

A decorative graphic of a circuit board with yellow lines and circular components, framing the central text.

presentado en

Pista 2

Claves para comprender la terminología de la IA (y detectar las tonterías)

Qué veremos en este módulo

- Una taxonomía de la “IA B.S.” (afirmaciones exageradas o engañosas).
- Afirmaciones sobre la precisión de la IA.
- Seguridad de la IA y sus narrativas pesimistas.
- Comparaciones entre la IA y las capacidades humanas.



Gen-AI tools deliver 'unprecedented, near-perfect' data accuracy

Latest results deliver 99% accuracy, ready for enterprise decisioning following benchmark test.

📅 3 months ago Posted in

Precisión

PR Newswire

Seekr and Intel Collaborate to Deliver Trusted, Responsible AI Solutions for Enterprise Customers

NEWS PROVIDED BY
Seekr Technologies →
Apr 16, 2024, 22:13 ET

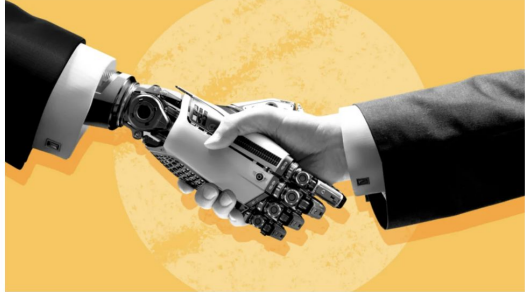
A new collaboration of compute and responsible AI combine to accelerate the deployment of foundation models with less error and bias.

Seguridad de la IA

CNN Health Watch Listen Live TV Sign in

AI may be as effective as medical specialists at diagnosing disease

By Jack Guy, CNN
🕒 3 minute read · Published 7:23 AM EDT, Wed September 25, 2019

A robotic hand is shaking a human hand. The background is a bright yellow circle.

IA > = Humano > =

"Mágico"

"Innovador"

"Revolucionario"

**"99 % de
precisión"**

"Sin errores"

**"Ahorros
enormes"**

**"Aumento de la
productividad
en un 50 %"**

PRECISIÓN DE LA IA: Preguntas que hay que plantearse

¿Quién ha calculado esta
cifra?

¿La propia empresa? ¿Un
estudio revisado por pares?
¿Investigadores contratados
por la empresa?

- **Hippo:** Hippo is a home insurance company that uses AI to identify fraudulent claims with 99% accuracy.

Forbes

FORBES > INNOVATION

BREAKING

New Tool Can Tell If Something Is AI-Written With 99% Accuracy

Arianna Johnson Forbes Staff

Johnson is a reporter on the Forbes news desk who covers explainers.

Follow

PRECISIÓN DE LA IA: Preguntas que hay que plantearse

¿Cuál es la referencia con la que se está comparando?

Ejemplo: si el método anterior tenía una precisión del 96 %, un 97 % puede no suponer una gran mejora.

- **Hippo:** Hippo is a home insurance company that uses AI to identify fraudulent claims with 99% accuracy.

Forbes

FORBES > INNOVATION

BREAKING

New Tool Can Tell If Something Is AI-Written With 99% Accuracy

Arianna Johnson Forbes Staff

Johnson is a reporter on the Forbes news desk who covers explainers.

Follow

PRECISIÓN DE LA IA: Preguntas que hay que plantearse

¿Cuáles son las tasas de falsos positivos, falsos negativos y precisión?

Existen muchas métricas diferentes para medir la «precisión».

- **Hippo:** Hippo is a home insurance company that uses AI to identify fraudulent claims with 99% accuracy.

Forbes

FORBES > INNOVATION

BREAKING

New Tool Can Tell If Something Is AI-Written With 99% Accuracy

Arianna Johnson Forbes Staff

Johnson is a reporter on the Forbes news desk who covers explainers.

Follow

PRECISIÓN DE LA IA: Preguntas que hay que plantearse

¿Se ha probado la precisión en todos los grupos? ¿La precisión difiere entre los grupos?

Un modelo puede tener una precisión del 97 % para un grupo de personas, mientras que para otro solo tiene una precisión del 80 %.

- **Hippo:** Hippo is a home insurance company that uses AI to identify fraudulent claims with 99% accuracy.

Forbes

FORBES > INNOVATION

BREAKING

New Tool Can Tell If Something Is AI-Written With 99% Accuracy

Arianna Johnson Forbes Staff

Johnson is a reporter on the Forbes news desk who covers explainers.

Follow

PRECISIÓN DE LA IA: Preguntas que hay que plantearse

¿A cuántas personas se aplica
la herramienta?

Una precisión del 99 % (y una
tasa de error del 1 %) para
decenas de millones de
personas puede significar
100 000 decisiones erróneas.

- **Hippo:** Hippo is a home insurance company that uses AI to identify fraudulent claims with 99% accuracy.

Forbes

FORBES > INNOVATION

BREAKING

New Tool Can Tell If Something Is AI-Written With 99% Accuracy

Arianna Johnson Forbes Staff

Johnson is a reporter on the Forbes news desk who covers explainers.

Follow

PRECISIÓN DE LA IA: Preguntas que hay que plantearse

¿Se ha probado el modelo en
el «mundo real»?

Un modelo puede tener una
alta precisión en los datos de
pruebas internas (con los que
también puede haberse
entrenado), pero una baja
precisión una vez aplicado en
el mundo real.

- **Hippo:** Hippo is a home insurance company that uses AI to identify fraudulent claims with 99% accuracy.

Forbes

FORBES > INNOVATION

BREAKING

New Tool Can Tell If Something Is AI-Written With 99% Accuracy

Arianna Johnson Forbes Staff

Johnson is a reporter on the Forbes news desk who covers explainers.

Follow

Las afirmaciones
sobre la
precisión suelen
tener una letra
pequeña
reveladora

🕒 JANUARY 25, 2023

✓✓ Editors' notes

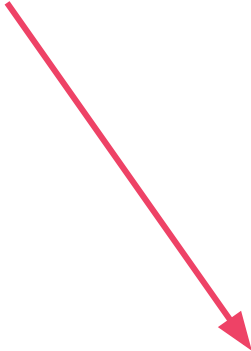
A new AI tool can predict mosquitoes' ages with 98% accuracy to speed malaria research

by SciDev.Net

However, the scientists stress that further research is needed since the study looked at only one specific type of mosquito, *Anopheles arabiensis*, obtained from only two countries.

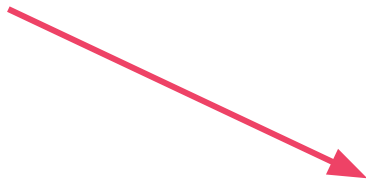


Las afirmaciones sobre la precisión suelen tener una letra pequeña reveladora



ShotSpotter collects are actual gunfire. ShotSpotter, Inc. continues to claim a supposed 97% "accuracy" rate. To that end, they paid a consultant to "audit" this number. But the audit actually confirmed that this is not an accuracy rate at all, reflecting no actual testing of the system. Instead, it showed that ShotSpotter just assumes that all of its alerts are gunshots unless local police happen to file a complaint with the company reporting a potential error. ShotSpotter's supposed "accuracy" claims are actually just a tally of customer complaints. And, according to Manes, CPD does not report any of

Las afirmaciones sobre la precisión suelen tener una letra pequeña reveladora



Amr Awadallah, the chief executive of Vectara, warns that its chatbot software doesn't always tell the truth. Cayce Clifford for The New York Times

Chatbots May 'Hallucinate' More Often Than Many Realize

When summarizing facts, ChatGPT technology makes things up about 3 percent of the time, according to research from a new start-up. A Google system's rate was 27 percent.



Share full article



By **Cade Metz**

Cade Metz has been watching chatbots hallucinate since 2017.

Published Nov. 6, 2023 Updated Nov. 16, 2023

[Leer en español](#)

≡ FORTUNE

CONFERENCES · A.I.

AI hallucinations will be solvable within a year, ex-Google AI researcher says—but that may not be a good thing: 'We want them to propose things that are weird and novel'

BY **ORIANNA ROSA ROYLE**

April 16, 2024 at 3:10 PM GMT+3



"intenciones"

"racional"

"Sensible"

"IA similar a la humana"

"Supera a los humanos"

"La IA comprende las emociones humanas"

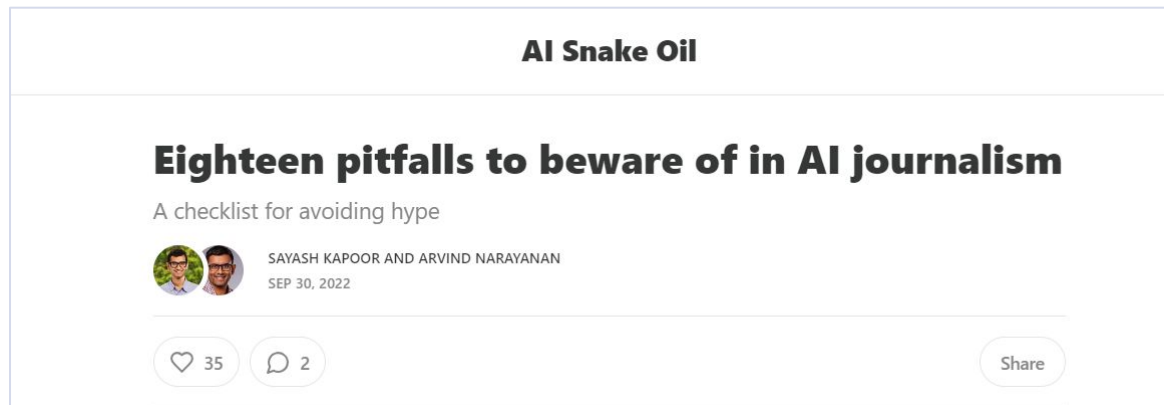
"Objetiva"

Cuando tratamos a la IA como si fuera humana



Es fácil reproducir clichés en nuestros reportajes

El substack AI Snakeoil de Sayash Kapoor y Arvind Narayanan es una lectura obligatoria para los periodistas que trabajan en este ámbito.



Cuatro errores comunes al informar sobre IA

Pitfall 1. Attributing agency to AI: Describing AI systems as taking actions independent of human supervision or implying that they may soon be able to do so.

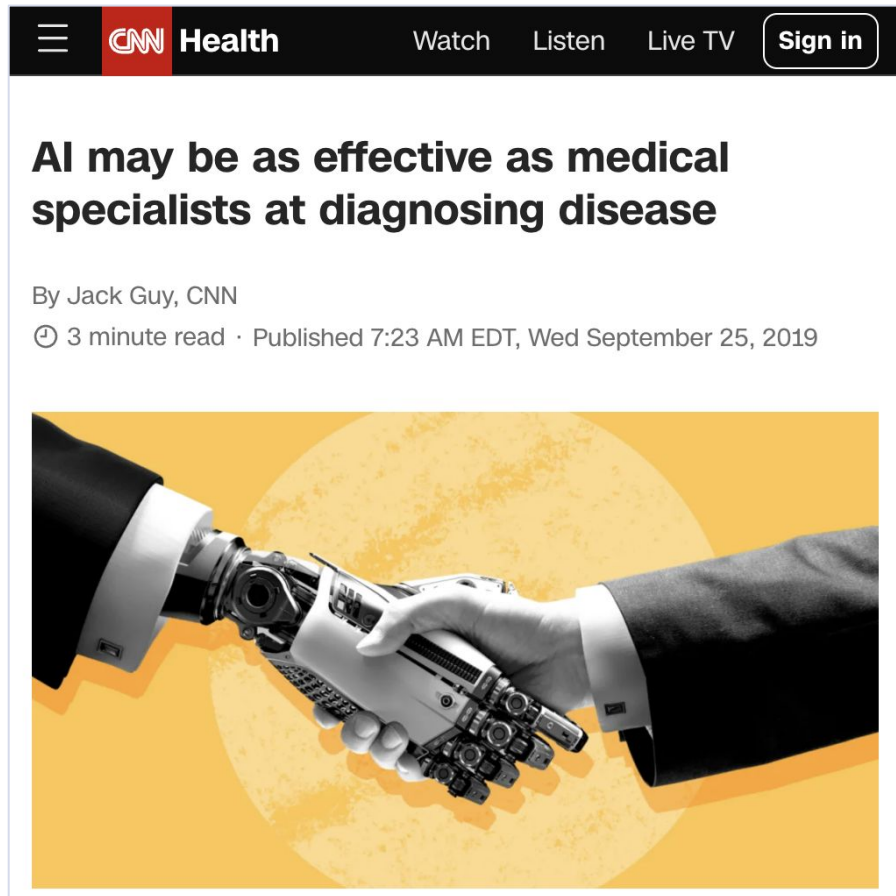
Pitfall 2. Suggestive imagery: Images of humanoid robots are often used to illustrate articles about AI, even if the article has nothing to do with robots. This gives readers a false impression that AI tools are embodied, even when it is just software that learns patterns from data.

Pitfall 3. Comparison with human intelligence: In some cases, articles on AI imply that AI algorithms learn in the same way as humans do. For example, comparisons of deep learning algorithms with the way the human brain functions are common. Such comparisons can lend credence to claims that AI is “sentient”, as Dr. Timnit Gebru and Dr. Margaret Mitchell note in their [recent op-ed](#).

Pitfall 4. Comparison with human skills: Similarly, articles often compare how well AI tools perform with human skills on a given task. This falsely implies that AI tools and humans compete on an equal footing—hiding the fact that AI tools only work in a narrow range of settings.

**Es importante detectar
estos errores en nuestro
trabajo y en las relaciones
públicas de las empresas o
los gobiernos.**

**Por qué muchas
empresas
promocionan sus
productos de IA
dándoles rasgos
humanos.**



**Evitemos la falsa
idea de que la IA
puede sentir o
pensar, o que es
naturalmente
objetiva y racional.**



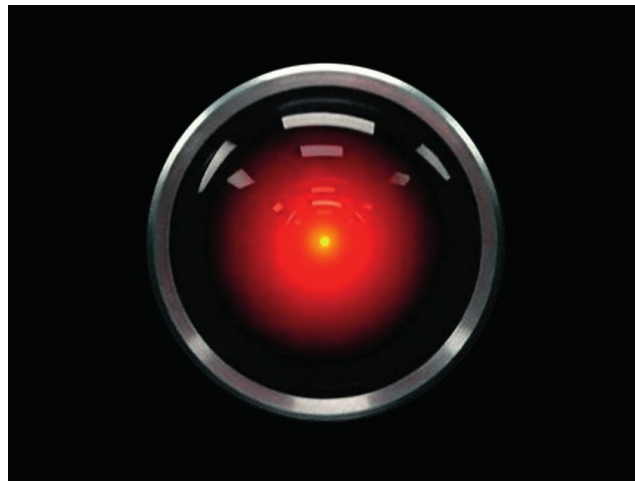
Definiciones

- IA alineada con los valores: codificación de los «valores humanos» en un modelo de IA (normalmente utilizado en el contexto de los LLM).
- Seguridad de la IA*: alinear la IA con los valores para proteger a la humanidad de la amenaza «superhumana» y «existencial» de la IA.

** La seguridad de la IA se utiliza a veces para describir de forma más general los intentos de controlar los daños sociales causados por la IA.*

Del «catastrofismo de la IA» y el tecnosolucionismo

El catastrofismo sobre la IA distrae la atención de los daños actuales y concretos relacionados con la implementación de la IA.



An open letter with signatures from hundreds of the biggest names in tech, including Elon Musk, has urged the world's leading artificial intelligence labs to pause the training of new super-powerful systems for six months, saying that recent advances in AI present “profound risks to society and humanity.”

The letter comes just two weeks after the public release of OpenAI's **GPT-4**, the most powerful AI system ever released, which has led researchers to slash their expectations for when AGI—or artificial general intelligence that surpasses human cognitive ability—will arrive. Many experts fear that, as an AI arms race heats up, humanity is sleepwalking into catastrophe.

Del «catastrofismo sobre la IA» y el tecnosolucionismo

El pesimismo sobre la IA distrae la atención de los daños actuales y concretos relacionados con la implementación de la IA.



TECH BY VICE

'I'm a Doomer': OpenAI's New Interim CEO Wants to Slow AI Progress Down

Emmett Shear has called himself a "doomer" and "AI safetyist" in posts on X, and railed against e/acc ideology.



Listen to this article now

6 min listen

00:00



-05:34

Powered by [Trinity Audio](#)

 By [Jordan Pearson](#)

November 20, 2023, 8:07pm   





AI Snake Oil



Subscribe

Sign in

AI safety is not a model property

Trying to make an AI model that can't be misused is like trying to make a computer that can't be used for bad things



ARVIND NARAYANAN AND SAYASH KAPOOR

MAR 12, 2024

Seguridad de la IA

- Posiciona a la IA como una amenaza existencial y sobrehumana.
- La solución es técnica.

Ética de la IA

- Se centra en los daños actuales de la IA.
- La solución es sociotécnica (es decir, no solo se trata de la tecnología, sino del contexto de cómo y dónde se implementa).

Auditoría de IA

- La auditoría de los sistemas de IA puede ser una herramienta valiosa para la rendición de cuentas, pero debemos mantener un escepticismo saludable frente a ciertas afirmaciones.
- Empresas y gobiernos suelen decir que sus modelos han sido auditados por seguridad o sesgos. Sin embargo, eso no significa que sean éticos, justos o imparciales.
- Antes de dar por válida una auditoría, conviene preguntar: ¿Quién la realizó? ¿Fue interna o externa? ¿Independiente o contratada por la propia empresa? ¿Hubo contradicción o contraste de resultados?

Auditoría de IA

FINANCIAL
ADVISORY SERVICES

Deloitte.

Combating welfare fraud with machine learning

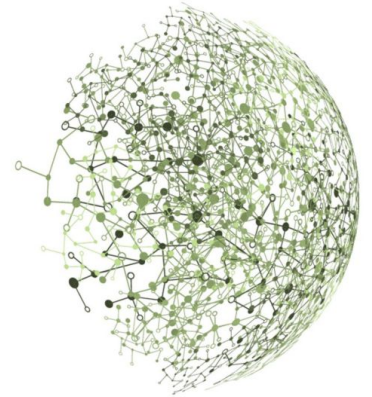
A government organisation in the Netherlands had to contend with cases of fraud. The organisation was responsible for paying allowances to citizens, but hundreds of millions of euros were being paid to ineligible persons. An internal department had identified that processes were not running effectively, and in the aftermath of negative publicity it was clear that the method of tackling fraud needed to be improved.

NEWS > TECHNOLOGY

Dutch scandal serves as a warning for Europe over risks of using algorithms

The Dutch tax authority ruined thousands of lives after using an algorithm to spot suspected benefits fraud — and critics say there is little stopping it from happening again.

Deloitte.



Municipality of Amsterdam

Reporting Analysis Algorithm with the application for a Living allowance

Published version (shortened version for the algorithm register)

MAKING AN
IMPACT THAT
MATTERS
Since 1945

Consejos para ver más allá del bombo publicitario de la IA

- No se crean las afirmaciones de las empresas sin más. Pidan pruebas concretas.
- Entiendan qué métricas utiliza una empresa para respaldar sus afirmaciones de precisión.
- Céntrate en los daños concretos y presentes, en lugar de distraerte con el pesimismo «existencial».
- No hables solo con la empresa, habla con expertos y críticos que cuenten con financiación independiente.

Actividad:
Analizar una
presentación de
relaciones públicas
15 minutos

Una startup ha desarrollado un software de inteligencia artificial que lee radiografías de tórax y afirma tener una precisión del 99 %. El director ejecutivo de la startup afirma que, tras realizar una prueba en un hospital, el software ha demostrado que salvará vidas y reducirá costes, ya que las enfermeras podrán detectar el cáncer de pulmón u otros problemas con la misma eficacia que los médicos expertos.

datos

+

Calcular



Modelos de IA



Aplicaciones

¿Con qué datos se entrenó este software?

¿De quién proceden esos datos?

¿Dieron su consentimiento?

¿Cuáles eran sus características demográficas?

¿Cuánto costó desarrollar el software?

¿Dónde se almacenan los datos de los pacientes?

¿Qué ocurre una vez que se suben a este sistema?

¿Qué hacen exactamente los algoritmos de este sistema?

¿Quién dice que tiene una precisión del 99 %?

¿Se probó el software en condiciones similares a las del mundo real?

¿Qué medidas de control se han implementado para detectar errores?

¿Cómo ahorra dinero el sistema y para quién? ¿Se ha demostrado? ¿Qué opinan las enfermeras y los médicos sobre este sistema?

¿Qué opinan las enfermeras y los médicos sobre este sistema?

¿Pueden los pacientes optar por no participar?

¿Cuánto cuesta el software?
¿Cómo se le paga a la startup?

El gobierno de una ciudad planea utilizar la IA para evaluar las solicitudes de viviendas para personas con bajos ingresos. El gobierno afirma que el sistema aceptará automáticamente las solicitudes legítimas y marcará las fraudulentas para su investigación, lo que eliminará por completo cualquier sesgo del proceso.

datos

+

Calcular



Modelos de IA



Aplicaciones

¿Con qué datos se entrenó este modelo?

¿De quién procedían esos datos?

¿Dieron su consentimiento?

¿Cuáles eran sus características demográficas?

¿Cómo se entrenó este sistema?

¿El gobierno está colaborando con una empresa privada, por ejemplo, un proveedor de servicios en la nube, para entrenar este modelo?

¿Con qué frecuencia se actualizará?

¿Qué tipo de algoritmo se está utilizando?

¿Se puede acceder al algoritmo?

¿Se ha auditado el software para comprobar su precisión y sesgos?

¿Qué fiabilidad tiene el software? ¿Es suficiente para este fin?

¿Qué tareas realiza exactamente y qué tienen que seguir haciendo los humanos?

¿Quién lo utiliza y con quién?

Desde la implementación del software, ¿ha habido algún cambio en quién es aceptado o rechazado?

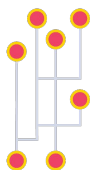
¿Realmente necesita la IA, o existe alguna alternativa más económica, transparente y fiable?

Recursos

- [AI Snake Oil](#)
- [Eighteen pitfalls to beware of in AI journalism](#)
- [How to report better on artificial intelligence - Columbia Journalism Review](#)
- [3 Easy Ways to Evaluate AI Claims - IEEE Spectrum](#)
- [Emily M. Bender – Medium](#)
- [On the Dangers of Stochastic Parrots | Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency](#)



AI Spotlight Series



**The AI
Accountability
Network**



Pulitzer Center