente de ojo de pez en un soporte hecho a medida, en la cabeza.

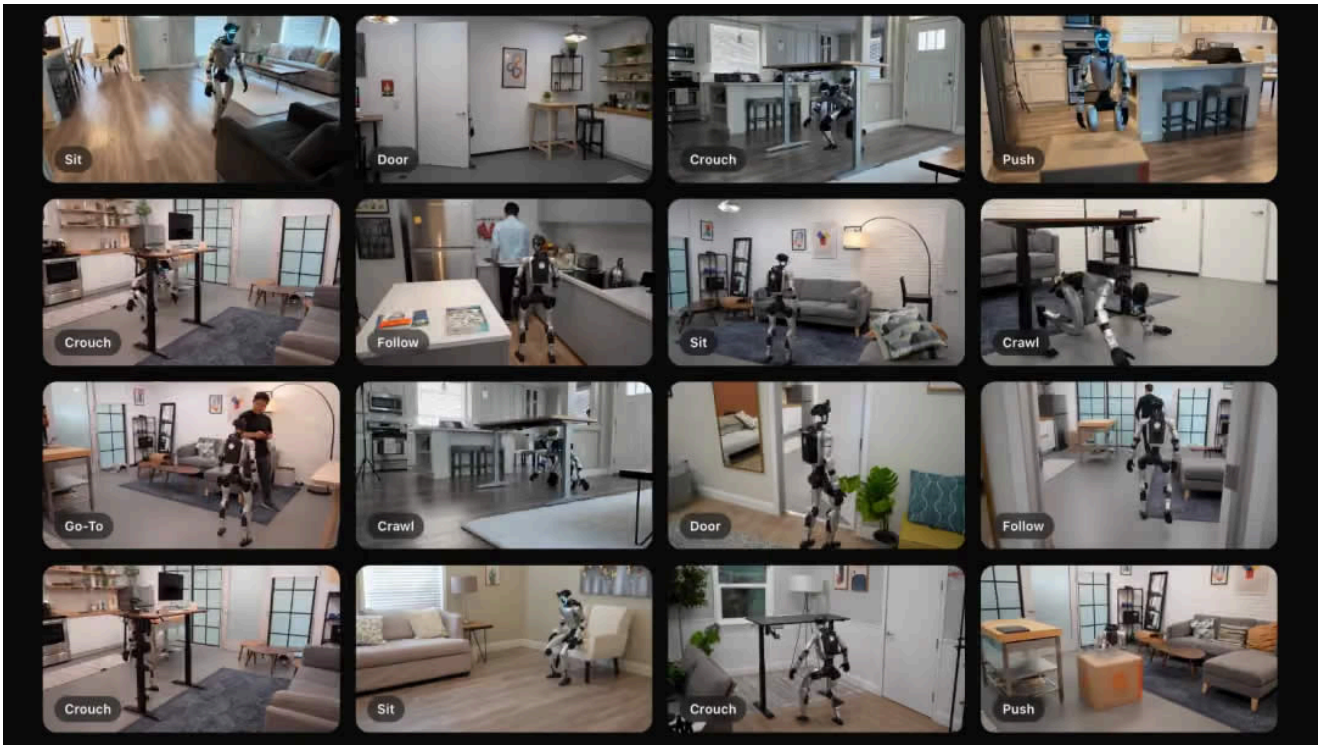
Alguien dijo: ve a sentarte en la silla.

La máquina giró, cruzó la habitación y se sentó. Los investigadores de Apple corrieron esa tarea y otras cinco, 360 veces, en seis casas y oficinas donde el robot no había entrado. Funcionó el 91% de las veces. Ningún dato de entrenamiento salió de una persona con traje de captura de movimiento, ni de alguien con un control en una habitación real.

HALO, Humanoid Autonomy, Learned in simulation, running Onboard, es el nombre que Apple le puso al sistema en el paper Perceptive Humanoid Motion from Simulation. Ocho autores, todos de Apple. Siyuan Hu es el primero. Vladlen Koltun, el último. Koltun dirigió durante años el Intelligent Systems Lab de Intel, antes de que Intel lo cerrara. Es la pieza de IA física más convincente que Apple mostró en público. No es un producto.

Este mes cayeron dos documentos, con 48 horas de diferencia, apuntando para lados opuestos. El 4 de noviembre el grupo de investigación de aprendizaje automático de Apple publicó el paper de esa demo. El 2 de noviembre, The Information reportó que el robot que

Apple sí quiere vender, un aparato de mesa de US$1.000 con el código J595, se corrió de 2027 a tan tarde como 2028. No lo trabó el hardware. Lo trabó el software.



Qué resolvió HALO

Todo laboratorio de humanoides choca con la misma pared, y no son los motores ni las baterías. Es la supervisión. Para que una máquina pase de una imagen y una frase a comandos de articulaciones hacen falta ejemplos pareados de qué hacer. Nadie tiene un corpus de eso a escala. Las dos formas de armarlo son la teleoperación y la captura con un wearable. El paper las despacha en el abstract: hacen falta al menos una hora de tiempo humano para producir una hora de datos.

Ese ratio es la razón de que el progreso de los humanoides haya sido lento y caro. Es lo que HALO ataca.

El pipeline arranca con Kimodo, el modelo de difusión de movimiento cinemático que NVIDIA Research publicó en marzo, entrenado con 700 horas de captura óptica. Con descripciones de texto, waypoints y algunos fotogramas clave, Kimodo genera movimiento de cuerpo entero. Esos movimientos caen en habitaciones generadas por Infinigen, la herramienta de escenas de procedimiento del laboratorio de visión de Princeton: 1.000 habitaciones, 45% living, 30% oficinas, 25% dormitorios. Cada una se renderiza en Isaac Sim, de NVIDIA, con ray tracing, desde un punto de vista que coincide con la cámara de ojo de pez del robot. Después se degrada a propósito, con desenfoque, ruido, exposición y cambios de color al azar, para cerrar la brecha entre la simulación y un living real a las cuatro de la tarde.

Salieron 162.698 episodios, o 296 horas de datos de entrenamiento a 30 cuadros por segundo, en seis habilidades. El resumen de Apple cabe en una frase: el corpus escala con cómputo, no con trabajo humano.

Es un reclamo de economía tanto como de robótica. Apple dice que el insumo escaso del desarrollo de humanoides se puede comprar a un proveedor de nube, en vez de contratarlo.

Tres modelos, tres relojes

HALO no es un modelo. Son tres, a velocidades muy distintas. El corte es el diseño.

Un orquestador toma el pedido, “ve a sentarte en la silla”, y lo parte en subtareas. Decide cuándo seguir y cuándo reintentar. El modelo es de estantería: el default de Apple es uno de 35.000 millones de parámetros de la familia Qwen, de Alibaba, cuantizado a 4 bits para que entre. Corre más o menos una vez cada dos segundos, lo más rápido que un modelo de ese tamaño puede ir en una mochila.

Un generador de movimiento corre unas tres veces por segundo. Junta un codificador de visión Qwen congelado con un transformador de difusión de 49 millones de parámetros. Lee el cuadro actual de la cámara y la subtarea actual, y planea cinco segundos de movimiento de cuerpo entero.

Un tracker corre a 50 Hz. Lo entrenaron con aprendizaje por refuerzo en Isaac Lab. Consume imágenes de profundidad, propiocepción y la IMU del robot, y saca objetivos de posición para 29 articulaciones.

El corte de cómputo importa. Los modelos corren en el robot, pero no en nada que haya puesto Unitree. Apple atornilló una Mac mini con M4 Pro y 64 GB de memoria en una caja, en la espalda del G1. Esa máquina corre el orquestador y el generador. El tracker corre en el Jetson Orin NX del propio G1. El robot no va atado. También carga una computadora de escritorio.

Qué no cubre el 91%

Por separado, las seis habilidades aguantan. Agacharse bajo una mesa funcionó el 95% de las veces en simulación y el 100% en el mundo real. Pasar de costado por una puerta entreabierta, 91 y 93. Caminar hasta un objeto nombrado, 98 y 92. Sentarse en una silla es la más floja, 84 y 85. El promedio de las seis queda en 91% en los dos lados.

Encadenarlas es donde se cae. Pasar una puerta y después ir a un objeto: 10 de 10. Agacharse bajo una mesa y después sentarse: 60%. Agacharse, caminar hasta una caja y empujarla: 50%. Esas pruebas usaron un solo entorno real. Los autores son honestos: la caída es más fuerte de lo que predecirían las tasas de una sola tarea.

Los huecos van más hondo. Ninguna de las seis habilidades agarra algo con las manos. El robot no tiene tacto. El tracker, como dice el paper, sigue el plan cinemático sin saber para qué es el plan: va a ejecutar un plan que ya salió mal. Y sumar una séptima habilidad no es cuestión de más cómputo. Alguien tiene que escribir los prompts y armar los fotogramas clave. El repertorio crece solo tan rápido como una persona puede especificar tareas nuevas.

Tampoco los datos son tan libres de humanos como dice el titular. La primera etapa de entrenamiento del tracker usa LAFAN1, un dataset público de captura de movimiento, redirigido al G1. Los fotogramas de agacharse y gatear salieron de un video de celular de una persona pasando bajo una mesa, levantado a 3D. El set de empujar la caja incluye una grabación de captura. Lo que HALO eliminó es el impuesto de una hora de teleoperación por cada hora de datos, no el insumo humano.

Las ablaciones son donde están las lecciones útiles. Entrenar el generador en una habitación en vez de 1.000 baja la tarea de sentarse al 43% en habitaciones simuladas que no vio. Cambiar el decodificador de difusión por una regresión simple: 19%. Sacar la profundidad del tracker: 74%. Apple publicó en GitHub el pipeline de datos, el código de entrenamiento y el de despliegue, para MuJoCo y para el G1 real. No publicó los pesos entrenados ni el dataset de 296 horas. No corrió una comparación cabeza a cabeza contra otros sistemas publicados, en condiciones parejas. Es lo que un lector escéptico más querría.

El aporte de Apple es el armado, no las piezas

Mirá lo que hay adentro del resultado. Un cuerpo chino. Un modelo de movimiento de NVIDIA. Un simulador de NVIDIA. Un orquestador de Alibaba. Un generador de escenas de un laboratorio universitario de Nueva Jersey. Una cámara de commodity en un soporte impreso en 3D.

El aporte de Apple es el pipeline que los conecta, la decisión de congelar el codificador de visión y entrenar solo las partes chicas, la cuantización que metió un modelo de 35.000 millones de parámetros en una mochila a batería, y el criterio de elegir un robot que se puede pedir online en vez de fabricar uno. Eso último no es pereza. Usar un G1 de stock saca el hardware como variable experimental. Es lo que se hace cuando la pregunta es de software.

Algo cambió, igual. Durante veinte años la ventaja de Apple en cualquier categoría era que era dueña de la pieza que definía: el chip, el sistema operativo, el diseño industrial, el canal de venta. Acá no es dueña de ninguna. Es dueña de la integración, y la hizo lo bastante bien para un resultado de punta. Es una empresa distinta de la que armó el iPhone. Una que puede sacar un paper brillante y un producto tarde en las mismas dos semanas.

Por qué el J595 se corrió a 2028

El producto es una pantalla de nueve pulgadas en un brazo que gira. Se da vuelta hacia quien habla, hace FaceTime, y prende las luces cuando alguien entra a la habitación. El precio

objetivo ronda los US$1.000.

Viene de lejos. Había un prototipo en 2018. Tim Cook vio una demo en 2019 y dejó que el equipo siguiera. El proyecto se aprobó formalmente en 2022. Cuando Apple mató el auto, en febrero de 2024, el grupo de robótica salió de la organización de IA y pasó a John Ternus, vicepresidente senior de ingeniería de hardware. Recién ahí dejó de ser un proyecto lateral y pasó a prioridad estratégica.

El hardware está congelado hace unos dos años. El software no. Según The Information, la discusión interna no es si el robot puede funcionar. Es qué debería poder hacer la primera versión. El equipo de hardware quería sacar el producto. El de software fue achicando el alcance, de vuelta a FaceTime y control básico de la casa, porque la integración más profunda con Siri nunca llegó. El hueco de IA de Apple, discutido en general durante tres años, acá tiene un costo concreto: es la razón de que una máquina que debía estar a la venta el año que viene pueda llegar en 2028.

Una fecha corrida no es nueva en Apple. Sí lo es un producto cuya lista de funciones la escribió el límite de su propio equipo de IA.

El producto de US$299 que hace todo menos moverse

El viernes Apple empieza a vender HomeView. El hardware es una pantalla de siete pulgadas sobre una base de parlante, a US$299, para la mesa o la pared. Lo único que no puede hacer es girar hacia vos.

La pregunta comercial que nadie en Apple respondió en público: por qué un hogar pagaría tres veces más por un HomeView que gira.

La respuesta interna, según lo reportado, es que el J595 es primero una terminal de recolección de datos, y después un producto. Una pantalla que se mueve produce parejas que una pantalla fija no puede. Aprende dónde se sienta una persona respecto de adónde apunta. Aprende qué movimiento físico sigue a qué pedido hablado, y qué cambio en una habitación debería hacerla actuar sin que se lo pidan. Esas parejas son la materia prima de lo que Apple arme después. No se pueden generar en un simulador: el punto de los últimos dos años de investigación es que los humanos simulados no son los reales.

Es un argumento coherente. También es el que da toda empresa cuando el producto llega tarde. Se va a probar si alguien lo compra. Mil dólares es un precio raro para un aparato de recolección de datos que, además, tiene que ser una buena máquina de FaceTime en una cocina donde ya hay una de US$299.

Diez años de robótica, cada intento más chico

Apple lleva en esto más tiempo que la mayoría de las empresas que hoy venden humanoides. El Project Titan arrancó en 2014 con meta de autonomía completa, corrió una década y terminó en febrero de 2024 sin un vehículo. Lo que sobrevivió no fue un producto. Fue un conjunto de capacidades: visión por computadora, navegación autónoma, control de movimiento, fusión de sensores. Los ingenieros que las tenían pasaron, en su mayoría, a robótica.

Un año después Apple publicó ELEGNT, un marco para lo que sus autores llamaron diseño de movimiento expresivo y funcional en robots que no tienen forma humana. La demo era una lámpara de escritorio en un brazo, al estilo de la lámpara de Pixar: se inclinaba cuando tenía curiosidad, se echaba atrás ante una pregunta difícil, empujaba hacia una persona para sugerirle que tomara agua. Era fácil archivarlo como un estudio de diseño. No lo era. ELEGNT y J595 hacen la misma pregunta: si el primer trabajo de un robot es parecer humano, o hacer lo que una pantalla no puede. Apple la respondió igual las dos veces.

El patrón es consistente: primero la investigación, el producto tarde. Titan guardó capacidades. ELEGNT guardó un lenguaje de interacción. HALO guarda un método de entrenamiento. El J595 es el primer intento de gastar algo de eso, y el gasto es cauteloso: un precio bajo, y una forma que no se puede caer.

La habitación, otra vez

El G1 sigue parado en ese living, esperando la próxima instrucción. El cuerpo salió de Hangzhou. El modelo de movimiento, de Santa Clara. El orquestador también salió de Hangzhou, achicado a 4 bits por ingenieros de Apple para que entrara en una mochila. Las habitaciones de entrenamiento salieron de Nueva Jersey. El código salió de Cupertino, y ahora está en GitHub para quien lo quiera bajar.

Lo que aportó Apple es el criterio de cómo juntar esas piezas. Por la evidencia del paper, ese criterio es muy bueno.

Apple puede hacer inteligente a un robot. Lo probó en 360 ensayos. Si puede meter uno en una caja y en una góndola antes de que las empresas que ya venden el cuerpo, los motores y las articulaciones resuelvan el resto, es otro problema. Por US$13.500, Unitree vende el mismo cuerpo que usaron los investigadores de Apple. El robot de Apple, una pantalla en un brazo que gira, va a costar mil dólares y no se espera antes de 2028.

La brecha entre esos dos hechos no es un problema de tecnología. Es un problema de sacar el producto. Y en eso Apple solía ser la mejor del mundo.

Fuente: Robots Daily

roboticaasialatam.com/2026-11-10-apple-s-1-000-home-robot-is-delayed-to-2028-its-humanoid-res-3bf8e7/