AI DAILY REPORT

Informe nocturno de IA del 16 de octubre de 2026

Acceso para evaluación externa, verificación de robots en campo y agentes de investigación

2026. 10. 16

Una empresa desarrolladora de IA abre herramientas internas de evaluación de riesgos a evaluadores externos, mientras los agentes de programación conectan alertas de seguridad con borradores de corrección. Los robots amplían su campo de aplicación a tareas logísticas y procesos industriales, y en medicina y ciencia asisten en la búsqueda de candidatos y el diseño de ensayos clínicos. El lanzamiento de productos, los resultados de investigación y los futuros planes de prueba piloto se encuentran en etapas diferentes.

1 Anthropic otorga acceso a evaluación interna de riesgos a evaluadores externos

2 GitHub conecta alertas de seguridad con borradores de corrección

3 OpenGCUA integra clics y comandos en un único programa

4 Data World entrena el análisis con datos virtuales de estructura causal

5 Liquid AI ofrece un modelo que devuelve probabilidades en vez de explicaciones

6 Voxel51 integra búsqueda de datos de sensores y revisión de etiquetas

7 DeepMind aplica agentes a la búsqueda de candidatos antimicrobianos

8 Conversión de criterios de ensayos oncológicos en reglas de búsqueda

9 Ensayo clínico para evaluar entrevistas con tutores por IA previas a consulta

10 KinoGen genera videos de movimiento adaptados al robot como datos de entrenamiento

11 Maum AI impulsa pruebas piloto de robots patrulla cuadrúpedos en vías peatonales

12 Yujin Robot se expande del transporte de mercancías a tareas de agarre y guardado

13 T-Robotics participa en investigación para automatizar procesos de módulos ópticos

14 Francia prepara para 2027 la formación de una unidad de robots de combate

15 Codex y Claude Code mejoran el control de interrupciones e informes de fallos

Supervisión de IA

1 Anthropic otorga acceso a evaluación interna de riesgos a evaluadores externos

Anthropic ha comenzado a conceder a seis empleados de la organización sin fines de lucro de evaluación de IA METR acceso a equipos informáticos, espacios de oficina y herramientas internas de la empresa. Según se explica, el propósito es permitirles examinar modelos y mecanismos de seguridad a un nivel similar al del equipo interno de evaluación de riesgos. Este enfoque va más allá de limitarse a probar las respuestas del modelo ya finalizado, permitiendo que evaluadores externos examinen también el proceso mediante el cual la compañía determina los riesgos.

METR puede publicar de forma independiente sus análisis sobre riesgos, incidentes y prácticas operativas, y Anthropic no puede rechazar su divulgación. No obstante, la empresa puede solicitar la redacción de categorías restringidas predefinidas, como información de clientes, secretos legales y arquitectura de modelos. METR puede emitir una opinión pública sobre si dicha solicitud fue justificada. Este anuncio se refiere a la concesión de acceso y derechos de publicación por parte de la empresa, y no implica que ya se hayan publicado nuevos resultados de evaluación.

Cuaderno de IA de hoy

Acceso a evaluación externa

6 empleados de METR

Acceso a herramientas internas

Publicación garantizada



Informe integral de IA de KIMKJ

Automatización del desarrollo

2 GitHub conecta alertas de seguridad con borradores de corrección

GitHub ha lanzado oficialmente Agentic Autofix, que delega las alertas de seguridad del código en agentes de Copilot. El agente lee los archivos pertinentes, genera una propuesta de corrección y vuelve a ejecutar los análisis de CodeQL. Si la alerta persiste, repite el proceso de corrección y abre un borrador de pull request que detalla los cambios y el proceso de verificación. El pull request es una solicitud para revisar el código modificado, preservando el espacio para la supervisión humana.

Es posible seleccionar una alerta individual de la rama predeterminada o agrupar hasta 25 alertas en una única solicitud de corrección. No se aplica a alertas detectadas exclusivamente en otras ramas o en solicitudes de corrección. Requiere productos de seguridad designados de Enterprise Cloud, activación por parte de administradores y configuración de facturación de Copilot. Se generan costes según los créditos de IA y el tiempo de ejecución de Actions, sin que se hayan publicado datos sobre la precisión de la corrección automática ni sobre la tasa real de fusión.

Cuaderno de IA de hoy

De alerta a borrador

Hasta 25 alertas

Reinspección tras corrección

Revisión de código por humanos



Informe integral de IA de KIMKJ

Control de pantalla

3 OpenGCUA integra clics y comandos en un único programa

El equipo de investigación de OpenGCUA ha propuesto un método de uso de ordenadores que representa los clics en pantalla mediante código y combina información de accesibilidad, API y comandos de shell en un mismo programa. Se incorporan bucles y bifurcaciones condicionales dentro del código, permitiendo que el propio modelo elija cuándo volver a observar la pantalla. Esto supone una transición desde capturar pantallas a intervalos fijos para actuar hacia un esquema en el que el modelo diseña conjuntamente la observación y la ejecución.

En la evaluación realizada por los autores, la puntuación parcial media de ocho modelos en OSWorld-v2 aumentó en 19,6 puntos, y los pasos por tarea se redujeron en un promedio del 53,0% en comparación con el método de referencia de procesamiento por lotes en pantalla. Estos datos corresponden a resultados autoinformados bajo dichas condiciones de prueba. El entorno de ejecución y las herramientas de desarrollo publicados se hallan en fase preliminar, y la comparación en dispositivos reales se limita a cuatro tareas en macOS y ejecuciones seleccionadas. No puede asumirse que se obtengan las mismas mejoras en la totalidad de las tareas cotidianas.

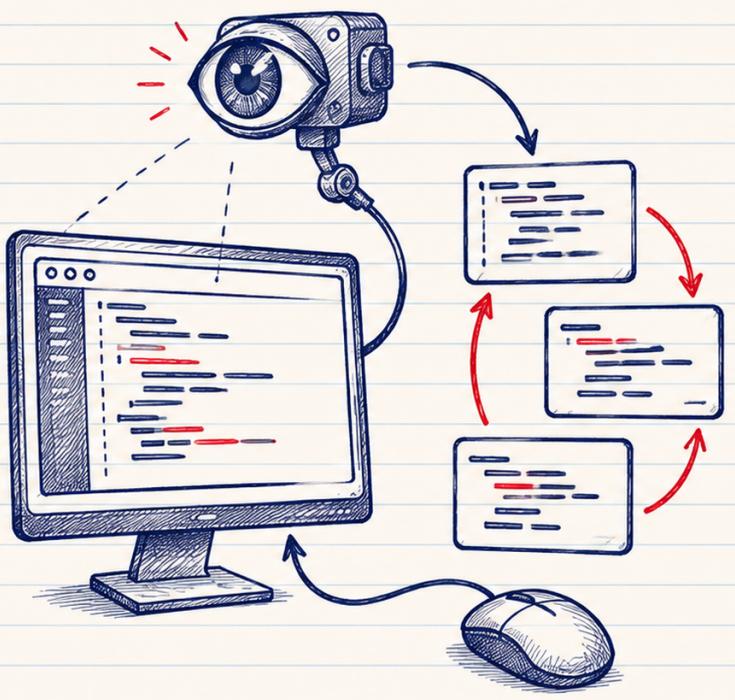
Cuaderno de IA de hoy

Unir clics mediante código

Aumento medio de 19,6 puntos

Reducción del 53,0% de pasos

Validación física limitada



Informe integral de IA de KIMKJ

Análisis de datos

4 Data World entrena el análisis con datos virtuales de estructura causal

Data World define primero las relaciones y reglas con las que se generan los datos, antes de crear archivos sin ningún criterio. A ese estado de referencia le añade valores faltantes, identificadores que no coinciden y cambios de formato para producir archivos de análisis. Un agente auditor independiente comprueba si las relaciones se mantienen y si la respuesta correcta puede deducirse a partir de los archivos observables. Los investigadores construyeron 2135 tareas de aprendizaje por refuerzo a partir de 1037 mundos virtuales de entrenamiento.

Los investigadores afirmaron que los modelos entrenados con estas tareas mejoraron en tres evaluaciones de análisis de datos. Sin embargo, no se realizaron experimentos que eliminaran los componentes uno por uno para comprobar su efecto, ni se hizo una comparación directa entre las tareas sintéticas y los datos reales. En las solicitudes reales, con frecuencia el objetivo es ambiguo o falta información necesaria. Que suba la puntuación de cálculo y que se produzca un análisis útil para la toma de decisiones en el mundo real son cuestiones que deben verificarse por separado.

Cuaderno de IA de hoy

Datos creados a partir de reglas

1037 mundos de entrenamiento

2135 tareas de aprendizaje

Aplicación real: verificar aparte



Informe integral de IA de KIMKJ

Modelos ligeros

5 Liquid AI ofrece un modelo que devuelve probabilidades en vez de explicaciones

Liquid AI ha incorporado d1-ultra, un modelo destinado a la toma de decisiones. Al recibir texto e imágenes de entrada, en lugar de generar textos explicativos extensos, devuelve las probabilidades correspondientes a cada opción. Diseñado para emitir juicios en un único cálculo, se orienta a tareas que requieren clasificación, inspección y asignación de puntuaciones más que redacción de textos. Añade un modelo superior a la gama d1 existente y se suministra mediante API.

El precio de referencia es de 0,042 dólares por cada millón de tokens de entrada, sin tarifa por tokens de salida. No obstante, se aplica la condición de que, si se formulan múltiples preguntas sobre el mismo material, los datos de entrada se facturan por separado para cada consulta. Esta es la razón por la cual no debe calcularse el coste total de procesamiento basándose únicamente en la ausencia de costes de salida.

Aunque las probabilidades devueltas permiten clasificar entre procesamiento automatizado y revisión humana, una probabilidad alta no garantiza por sí sola la exactitud del dictamen real.

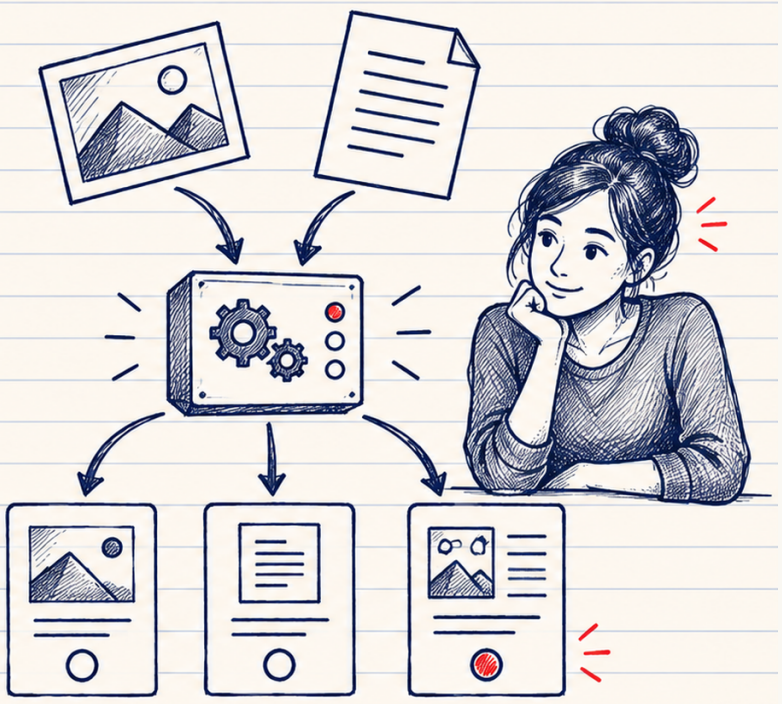
Cuaderno de IA de hoy

Probabilidad en vez de texto

Entrada de texto e imagen

Probabilidad por opción

Cobro de entrada por consulta



Informe integral de IA de KIMKJ

IA de campo

6 Voxel51 integra búsqueda de datos de sensores y revisión de etiquetas

Voxel51 ha presentado la versión 3.0 de su plataforma de datos para el desarrollo de IA de campo. Se trata de un producto que permite examinar conjuntamente los datos registrados por múltiples sensores a lo largo del tiempo, enlazando la búsqueda, selección, etiquetado y evaluación. Permite que los desarrolladores identifiquen escenas en las que un robot ha fallado y las utilicen para el siguiente ciclo de entrenamiento y evaluación dentro del mismo entorno. Se centra en ayudar a los desarrolladores, que antes gestionaban los datos de sensores por separado, a analizar conjuntamente el contexto de las escenas reales.

También se ha incorporado una función en la que un modelo de visión y lenguaje genera etiquetas a partir de instrucciones en lenguaje natural y ejemplos, permitiendo que un revisor humano examine los resultados. Las etiquetas son descripciones asignadas a objetos y situaciones en imágenes, por lo que una anotación errónea repercute en el entrenamiento y la evaluación. La generación automática no equivale a una verificación de precisión. La versión comercial FiftyOne Enterprise pasa a denominarse Voxel51, mientras que el proyecto de software de código abierto conserva el nombre FiftyOne.

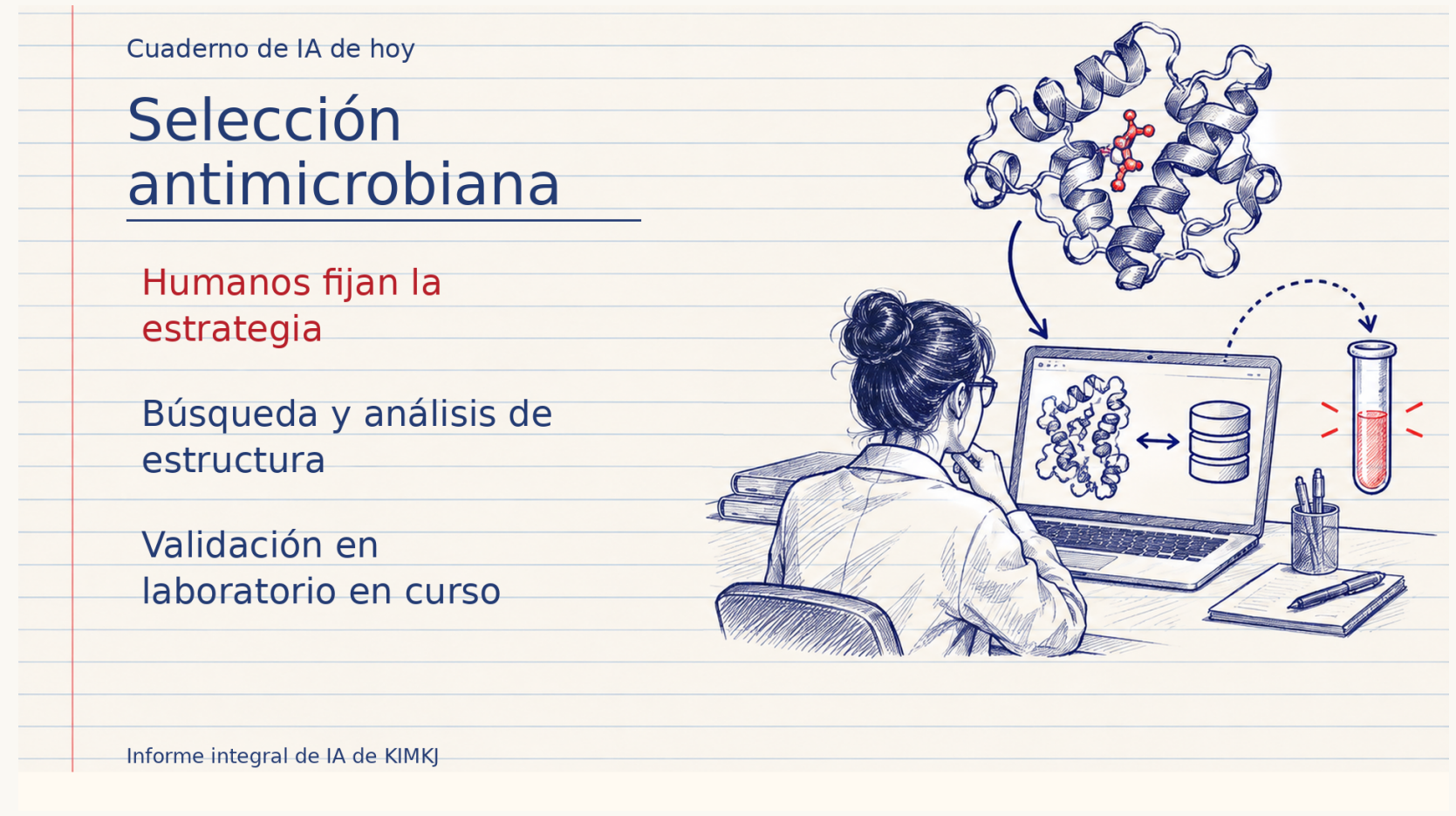


Investigación científica

7 DeepMind aplica agentes a la búsqueda de candidatos antimicrobianos

Google DeepMind ha revelado un proceso de investigación que emplea Antigravity para identificar péptidos antimicrobianos candidatos. Los péptidos son moléculas compuestas por cadenas de aminoácidos, y el equipo busca entre ellos candidatos con potencial actividad antimicrobiana. En este flujo, los humanos establecen las preguntas y la estrategia científica, mientras que el agente ejecuta de forma sucesiva la búsqueda en bases de datos de proteínas, el análisis estructural y la revisión bibliográfica.

El candidato presentado fue identificado en Streptococcus pseudopneumoniae y se encuentra actualmente en fase de validación experimental junto con otros candidatos. También se presentó una herramienta tridimensional que permite a investigadores y agentes examinar conjuntamente estructuras de proteínas. Constituye un caso de investigación en el que se conectaron múltiples herramientas analíticas para acotar objetivos computacionalmente prometedores. Esto no demuestra la eficacia antimicrobiana ni la seguridad de un fármaco para uso en humanos, debiendo distinguirse entre el hallazgo de candidatos y los resultados experimentales.



Investigación médica

8 Conversión de criterios de ensayos oncológicos en reglas de búsqueda

Se ha publicado un estudio que transforma los criterios de elegibilidad de ensayos clínicos oncológicos en reglas de consulta de bases de datos mediante modelos de lenguaje. Una vez estructurada la información del paciente, el sistema localiza los ensayos correspondientes con base en dichas reglas sin consultar reiteradamente al modelo de lenguaje. Los investigadores emplearon 746 ensayos relativos a cáncer de mama, cáncer de colon y cáncer de pulmón no microcítico, junto con 24 casos sintéticos de pacientes elaborados por especialistas médicos. No se trata de un estudio que evalúe resultados terapéuticos en pacientes de práctica clínica real.

Excluyendo el tiempo de preparación de los datos del paciente, la mediana del tiempo de emparejamiento fue de 0,459 segundos por paciente. De 240 pares paciente-ensayo contrastados con el consenso médico, 230 pares coincidieron, lo que representa una concordancia del 95,8% y una sensibilidad del 83,3%. Esto implica que hubo casos en los que se omitieron candidatos elegibles. Se trata de un artículo previo a la revisión por pares, y en las condiciones de emparejamiento no se incluyeron valores analíticos de laboratorio ni comorbilidades. Se requiere validación adicional antes de su aplicación en el reclutamiento real de pacientes.



Investigación médica

9 Ensayo clínico para evaluar entrevistas con tutores por IA previas a consulta

Se ha registrado un ensayo clínico en el que los tutores realizan una entrevista previa con IA antes de la consulta psiquiátrica para transmitir información organizada al personal médico. El tutor completa de forma remota una única entrevista adaptativa antes de la visita ambulatoria programada. Una entrevista adaptativa es un método en el que las preguntas subsiguientes varían en función de las respuestas anteriores. El sistema genera un informe estructurado para el equipo médico y candidatos para el cribado diagnóstico de nueve trastornos predefinidos.

El estudio está planificado para 120 participantes y evaluará el rendimiento como apoyo al diagnóstico, la calidad de recopilación y documentación de datos, la usabilidad y la seguridad. El médico responsable consulta la información generada por la IA pero mantiene la potestad de decidir la evaluación clínica final y el diagnóstico. Los datos disponibles hasta el momento corresponden al diseño del ensayo y la información de registro, sin que existan resultados clínicos medidos. Tampoco se especifican el modelo de lenguaje utilizado ni los nombres concretos de los nueve trastornos, por lo que no es posible determinar de antemano su eficacia.

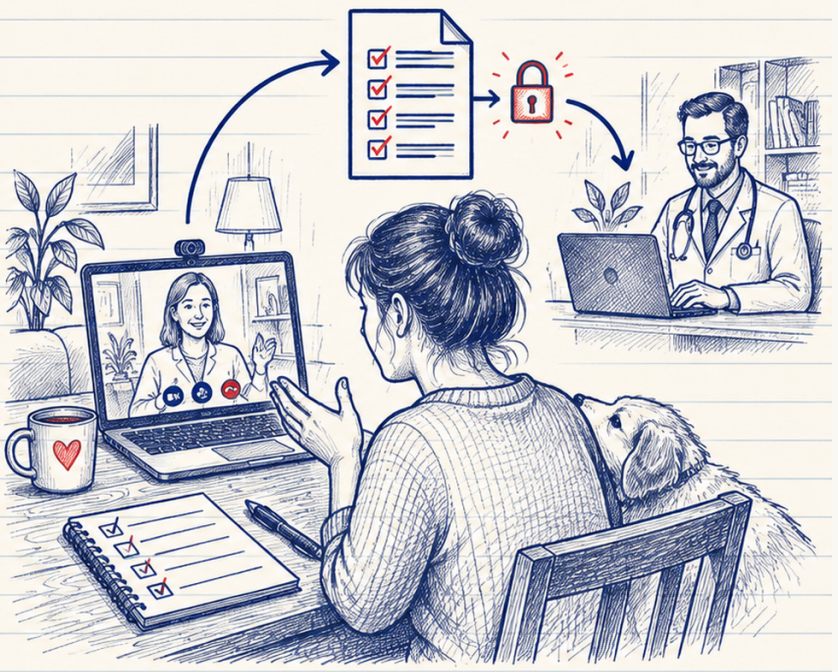
Cuaderno de IA de hoy

Entrevistas previas con IA

Planificado para 120 personas

Cribado de 9 trastornos

Diagnóstico médico final



Informe integral de IA de KIMKJ

Investigación robótica

10 KinoGen genera videos de movimiento adaptados al robot como datos de entrenamiento

El equipo de KinoGen ha propuesto un método que genera videos considerando la geometría del robot y de los objetos, reconstruyendo a partir de ellos las trayectorias de movimiento que el robot debe seguir. Al utilizar videos sintéticos generados por simulación desde dos puntos de vista, se prescinde de la etapa de capturar previamente movimientos humanos para adaptarlos a la estructura del robot. El sistema está diseñado para regenerar el video tomando como referencia la escena correspondiente cuando el usuario modifica la postura objetivo del objeto.

En un artículo previo a la revisión por pares, los autores evaluaron ocho tareas que combinaban locomoción y manipulación. Afirmaron haber reducido la penetración en objetos y el deslizamiento de pies en comparación con los métodos de referencia, conservando más del 90% de los contactos indicados. El esquema de control entrenado con trayectorias sintéticas se aplicó también a un robot Unitree G1 equipado con manos diestras. No obstante, no se especifican la tasa concreta de éxito en robots reales ni los tiempos de generación, por lo que aún no es posible evaluar su productividad en entornos reales.

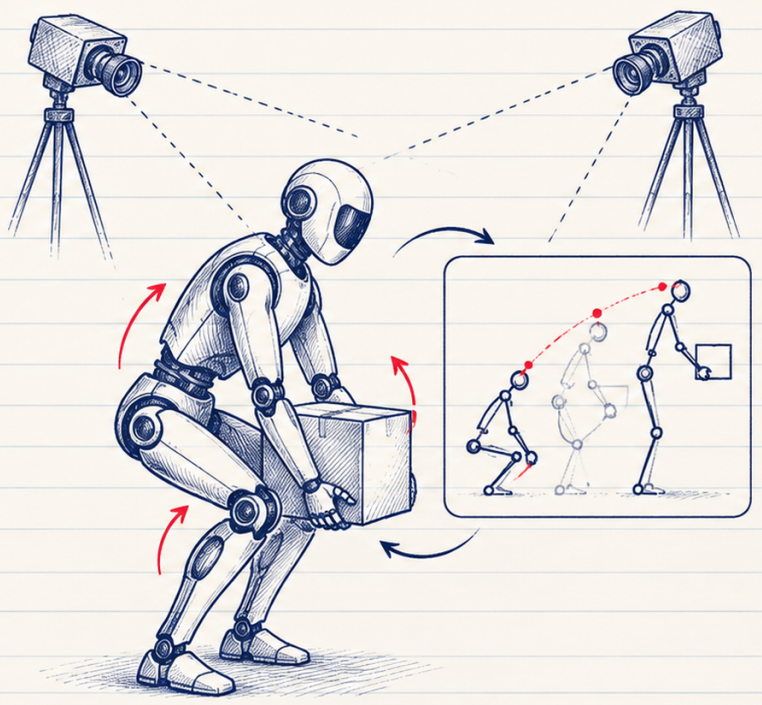
Cuaderno de IA de hoy

Movimiento según el robot

Videos de dos puntos de vista

8 tareas de locomoción y manipulación

Éxito en robot real no indicado



Informe integral de IA de KIMKJ

Robótica surcoreana

11 Maum AI impulsa pruebas piloto de robots patrulla cuadrúpedos en vías peatonales

El proyecto de robots patrulla cuadrúpedos de exterior presentado por Maum AI ha sido deliberado por el Comité de Deliberación de Exenciones Regulatorias de Convergencia Industrial. La empresa explicó que el sistema existente de certificación de seguridad operativa está centrado en robots móviles sobre ruedas, lo que dificulta su aplicación a robots cuadrúpedos cuya postura y puntos de apoyo varían al caminar. Conviene distinguir el hecho de haber sido sometido a deliberación del inicio efectivo de las operaciones.

Sujeto a la confirmación de la exención regulatoria para pruebas y al cumplimiento de condiciones adicionales, se prevé probar seis unidades del robot Haechi en senderos fluviales y parques de la ciudad de Cheonan. Los robots ejecutan IA de forma local en el dispositivo para reconocer el entorno y situaciones de peligro, transmitiendo video e información de riesgos al centro de control. Los principales objetivos de prueba son la detección de peligros en riberas fluviales y la respuesta ante incidentes personales en parques nocturnos. Aunque se proyecta utilizar los datos de navegación y decisión recopilados para el entrenamiento de otros robots, aún no se dispone de resultados de las pruebas previstas.

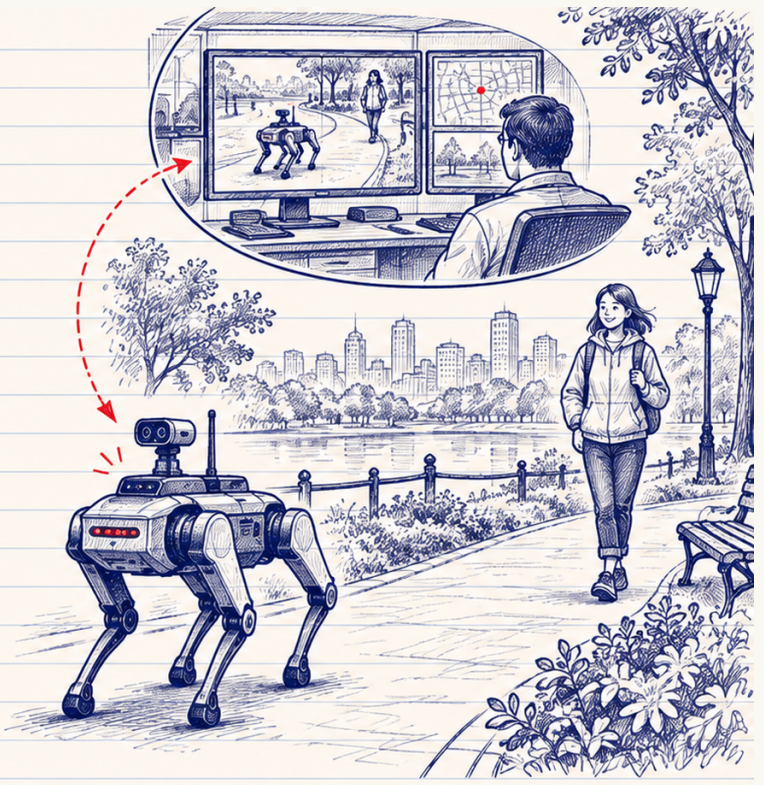
Cuaderno de IA de hoy

Pruebas de patrulla cuadrúpeda

Plan para 6 robots Haechi

Exención y condiciones previas

Senderos y parques en Cheonan



Informe integral de IA de KIMKJ

Robótica surcoreana

12 Yujin Robot se expande del transporte de mercancías a tareas de agarre y guardado

Yujin Robot expande sus áreas de negocio hacia la automatización de tareas en entornos de fabricación y logística. Planea desarrollar un robot con características de humanoide adaptado a instalaciones logísticas combinando sus propios robots móviles autónomos con brazos robóticos y sensores de visión de terceros. El propósito es trascender el mero transporte de cajas a su destino para realizar tareas de recolección y clasificación de mercancías o manipulación de maquinaria. Se concibe montando dispositivos de trabajo de la parte superior del cuerpo sobre una plataforma móvil.

La empresa ha fijado como objetivos de aplicación el picking, la clasificación y la carga (colocación en contenedores) de productos, así como la operación de equipos. Empleará centros logísticos reales como escenarios de prueba para validar la capacidad de trabajo continuo y la viabilidad económica, ampliando además la colaboración con instituciones de investigación como el Instituto de Tecnología de Georgia. El anuncio de ampliar las operaciones no implica haber alcanzado una automatización total. La estabilidad con la que operen en procesos que requieren intervención manual y si la eficacia compensa los costes dependerán de la verificación en campo.

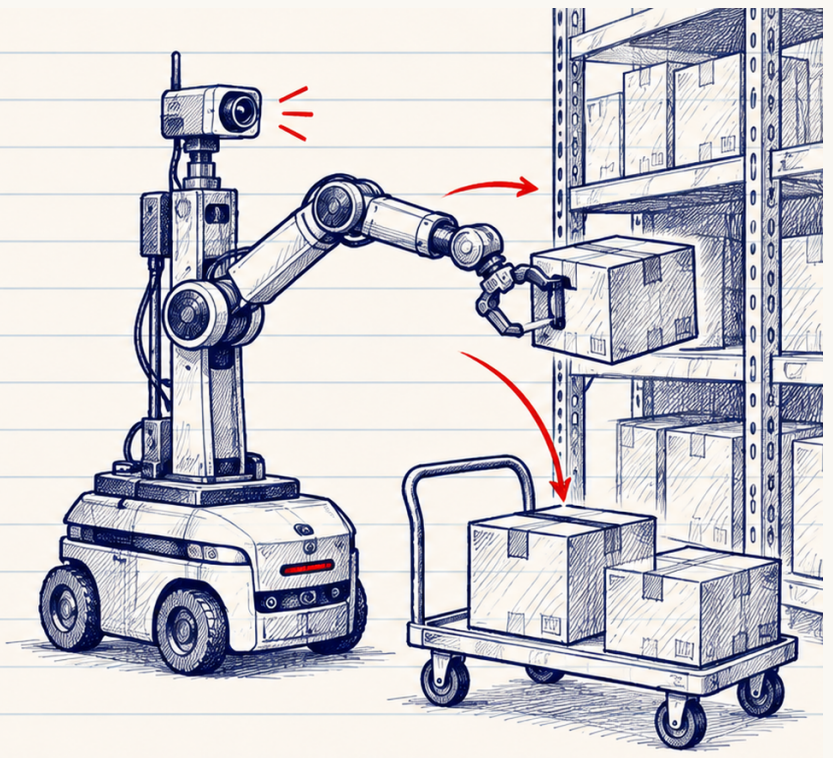
Cuaderno de IA de hoy

Del traslado a la manipulación

Brazo acoplado a robot móvil

Meta: picking, clasificación y carga

Validación económica en campo



Informe integral de IA de KIMKJ

IA industrial

13 T-Robotics participa en investigación para automatizar procesos de módulos ópticos

T-Robotics participa en un proyecto de investigación y desarrollo que aplica IA física al proceso de curado de módulos ópticos. La IA física se ocupa de equipos físicos y movimientos en el mundo real. El proyecto busca automatizar el proceso sin presencia humana y desarrollar tecnología de mantenimiento predictivo que detecte anomalías con anticipación a partir de los datos de procesos y del estado de las instalaciones. Se trata de un estudio orientado a pasar de reparar equipos tras su avería y detención a identificar indicios de anomalías de manera anticipada.

El proyecto cuenta con una inversión de 10.200 millones de wones en fondos de I+D gubernamentales y se desarrollará hasta diciembre de 2029 con la participación conjunta de grandes empresas surcoreanas y KAIST, entre otros. Esta cifra representa la subvención total para el proyecto de investigación conjunto y no corresponde a los ingresos ni a los contratos de T-Robotics como entidad única. Actualmente se encuentra en la etapa inicial de I+D, siendo la estabilidad de producción y la mejora de eficiencia metas previstas. Se requerirán futuras pruebas piloto en campo y validaciones comerciales para juzgar los resultados reales.

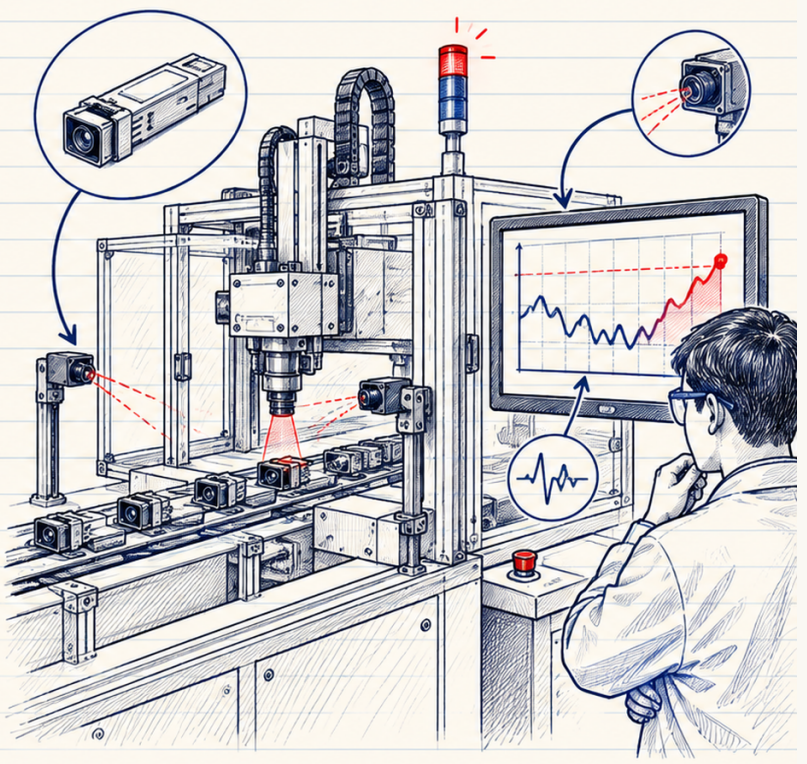
Cuaderno de IA de hoy

Detección previa de anomalías

Proyecto de 10.200 millones ₩

Desarrollo hasta diciembre 2029

Fase inicial y validación



Informe integral de IA de KIMKJ

IA militar

14 Francia prepara para 2027 la formación de una unidad de robots de combate

El proyecto Pendragon, que el Ejército de Tierra francés hizo público el pasado abril, es una investigación para operar conjuntamente robots de combate terrestres y drones. Es una iniciativa impulsada de forma conjunta por la Agencia de Defensa de Inteligencia Artificial y el Mando de Combate del Futuro, y plantea conectar en red vehículos terrestres no tripulados y drones. El Ejército planea formar hasta 2027 la primera unidad de robots de combate, que agrupará más de veinte equipos. Conviene distinguir entre las demostraciones actuales y los objetivos futuros de formación de la unidad.

El software de control autónomo que se aplica apoya las decisiones del mando, y la decisión final sobre el uso de armas letales queda bajo control humano. Es decir, se han diseñado por separado la tecnología para que varios equipos se muevan juntos y la autoridad para aprobar el uso de la fuerza. La etapa que se puede confirmar actualmente es la de demostraciones y pruebas; aún no se ha completado el despliegue operativo de una unidad regular. Tampoco este material presenta cifras de rendimiento independientes sobre la capacidad de operación en el terreno.

Cuaderno de IA de hoy

Preparación de la unidad de robots

Objetivo de formación: 2027

Vehículos y drones conectados

El uso final de la fuerza, humano



Informe integral de IA de KIMKJ

Operación de agentes

15 Codex y Claude Code mejoran el control de interrupciones e informes de fallos

Codex 0.163.0 ha añadido funciones para ejecutar herramientas independientes de solo lectura en paralelo y redirigir el curso de acción mediante instrucciones de seguimiento durante una respuesta. Al interrumpir una tarea, se detienen tanto el código en ejecución como las llamadas a herramientas que contiene. Al reanudar una conversación guardada, se restauran los permisos y autorizaciones de red previos, pero supeditados a los nuevos valores definidos por el usuario y a las directivas de gestión vigentes.

Claude Code 2.1.301 ha incorporado una opción para finalizar la solicitud si los límites de uso tardan demasiado en restablecerse durante los reintentos desatendidos. Asimismo, se comunicó la corrección de un error que reportaba éxito tras saltarse pasos de trabajo después de leer avisos de permisos, así como de un problema en el que se generaban costes de resumen sin confirmación al cambiar de modelo. En ambos productos, los cambios no solo atañen a la velocidad de ejecución, sino también a la interrupción, los permisos, los costes y la finalización real. No se divulgaron cifras concretas sobre la reducción de la tasa de fallos.

Cuaderno de IA de hoy

Pausa y finalización precisas

Cambio inmediato por instrucción

Termina esperas largas de límite

Corrige falso éxito al saltar



Informe integral de IA de KIMKJ