



La IA necesita un cuerpo. La era de la inteligencia física

Descripción

Introducción

Hace unos años, ver a un robot bípedo caminar sin caerse ya parecía extraordinario. Después aprendieron a correr, saltar, levantarse del suelo e incluso realizar movimientos que hasta entonces asociábamos casi exclusivamente con la coordinación humana. Hoy los vídeos de robots bailando, practicando artes marciales o trabajando aparentemente con soldadura en fábricas circulan con tanta frecuencia que corremos el riesgo de normalizar algo que, en realidad, representa un cambio tecnológico considerable.

Pero quizá estemos mirando la parte equivocada.

El verdadero salto no consiste en construir máquinas que se parezcan a nosotros. Tampoco en introducir un modelo de lenguaje dentro de un cuerpo metálico. La transformación más profunda aparece cuando una inteligencia artificial deja de limitarse a analizar información sobre el mundo y adquiere capacidad para percibirlo, anticipar qué puede ocurrir y actuar físicamente dentro de él.

A ese conjunto de tecnologías comienza a denominársele *Physical AI* o inteligencia artificial física.

El término está siendo impulsado especialmente desde la industria tecnológica. NVIDIA, por ejemplo, lo utiliza para describir sistemas autónomos capaces de percibir, comprender, razonar y ejecutar acciones en el mundo físico. BMW emplea ya la misma

expresión para referirse a la convergencia entre inteligencia artificial digital, máquinas y robots dentro de sus fábricas. No existe, sin embargo, una frontera científica universal que separe perfectamente la *Physical AI* de conceptos anteriores como *embodied intelligence*, robótica cognitiva o aprendizaje robot. Buena parte de sus fundamentos llevan décadas desarrollándose.

Lo nuevo es la convergencia. Modelos multimodales, visión artificial, aprendizaje por refuerzo, simulación, sensores, sistemas de control, grandes conjuntos de datos, modelos del mundo y robots cada vez más capaces están empezando a formar un mismo ecosistema.

Por eso, más que preguntarnos cuándo tendremos un robot humanoide en casa deberíamos entender qué es lo que cambiará cuando una inteligencia artificial deje de generar respuestas sobre el mundo y tenga que aprender a desenvolverse dentro de él. Esto nos obliga a unir robótica, inteligencia artificial, psicología, neurociencia y ciencias del comportamiento, porque tener un cuerpo cambia la perspectiva a nuestro problema.

De programar movimientos a aprender a actuar

Durante buena parte de la historia de la robótica industrial, el principio era relativamente sencillo, diseñar una máquina para ejecutar de forma extraordinariamente precisa una tarea relativamente bien definida.

Un brazo robot puede soldar miles de puntos con una regularidad imposible para una persona. Pero esa competencia no implica necesariamente que comprenda qué está haciendo. Su entorno, las posiciones de las piezas y las secuencias de operación suelen encontrarse fuertemente controlados.

El reto cambia cuando queremos que una máquina trabaje en escenarios variables. Es entonces ya no basta con programar cada movimiento.

Un ejemplo que permite apreciar la diferencia es el de Tuomas Haarnoja y colaboradores, publicado en *Science Robotics* en 2024, sobre un pequeño robot bípedo aprendía mediante *deep reinforcement learning* a jugar una versión simplificada de fútbol uno contra uno. El sistema desarrolló capacidades como caminar, girar, golpear el balón, levantarse después de caer y adaptar su comportamiento a situaciones cambiantes del juego. Gran parte del aprendizaje se realizó primero en simulación y posteriormente se transfirió al robot real.

La importancia científica no está en conseguir que un robot juegue al fútbol, sino en el procedimiento. En lugar de escribir manualmente una secuencia para cada circunstancia posible, podemos permitir que el sistema explore, reciba información sobre las consecuencias de sus acciones y vaya desarrollando estrategias que resuelvan el problema.

No deberíamos decir que aprende como un ser humano. Los mecanismos biológicos y computacionales son profundamente diferentes. Pero sí existe una semejanza funcional interesante. Su comportamiento deja de depender exclusivamente de instrucciones explícitas y empieza a depender de experiencia, feedback y adaptación.

A este enfoque se han añadido otras vías. Una de ellas es la imitación. Determinados sistemas permiten capturar movimientos realizados por personas y utilizarlos para entrenar humanoides. En 2024, por ejemplo, se presentó un sistema de teleoperación capaz de trasladar en tiempo real movimientos humanos captados mediante una cámara RGB a un humanoide de tamaño completo. Además de permitir controlar el robot, este tipo de interacción proporciona datos que posteriormente pueden utilizarse para aprendizaje por imitación.

La idea recuerda inevitablemente al aprendizaje por observación humano, pero debemos manejar la comparación con prudencia. Un niño no se limita a reproducir trayectorias articulares: interpreta intenciones, contexto social, objetivos y consecuencias. El robot puede aprender regularidades observables sin poseer necesariamente nada equivalente a esa comprensión.

Aun así, el cambio es importante. Estamos pasando progresivamente de programar comportamientos a entrenar políticas de comportamiento.

Y aquí aparecen los modelos *Vision-Language-Action* o VLA, una de las tecnologías que puede acelerar enormemente el proceso. Un modelo de visión y lenguaje puede relacionar una imagen con una instrucción. Un VLA intenta añadir la acción como una tercera dimensión.

Simplificando mucho, veo y recibo una instrucción por lo que produzco una acción física.

RT-2, desarrollado por investigadores de Google DeepMind y colaboradores, fue una demostración temprana de esta idea. El modelo combinaba datos de interacción robótica con conocimiento adquirido mediante entrenamiento sobre información visual y lingüística a gran escala, buscando mejorar la generalización ante objetos o instrucciones no

encontrados exactamente durante el entrenamiento (Zitkovich et al., 2023).

Desde entonces, la investigación ha crecido rápidamente. Una revisión publicada en *IEEE Access* en 2025 describe precisamente los VLA como un intento de integrar visión, lenguaje y acción para obtener políticas capaces de generalizar entre tareas, objetos, entornos e incluso diferentes cuerpos robóticos. Pero la propia revisión advierte de que el campo continúa en una fase relativamente inicial y presenta importantes problemas de datos, evaluación, transferencia y despliegue real (Kawaharazuka et al., 2025).

Google DeepMind ha continuado avanzando en esa dirección. En julio de 2026 presentó Gemini Robotics 2, un sistema que, según la compañía, combina modelos VLA con razonamiento aplicado al entorno físico para controlar desde manipuladores hasta humanoides completos. Los resultados mostrados incluyen manipulación, coordinación corporal y adaptación entre diferentes plataformas. Son resultados técnicamente relevantes, pero conviene mantener una cautela básica: proceden fundamentalmente de evaluaciones y demostraciones de la propia compañía y no equivalen todavía a una validación independiente de autonomía general.

Estamos, por tanto, ante un cambio prometedor, pero todavía incompleto. El robot ya no necesita únicamente saber cómo mover una articulación, necesita relacionar lo que ve con lo que le pedimos, con aquello que puede hacer y con las consecuencias probables de hacerlo. Y ahí el problema empieza a parecerse menos a la automatización tradicional y más a una pregunta sobre inteligencia.

Tener cuerpo cambia la inteligencia

Una taza puede describirse de muchas maneras. Podemos identificar su material, su color, su volumen aproximado o reconocer que pertenece a la categoría «taza».

Pero para un organismo que actúa existe otra información fundamental, como qué puedo hacer con ella. Por ejemplo, puedo agarrarla, puedo levantarla, puedo beber de ella, puedo apoyarla sobre una mesa y puedo sujetarla por el asa dependiendo del tamaño de mi mano.

Esta forma de entender la percepción conecta con el concepto de *affordance*, desarrollado originalmente por el psicólogo James J. Gibson y posteriormente incorporado a la neurociencia y la robótica. Una *affordance* no es simplemente una propiedad del objeto: describe posibilidades de acción que aparecen en la relación entre un agente y su entorno.

La investigación en robótica lleva años intentando utilizar esta idea para que las máquinas no solo reconozcan objetos, sino que aprendan también posibles interacciones con ellos (Jamone et al., 2018).

Aquí aparece una cuestión importante. Comprender físicamente el mundo quizá no consista únicamente en saber qué son las cosas, sino también en aprender qué podemos hacer con ellas.

Ese «podemos» depende además del cuerpo.

Una puerta ofrece posibilidades distintas a una persona, un dron, una silla de ruedas o un robot de seis metros. Un escalón aparentemente trivial para nosotros puede ser un problema complejo de equilibrio para otra morfología.

Esto conecta con uno de los grandes debates de las ciencias cognitivas. La denominada *embodied cognition* o cognición corporizada agrupa varias teorías que cuestionan la idea de que la cognición pueda entenderse completamente como un proceso abstracto separado del cuerpo y del entorno.

Margaret Wilson resumió en 2002 diferentes versiones de esta perspectiva: la cognición suele estar situada, ocurre bajo presión temporal, aprovecha estructuras del entorno y está estrechamente relacionada con la acción. Lawrence Barsalou desarrollaría posteriormente su teoría de *grounded cognition*, destacando el papel de sistemas perceptivos, motores y situacionales en numerosos procesos cognitivos.

Conviene evitar aquí una extrapolación excesiva. Estas teorías intentan explicar aspectos de la cognición humana. No demuestran que una máquina necesite un cuerpo humano para ser inteligente ni que un robot que interactúa físicamente con el entorno posea una cognición equivalente a la nuestra, pero sí aportan una lección enormemente útil, que el cuerpo condiciona el problema que la inteligencia tiene que resolver.

Una mano de cinco dedos, una pinza industrial y una ventosa no interactúan con el mismo mundo funcional, aunque observen exactamente la misma escena. La propia morfología puede incluso simplificar determinados problemas de control. La investigación en *morphological computation* lleva años explorando cómo determinadas propiedades físicas, como la elasticidad, la geometría, la distribución de masas o la dinámica mecánica, pueden contribuir al comportamiento sin que cada detalle deba ser calculado centralmente.

En otras palabras, no toda la inteligencia está necesariamente «en el algoritmo». Parte de la solución puede estar distribuida entre control, cuerpo y entorno.

Esto nos permite proponer un marco sencillo para entender la inteligencia física. No es un modelo científico establecido, sino una propuesta conceptual de trabajo. Y podemos pensar en cuatro capas.

La primera es el cuerpo, con sensores, articulaciones, motores, manos, ruedas, equilibrio, energía o resistencia.

La segunda es la percepción, es decir, detectar objetos, personas, espacios, estados y cambios relevantes.

La tercera es un modelo del mundo, lo que viene a plantear la capacidad para anticipar que una determinada acción probablemente producirá determinadas consecuencias.

Y la cuarta es la acción, o seleccionar un comportamiento, ejecutarlo, observar el resultado y corregirlo.

El ciclo completo sería percibir → interpretar → actuar → observar → consecuencias → adaptar.

Eso distingue bastante bien una parte importante del problema físico frente al puramente digital. Un modelo de lenguaje puede equivocarse describiendo cómo colocar una pieza, un robot puede dejarla caer. Un modelo puede producir una frase incorrecta y generar

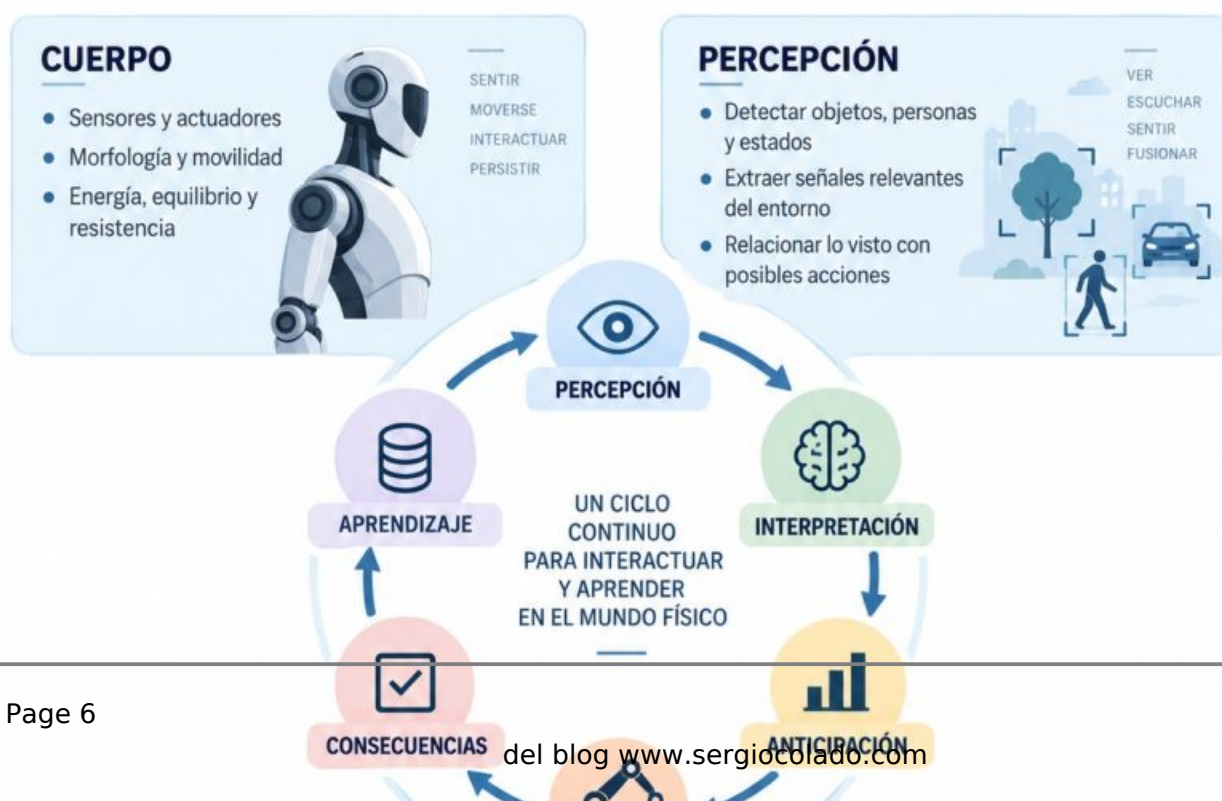
a persona.

se equivoca
lumbre,
lad no

De la IA digital a la inteligencia física

Modelo conceptual propuesto por el autor

Es un marco explicativo, no una taxonomía científica establecida.



De la demostración espectacular a la fábrica real

Por eso resulta especialmente interesante observar qué ocurre cuando estos robots abandonan el laboratorio.

BMW proporciona actualmente uno de los casos mejor documentados. Durante 2025, Figure 02 trabajó en la planta de BMW en Spartanburg, Estados Unidos. BMW confirma que, durante unos diez meses, el sistema intervino en la producción de más de 30.000 BMW X3, manipuló más de 90.000 componentes y acumuló alrededor de 1.250 horas de operación. Su función era concreta: retirar y posicionar piezas de chapa para que posteriormente pudieran realizarse operaciones de soldadura.

Estas cifras son relevantes precisamente porque no describen una demostración de cinco minutos, pero tampoco significan que Figure 02 «fabricara 30.000 automóviles» ni que pudiera trabajar de forma general en cualquier punto de la fábrica. Realizaba una tarea definida dentro de un entorno industrial preparado.

Ese matiz explica bastante bien dónde nos encontramos. Existe suficiente capacidad para obtener valor real en determinadas aplicaciones, pero todavía no existe evidencia equivalente que permita hablar de robots humanoides verdaderamente generalistas.

BMW, de hecho, continúa experimentando. En Leipzig ha iniciado su primer piloto europeo con AEON, desarrollado por Hexagon. El robot participará inicialmente en actividades vinculadas al ensamblaje de baterías de alta tensión y fabricación de componentes.

AEON contiene además una lección de diseño extraordinariamente interesante. Tiene torso humanoide, brazos y capacidad para incorporar diferentes herramientas, pero no camina, se desplaza sobre ruedas.

Esto cuestiona una idea que damos demasiado fácilmente por supuesta, que el robot del futuro tendrá necesariamente que parecerse por completo a nosotros.

Las piernas humanas son extraordinariamente versátiles. Permiten superar escalones, caminar por terrenos irregulares y movernos por espacios contruidos para personas. Pero son también un problema complejo de control, equilibrio y consumo energético. Si una fábrica tiene un suelo perfectamente plano, quizá unas ruedas sean una solución mejor.

Lo mismo ocurre con las manos. Cinco dedos proporcionan una enorme versatilidad, pero una pinza especializada puede resultar mucho más fiable para determinadas operaciones.

Por eso humanoide y Physical AI no son sinónimos.

Un vehículo autónomo forma parte del problema de la inteligencia física sin tener aspecto humano. También un robot móvil, un manipulador adaptativo o determinados sistemas inteligentes distribuidos por un edificio. Y un robot con cabeza, torso, brazos y piernas puede poseer una autonomía muy limitada.

La forma humanoide tiene, sin embargo, una ventaja difícil de ignorar, que hemos construido el mundo para nuestros cuerpos. Puertas, herramientas, estanterías, interruptores, vehículos, pasillos, mesas y escaleras responden a nuestras dimensiones y capacidades.

Por eso imitar parcialmente nuestra morfología puede permitir utilizar infraestructuras existentes sin reconstruirlas alrededor del robot.

Mi interpretación es que probablemente veremos hibridación antes que imitación completa.

Robots que incorporarán aquellas características humanas que resulten funcionalmente útiles y conservarán soluciones mecánicas superiores cuando copiar nuestro cuerpo no aporte ventajas.

AEON sobre ruedas es un buen ejemplo.

El cuerpo está avanzando más rápido que la autonomía

Los vídeos pueden producir otra ilusión, la de creer que movilidad equivale a inteligencia.

Un robot que corre, mantiene el equilibrio o ejecuta una coreografía compleja demuestra un extraordinario logro de ingeniería. Pero eso no significa necesariamente que pueda llegar a una fábrica desconocida, comprender espontáneamente el proceso productivo y decidir qué hacer.

China permite observar esta diferencia a gran escala.

Una investigación publicada por Reuters el 27 de agosto de 2026 describe una industria con más de 150 compañías trabajando en humanoides y un considerable apoyo público. La capacidad mecánica de los robots chinos ha avanzado rápidamente: correr, bailar o

ejecutar secuencias complejas resulta cada vez más frecuente.

Sin embargo, Reuters recoge también una conclusión mucho menos espectacular: la inteligencia, la destreza y la fiabilidad necesarias para realizar trabajo industrial flexible continúan por detrás del hardware. Para muchas operaciones perfectamente estructuradas, brazos industriales convencionales siguen siendo más baratos, rápidos o precisos.

Esta diferencia ayuda a identificar los verdaderos cuellos de botella.

Uno es la generalización. Hacer correctamente mil veces una tarea conocida no equivale a resolver una situación nueva.

Otro es la destreza. Manipular objetos deformables, pequeños, resbaladizos o parcialmente ocultos continúa siendo mucho más difícil de lo que nuestra experiencia humana nos hace imaginar.

Otro es la transferencia. Lo aprendido por un robot no puede trasladarse automáticamente a cualquier cuerpo diferente. Cambian articulaciones, sensores, fuerza, dimensiones y espacios de acción.

Y otro problema enorme son los datos. Los modelos lingüísticos pudieron entrenarse utilizando cantidades inmensas de texto creado durante décadas. Internet contenía ya una parte gigantesca de su materia prima.

En robótica no existe todavía algo comparable a un «Internet de acciones». Una acción física contiene información visual, posiciones articulares, velocidades, fuerzas, contacto, equilibrio, estado del entorno y consecuencias temporales.

Obtener estos datos suele requerir robots reales, personas teleoperándolos o simulaciones suficientemente precisas. Por eso proyectos como Open X-Embodiment y políticas generalistas como Octo resultan relevantes: intentan reunir experiencia procedente de múltiples robots y reutilizarla para reducir la dependencia de entrenamiento específico. Octo, por ejemplo, fue entrenado utilizando unas 800.000 trayectorias procedentes de Open X-Embodiment y evaluado posteriormente en diferentes plataformas. Es un avance importante, pero continúa siendo minúsculo comparado con la escala y diversidad de datos disponibles para lenguaje o visión.

Aquí probablemente estará una de las grandes carreras tecnológicas de los próximos años centrada en cómo proporcionar experiencia física suficiente a las máquinas sin tener que

aprender cada comportamiento desde cero en el mundo real.

Simulación, gemelos digitales, teleoperación, vídeo humano, aprendizaje por imitación, datos sintéticos y modelos capaces de anticipar estados futuros intentan resolver precisamente ese problema.

Pero queda otro componente, nosotros. Porque cuando estos sistemas comparten espacios con personas, la cuestión ya no será únicamente cuánto puede hacer el robot, será también qué creemos nosotros que puede hacer.

Décadas de investigación en interacción humano-robot muestran que la confianza depende de múltiples factores, pero el rendimiento y la fiabilidad del sistema ocupan un lugar especialmente importante (Hancock et al., 2011; McKenna et al., 2024).

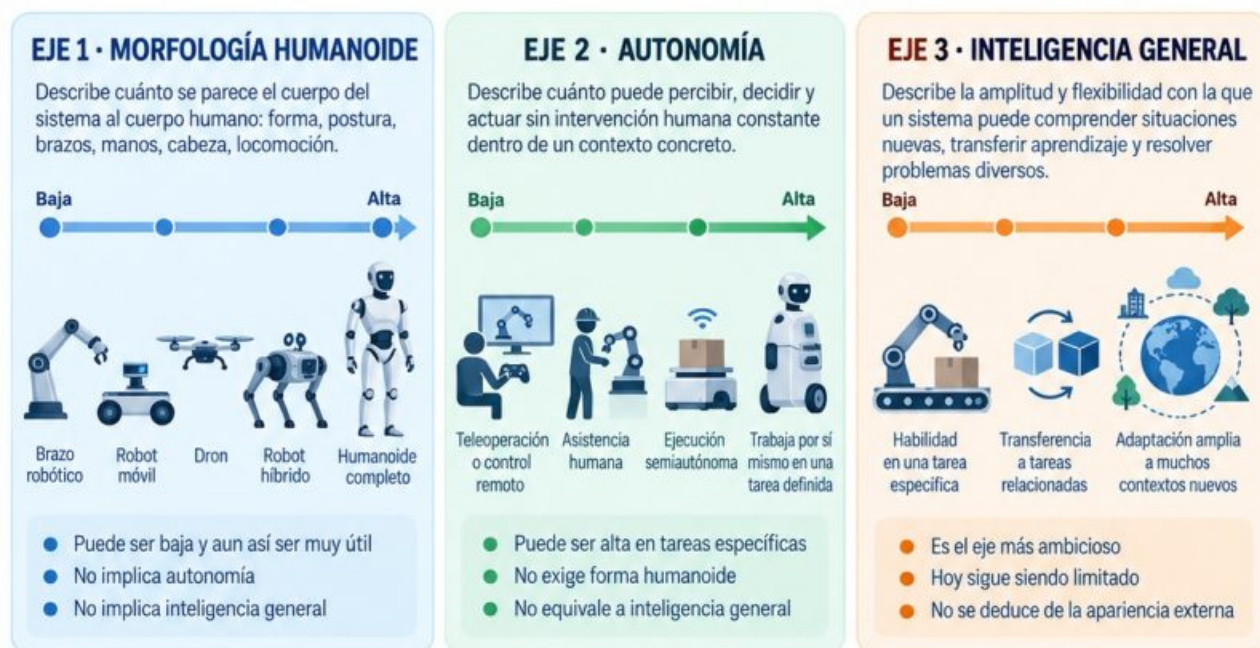
Y aquí aparece un riesgo psicológico interesante. Cuanto más natural sea el lenguaje de una máquina, más fácil puede resultar atribuirle conocimientos, intenciones o comprensión superiores a los que realmente posee. Un robot que responde con fluidez puede parecer más inteligente que otro silencioso incluso aunque ambos utilicen exactamente la misma política de control.

Por eso el objetivo no debería ser simplemente conseguir que confiemos en los robots, debería ser conseguir una confianza calibrada. Confiar cuando el sistema es competente,

Parecer humano no significa pensar como un humano

Tres ejes independientes para entender la robótica y la inteligencia física

Modelo explicativo propuesto por el autor. Sirve para evitar confundir apariencia, autonomía e inteligencia general.



ón física»,
en

Un robot puede parecer muy humano y ser poco autónomo.

También puede ser muy autónomo en una tarea concreta sin poseer inteligencia general.

Conclusión

No estamos todavía ante ejércitos de humanoides capaces de sustituir indistintamente cualquier trabajo humano. La evidencia disponible en 2026 no permite sostenerlo.

La mayor parte de los sistemas más avanzados continúa funcionando mejor en tareas delimitadas, escenarios preparados o demostraciones cuidadosamente diseñadas. Persisten problemas importantes de generalización, destreza, seguridad, autonomía y disponibilidad de datos. Pero sería igualmente equivocado minimizar lo que está ocurriendo.

Hace pocos años discutíamos principalmente si un robot podía aprender a caminar, recuperarse de una caída o adquirir una habilidad mediante aprendizaje por refuerzo. Hoy estamos intentando conectar esas capacidades motoras con modelos que relacionan visión, lenguaje, experiencia y acción. Y algunos de esos sistemas ya están abandonando el laboratorio.

Ese es el cambio importante. La inteligencia artificial ha pasado décadas aprendiendo a representar partes del mundo mediante números, imágenes, lenguaje y predicciones.

La *Physical AI* añade algo diferente, las consecuencias. Una máquina puede ahora percibir una situación, elegir una acción y comprobar qué ha producido físicamente esa decisión.

Eso convierte al mundo en parte del circuito de aprendizaje. No sabemos todavía hasta dónde llevará esta convergencia ni si los humanoides terminarán siendo su manifestación dominante. Probablemente convivirán durante mucho tiempo con robots especializados y arquitecturas híbridas que no tendrán ninguna necesidad de parecernos, pero el problema científico ya ha cambiado. Quizá la gran prueba de inteligencia para una máquina nunca haya sido mantener una conversación que parezca humana. Tal vez sea algo aparentemente mucho más sencillo: entrar en una habitación desconocida, comprender qué está ocurriendo, descubrir qué puede hacerse allí y actuar de forma útil sin romper nada ni hacer daño a nadie.

Cuando una inteligencia artificial pueda resolver sistemáticamente ese problema, ya no estaremos preguntándonos únicamente cuánto sabe sobre nuestro mundo, sino que estaremos empezando a descubrir cuánto ha aprendido a desenvolverse dentro de él.

Referencias

1. Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617-645. doi: 10.1146/annurev.psych.59.103006.093639.
2. BMW Group. (2026). *Leipzig debut: BMW Group introduces humanoid robots – a first in Germany*. Fuente corporativa sobre AEON, Figure 02 y los pilotos de Physical AI.
3. Colado García, S. (2017). Internet de las cosas aplicada a la mente humana (y animal). *Electroeficiencia*.
4. Colado García, S. (2019). *Influencia de la tecnología en el desarrollo del pensamiento y conducta humana*. Independently Published.
5. Colado García, S. (2020). El arriesgado camino de la evolución tecnológica. *Medicina Naturista*, 14(1), 18-21.
6. Colado García, S. (2021). *Multiversos digitales: La tecnología como palanca evolutiva*. Universo de Letras.
7. Figure AI. (2025). *F.02 contributed to the production of 30,000 cars at BMW*. Datos del despliegue industrial de Figure 02. Las cifras principales fueron posteriormente recogidas también por BMW.
8. Haarnoja, T., et al. (2024). Learning agile soccer skills for a bipedal robot with deep reinforcement learning. *Science Robotics*, 9(89). doi: 10.1126/scirobotics.adi8022.
9. Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, 53(5), 517-527. doi: 10.1177/0018720811417254.
10. Jamone, L., Ugur, E., Cangelosi, A., Fadiga, L., Bernardino, A., Piater, J., & Santos-Victor, J. (2018). Affordances in psychology, neuroscience, and robotics: A survey. *IEEE Transactions on Cognitive and Developmental Systems*, 10(1), 4-25. doi: 10.1109/TCDS.2016.2594134.
11. Kawaharazuka, K., Oh, J., Yamada, J., Posner, I., & Zhu, Y. (2025). Vision-Language-Action Models for Robotics: A review towards real-world applications. *IEEE Access*, 13, 162467-162504. doi: 10.1109/ACCESS.2025.3609980.
12. McKenna, P. E., Ahmad, M. I., Maisva, T., Nasset, B., Lohan, K., & Hastie, H. (2024). A meta-analysis of vulnerability and trust in human-robot interaction. *ACM Transactions on Human-Robot Interaction*, 13(3), Article 42. doi: 10.1145/3658897.
13. Reuters. (2026, 27 de agosto). *China can build kung fu-fighting robots. But it can't get them to do factory work*. Investigación sobre el desarrollo industrial y las limitaciones actuales de los humanoides chinos.
14. Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9, 625-636. doi: 10.3758/BF03196322.
15. Zitkovich, B., et al. (2023). RT-2: Vision-Language-Action Models transfer web knowledge to robotic control. *Proceedings of the 7th Conference on Robot Learning*,

PMLR 229, 2165-2183.