



Fuente:

Xataka

Gemini Robotics combina visión, lenguaje y acción para mejorar las capacidades de los robots. Se trata de un modelo de IA desarrollado por Google DeepMind que está basado en Gemini 2.0.

La robótica y la inteligencia artificial (IA) van de la mano. De nada serviría desarrollar robots humanoides capaces de levantar toneladas, con sensores de última generación, si no contáramos con un sistema inteligente que les permitiera interpretar el entorno y actuar en consecuencia. Sin IA, un robot moderno sería poco más que un montón de hardware sofisticado pero inútil. Son los algoritmos avanzados los que transforman esa potencia bruta en máquinas capaces de aprender, optimizar su rendimiento y responder de manera autónoma a los desafíos que se les presentan.

Desde ASIMO, el icónico robot de Honda de los 2000, hasta Sophia, Optimus de Tesla o Figure, la IA ha ido abriéndose camino en la robótica humanoide. Sin embargo, aún estamos lejos de ver máquinas que realmente igualen la versatilidad del cuerpo humano. Por más avanzados que sean, siguen teniendo problemas para moverse en entornos no controlados y manipular objetos cotidianos puede ser un auténtico desafío.

Gemini Robotics: la apuesta de Google para llevar la IA al mundo físico

Mientras tanto, en el mundo digital, la IA avanza a un ritmo completamente distinto. Ya es capaz de mantener conversaciones muy cercanas a las de una persona, superar exámenes con puntuaciones sorprendentes y resolver problemas complejos con una rapidez que hasta hace unos años parecía ciencia ficción. Un contraste que deja claro que, aunque la inteligencia artificial progresa a pasos agigantados, todavía queda mucho camino por recorrer en su integración con la robótica.

Estos desafíos están llevando a una nueva generación de modelos de IA diseñados específicamente para esta disciplina. Google, como era de esperar, no quiere quedarse atrás y ya trabaja en soluciones que prometen llevar a los robots humanoides un paso más allá. Su apuesta pasa por Gemini 2.0, que ahora tiene dos versiones pensadas para mejorar la interacción y el control de estas máquinas.

Por un lado, Gemini Robotics se enfoca en la visión, el lenguaje y la acción (VLA), lo que le permite tomar el control directo de los robots y mejorar su capacidad de respuesta en entornos dinámicos. Por otro, Gemini Robotics-ER está pensado para los expertos en robótica, dándoles las herramientas necesarias para desarrollar y ejecutar sus propios programas con habilidades de razonamiento avanzadas.

Google ha identificado tres cualidades esenciales que, según explican, deben tener los robots para ser realmente útiles para las personas.

- Generalidad. Un buen robot no solo debe ejecutar tareas predefinidas, sino también adaptarse a situaciones inéditas y resolver problemas sobre la marcha. Debe ser capaz de desenvolverse en entornos nuevos, manejar objetos desconocidos e interpretar instrucciones variadas sin depender de un entrenamiento previo. Según pruebas internas, su rendimiento en tareas imprevistas duplica con creces el de otros modelos de visión-lenguaje-acción de última generación.
- Interactividad. En un mundo en constante cambio, los robots deben ser capaces de comunicarse de forma natural y responder a instrucciones en tiempo real. Gemini Robotics comprende comandos en lenguaje cotidiano y en múltiples idiomas, adaptando su comportamiento según la conversación o el entorno. Además, monitoriza continuamente lo que ocurre a su alrededor y ajusta sus acciones en función de nuevas órdenes o cambios en el escenario.
- Destreza. Muchas tareas que los humanos realizamos sin esfuerzo requieren una motricidad extremadamente precisa, algo que la mayoría de los robots aún no han logrado dominar. Gemini Robotics, sin embargo, es capaz de realizar tareas complejas de varios pasos que exigen una manipulación minuciosa, como doblar origami o empacar un bocadillo en una bolsa Ziploc, demostrando un nivel de destreza superior.

Gemini Robotics no solo destaca en la resolución de tareas imprevistas, sino que su capacidad de generalización duplica con creces el rendimiento de otros modelos de visión-lenguaje-acción. Según el informe técnico de Google, es capaz de adaptarse a escenarios inéditos y tomar decisiones sin entrenamiento previo, acercando a los robots a una autonomía real.

Además, ha sido diseñado para funcionar con distintos tipos de robots. Aunque se entrenó principalmente con ALOHA 2, una plataforma de dos brazos, también ha demostrado controlar sistemas como los brazos Franka, usados en laboratorios, e incluso humanoides más avanzados como Apolo, desarrollado por Apptronik. Su flexibilidad lo convierte en un modelo adaptable a diversas aplicaciones, desde la industria hasta la asistencia.

Por ahora, no hay una fecha prevista para un despliegue generalizado de Gemini Robotics o Gemini Robotics-ER. La tecnología sigue en desarrollo y, por el momento, solo un grupo reducido de empresas está teniendo acceso a estas herramientas.

Google DeepMind está colaborando con Apptronik en la construcción de la próxima generación de robots humanoides, explorando cómo integrar estos modelos de IA en sistemas más avanzados. Además, algunos testers de confianza, como Agile Robots, Agility Robotics, Boston Dynamics y Enchanted Tools, ya están probando Gemini Robotics-ER, aunque no está claro si ese acceso se ampliará en el futuro.

Mientras tanto, Google DeepMind sigue trabajando en nuevos marcos y benchmarks de seguridad para evaluar los posibles riesgos de la IA en entornos físicos. Todo esto deja claro que, aunque el proyecto avanza, todavía queda un largo camino antes de que esta tecnología llegue al gran público.

Disponible en:

<https://www.xataka.com/robotica-e-ia/google-no-quiere-que-ia-se-quede-pa...> [1]

Links

[1] <https://www.xataka.com/robotica-e-ia/google-no-quiere-que-ia-se-quede-pantalla-gemini-robotics-su-plan-robots-actuen-mundo-real>