

El colapso de la IA: una amenaza real para la tecnología

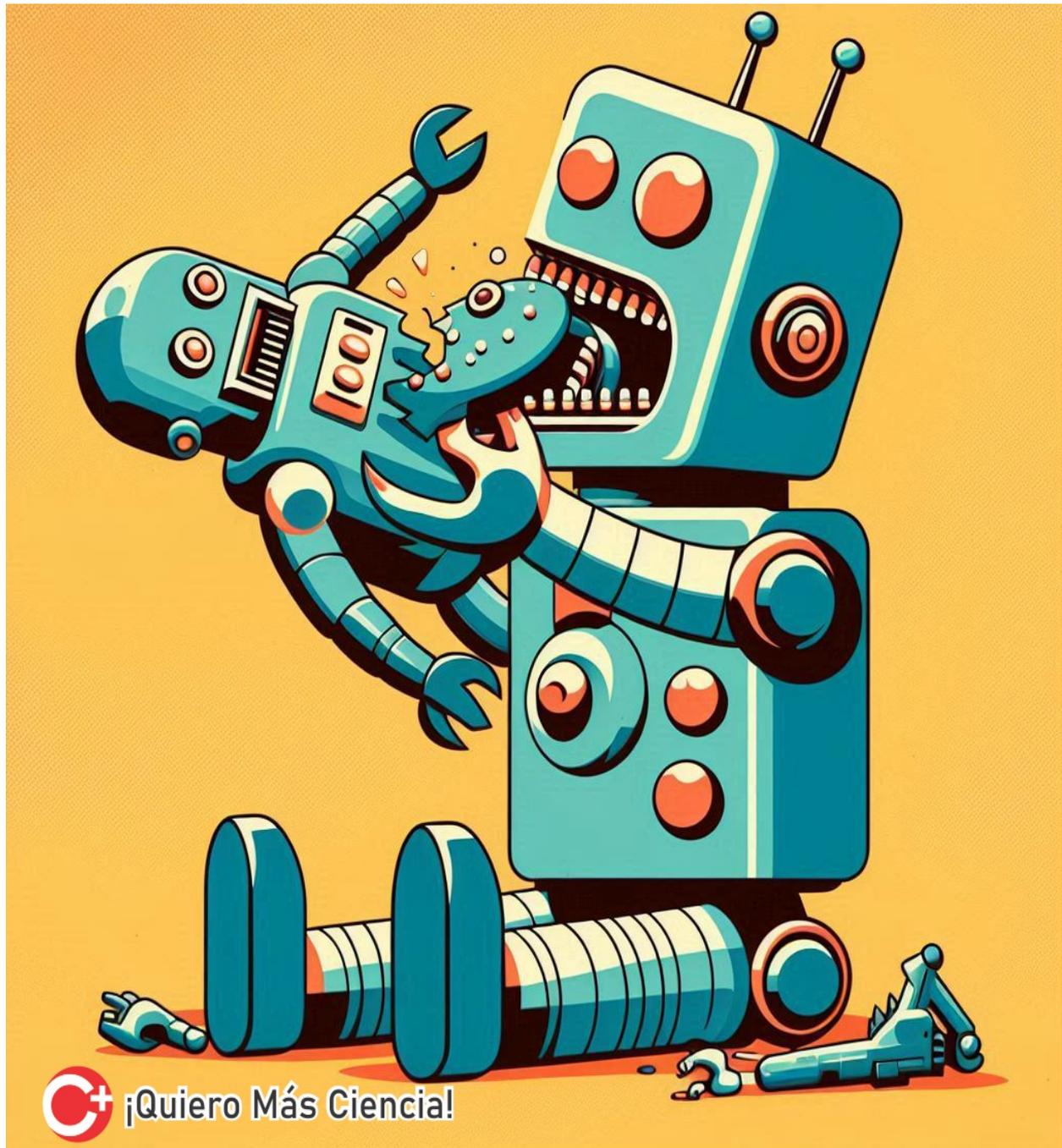
Description

Un modelo basado en datos exclusivamente sintéticos corre el riesgo de sufrir un colapso de la IA, perdiendo utilidad y confiabilidad en su proceso de toma de decisiones.

CONTENIDOS

La preocupación por la endogamia digital es un problema concreto

El colapso de los modelos de inteligencia artificial (IA) no es una idea sacada de la ciencia ficción. Sino una preocupación tangible entre los expertos en tecnología. Desde 2023, la discusión sobre este fenómeno ha ido ganando terreno. Cada vez más investigadores advierten sobre los peligros de entrenar IA utilizando datos generados por otras IA. La analogía con la endogamia es evidente: al igual que en los sistemas biológicos, [la falta de diversidad en los datos de entrenamiento puede llevar a una degradación en la calidad de los resultados](#). Esta es una de las principales preocupaciones que deben abordarse para garantizar la viabilidad de las futuras generaciones de IA.



Investigadores advierten que un colapso en los modelos de IA podría surgir si continúan utilizando datos generados por IA sin intervención humana adecuada.

Los datos humanos son insustituibles en el entrenamiento de IA

El aprendizaje automático, pilar de la IA moderna, depende en gran medida de datos de alta calidad. Según Aaron J. Snoswell, investigador en responsabilidad de la IA en la Universidad Tecnológica de Queensland, “sin datos humanos, los modelos de IA se vuelven cada vez más tontos”. Aunque la tentación de utilizar datos generados por IA es grande, debido a su bajo costo y la ausencia de controversias éticas, el riesgo de un “[entrenamiento regurgitivo](#)” es significativo. Este es un fenómeno, [donde los modelos aprenden de datos previamente generados por otras IA](#)

. El efecto resulta en una disminución progresiva en la utilidad, honestidad y diversidad cultural de las respuestas, Lo que podría comprometer gravemente la efectividad de estas tecnologías.

El colapso de la IA: La disminución de la calidad en los datos es inevitable sin intervención

Los intentos por filtrar contenido generado por IA han demostrado ser insuficientes. Las grandes tecnológicas, como OpenAI y Google, ya invierten cantidades considerables de recursos en la limpieza y filtrado de datos, eliminando hasta el 90% del contenido recopilado en algunas ocasiones. Sin embargo, Snoswell advierte que “será cada vez más difícil distinguir entre datos humanos y datos sintéticos”. Esto no solo implica un aumento en los costos y esfuerzos para mantener la calidad de los datos, sino también una potencial saturación de contenido de baja calidad que podría sofocar la innovación y reducir la capacidad de los modelos de IA para ofrecer resultados precisos y culturalmente relevantes.

Te Puede Interesar:

Investigación sobre los efectos del entrenamiento con datos de IA

La investigación realizada en 2023 reveló que [los modelos de IA entrenados únicamente con datos generados por otras IA mostraban una significativa pérdida de precisión y diversidad en sus respuestas](#). Snoswell señaló que “el aprendizaje de segunda generación es como copiar una copia defectuosa”, lo que lleva a una degradación acumulativa en la calidad de las respuestas de los modelos. Esta investigación destaca la importancia de mantener una fuente constante de datos humanos de alta calidad para evitar el colapso de los modelos. Es esencial continuar explorando métodos para integrar datos humanos frescos en el entrenamiento de IA y evitar la dependencia excesiva de contenido generado artificialmente.

El colapso de la IA: El agotamiento de los datos humanos podría ocurrir pronto

Uno de los mayores [desafíos que enfrenta la industria de la IA](#) es la escasez de nuevos datos humanos. Las estimaciones actuales sugieren que el conjunto de datos de texto generado por humanos podría agotarse tan pronto como en 2026, lo que plantea serias dudas sobre la sostenibilidad de la IA generativa. Empresas como OpenAI están buscando alianzas con gigantes de la industria que poseen grandes bases de datos propietarias, como [Shutterstock](#) y [Associated Press](#), para garantizar un flujo continuo de datos humanos. Sin embargo, esto solo puede ser una solución temporal, ya que la producción de nuevos datos humanos es limitada y no puede mantenerse al ritmo del crecimiento exponencial de la demanda de entrenamiento para modelos de IA.



El colapso de la IA podría ser inminente si no se toman medidas para asegurar una fuente constante de datos humanos de alta calidad.

Para seguir pensando

A medida que la cantidad de contenido generado por IA sigue aumentando, es crucial que la industria tecnológica reconozca la importancia de los datos humanos para el desarrollo continuo de la [inteligencia artificial](#). La homogeneización de las respuestas de IA y la posible pérdida de diversidad cultural no solo amenazan la calidad de los resultados, sino también la integridad del espacio digital. Como concluye Snoswell, “necesitamos proteger las interacciones humanas y los datos humanos, no solo por el bien de la tecnología, sino por el bien de la sociedad en su conjunto”. La preservación y el uso adecuado de los datos humanos serán fundamentales para evitar el

colapso de los modelos de IA y garantizar que estas tecnologías sigan siendo útiles, éticas y culturalmente relevantes en el futuro.