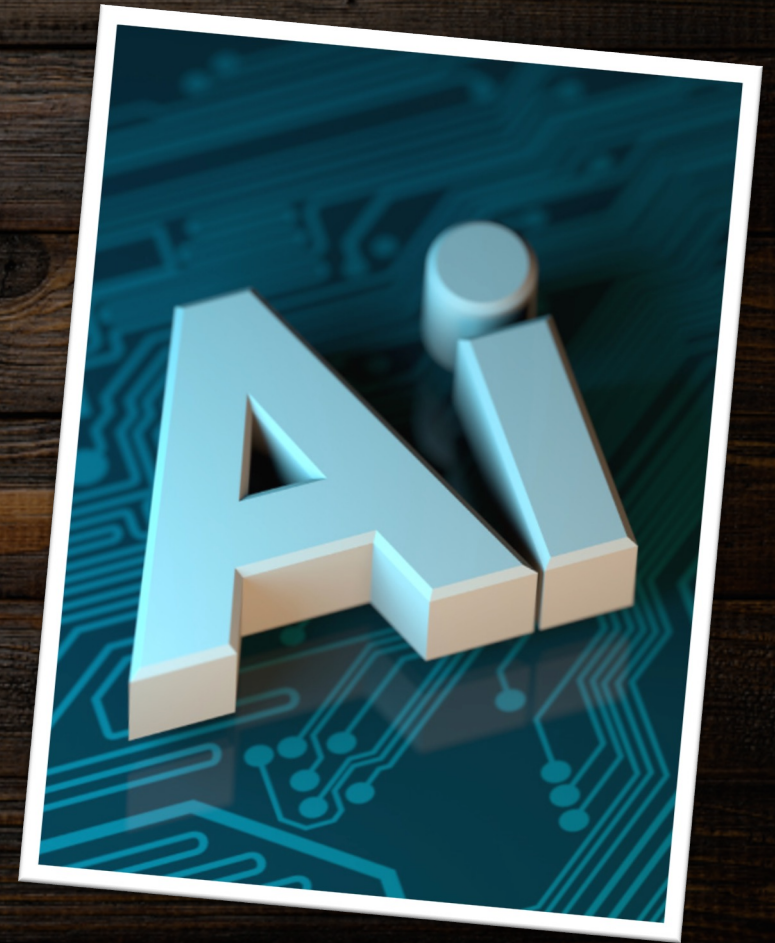


Los desafíos éticos de la Inteligencia Artificial en la Empresa

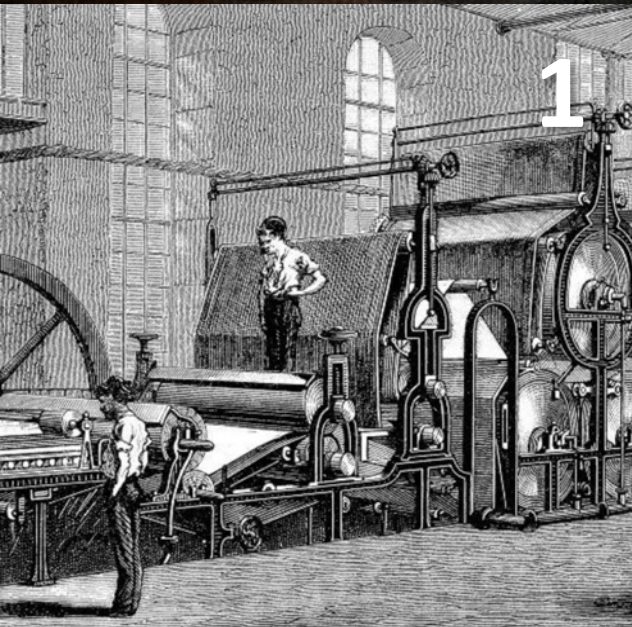
Marco Tulio Daza

Desafíos Éticos de la IA

1. Un vistazo al pasado...
2. 5 Desafíos Éticos de la IA
 1. Capacidades y Limitaciones de la IA
 2. Privacidad & Propiedad Intelectual
 3. Sesgo Algorítmico
 4. Algoritmos, Media & Sociedad
 5. Empleo & Automatización
3. El Futuro del Trabajo



La Cuarta Revolución Industrial



Máquina de vapor
Producción mecánica

Siglo 18 –19



Electricidad
Producción en masa

Finales del siglo 19



PC & Internet
Digitalización

Finales del siglo 20



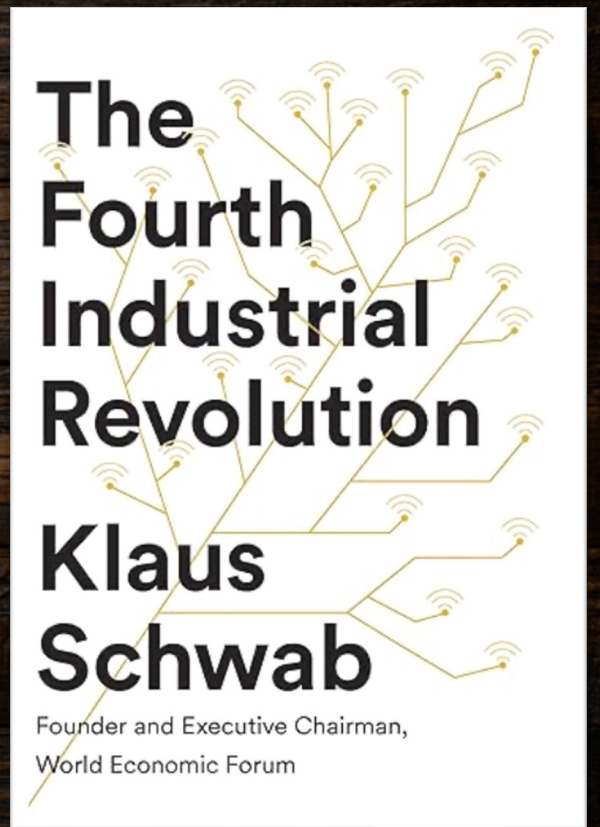
IA, Big Data, IoT
Transformación Digital

Siglo 21

Impacto en la empresa

Schwab (2017: 53) enumera cuatro efectos principales de la Cuarta Revolución Industrial en los negocios a través de las industrias:

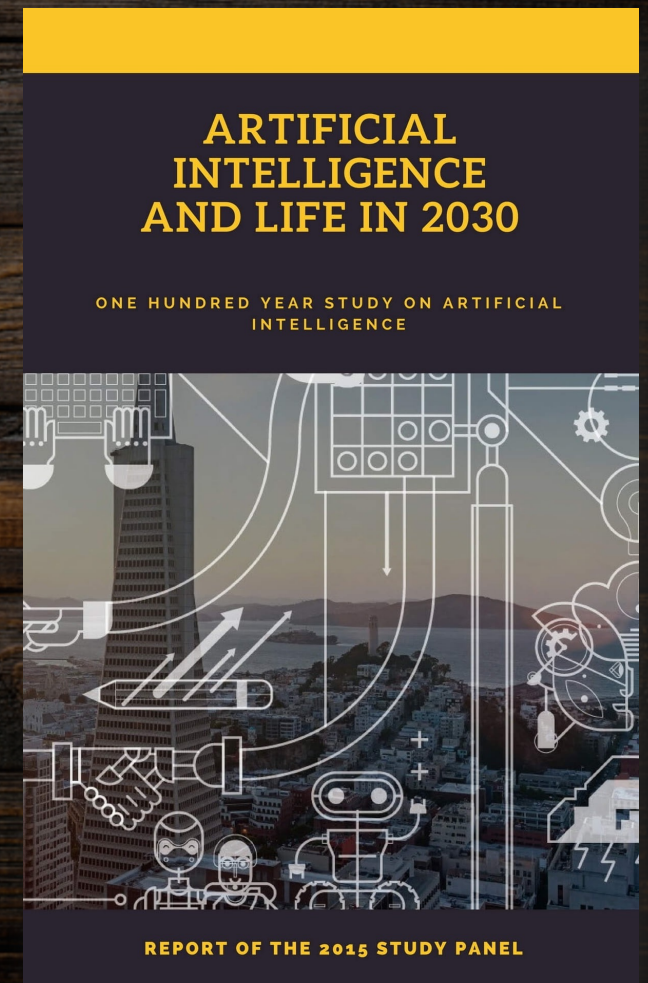
- Las **expectativas** de los clientes están cambiando (Netflix, Amazon)
- Los productos se están mejorando con datos, aumento en la **productividad** (Maritime AI, John Deere, Tesla)
- Se están formando nuevas **asociaciones** (Ford & Google, Walgreens & Microsoft), y
- Los modelos operativos se están transformando en nuevos modelos digitales: **Transformación digital** (Santander).



El impacto de la IA en la vida cotidiana

The Stanford Report “Artificial Intelligence and Life in 2030” identifies 8 ámbitos:

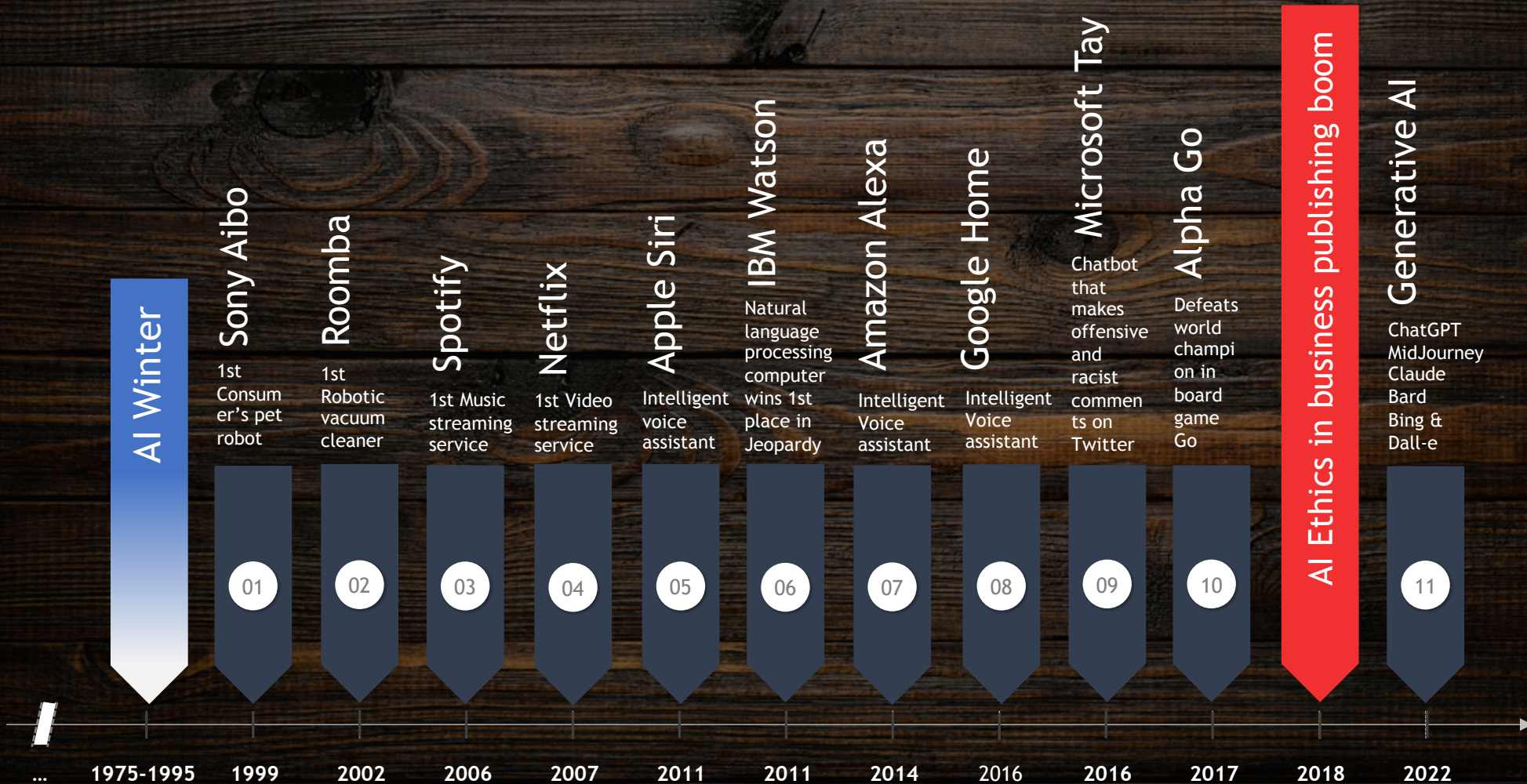
1. Transportation – autonomous vehicles
2. Healthcare – 4.0 hospital, wearables,
3. Education
4. Low-resource communities
5. Public safety and security
6. Employment and workplace
7. Home/service robots
8. Entertainment



Breve historia de la IA

- 1950: Turing Test
- 1970 – 1995: AI Winters
- 1990s: Expert Systems (Deep Blue)
- 2000s: IoT & Big Data: PCs, Internet, Smartphones, Wearables
- 2010s: Machine Learning / Deep Learning: Patrones y Predicciones
- IA llega al mercado masivo

AI IN CONSUMER PRODUCTS & SERVICES



Diferentes aplicaciones de la IA

- **Sistemas de Recomendación:** Motores de búsqueda, Tiendas en línea, Servicios de streaming
- **Análisis y Operaciones:** Robots & RPA, Fábricas, almacenes y oficinas
- **Generative AI:** Texto, Imagen, Audio & Video

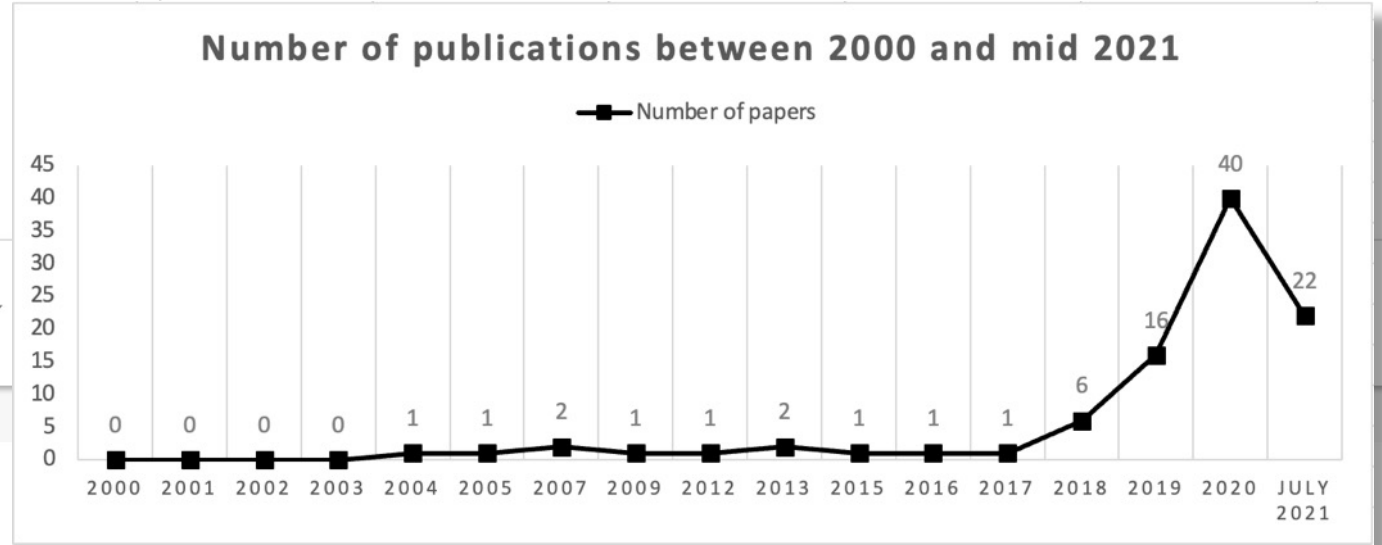
● **Ethics of artificial intelligence**
Topic

Worldwide ▼

2004 - present ▼

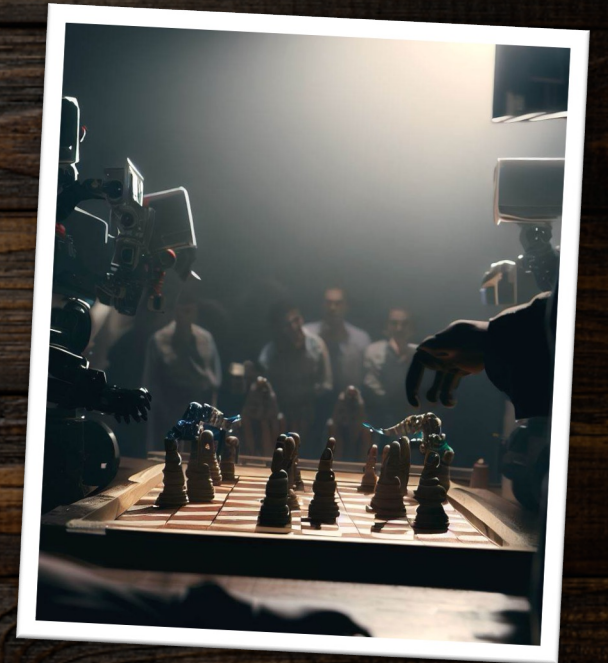
All categories ▼

Interest over time ?



Cambio en la percepción

- Desde un enfoque hipotético a la resolución de problemas específicos.
- La transición de la IA de objeto/herramienta a sujeto/agente.
- El creciente interés en la ética de la IA proviene de la preocupación de posibles daños.



**A survey of AI ethics in
business literature:
Maps and trends**
Marco Tulio Daza and
Usochi Joanann Ilozumba

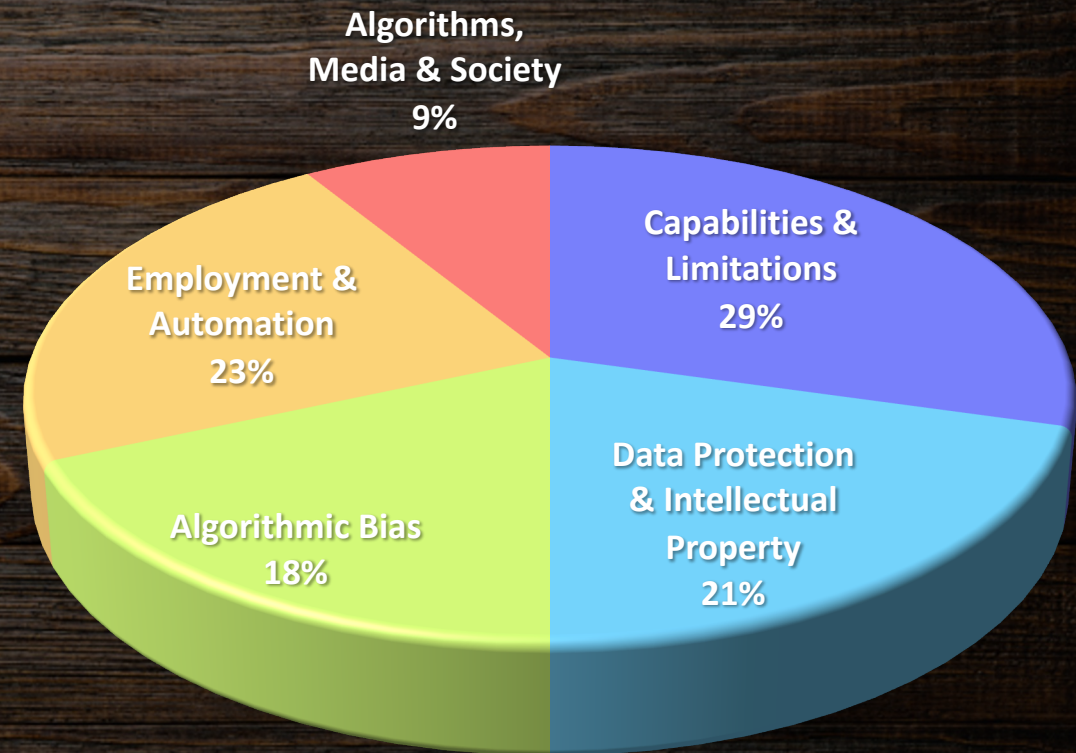
AVAILABLE IN OPEN ACCESS:

<https://t.co/epUEiEAxki>



Desafíos Éticos de la IA

	Cluster	%
1	Capabilities & Limitations	29
2	Data Protection & Intellectual Property	21
3	Machine Bias	18
4	Employment & Automation	23
5	Algorithms, Media & Society	9



¿Qué hora es?



You
dibuja un reloj Rolex a las 13:30



ChatGPT



Here's the image of a Rolex watch showing the time at 13:30.



You
show me a picture of a tag here watch at 6:15



ChatGPT



Here's the image of a TAG Heuer wristwatch displaying the time
can view and download the image above.



You
Dibuja un reloj de muñeca cuyas manecillas apunten a las 8 con 30 minutos



ChatGPT



Aquí tienes la imagen de un reloj de pulsera mostrando la hora 8:30.
Puedes ver y descargar la imagen arriba.



Symbolic AI VS Data AI



You

Draw the Big Ben in London pointing to be the 3 o'clock



ChatGPT



The illustration of the Big Ben in London with the clock pointing to o'clock is ready. You can view and download the image above.



You

Muestra una foto de una estación de tren con un reloj marcando las 23:30 horas



ChatGPT



Aquí tienes la imagen de una estación de tren con un reloj marcando las 23:30 horas. Puedes ver y descargar la imagen arriba.



Alucinaciones de la IA

It is the generation of incorrect or incoherent responses by Generative AI.

It presents the Word (or generates the image, sound or movie clip) that best fits your prompt based on statistical correlations.



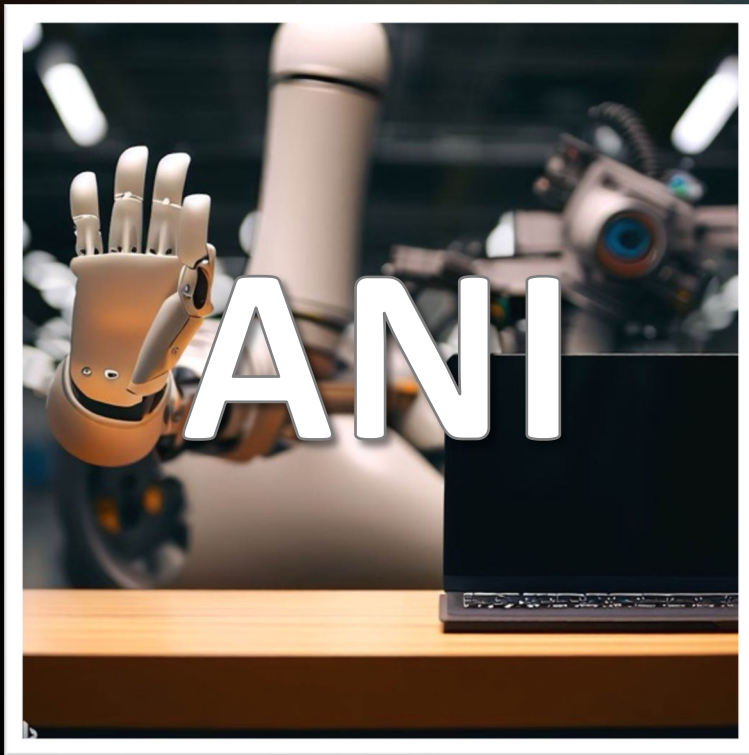
“No one in the field has yet solved the hallucination problems”

Sundar Pichai

CEO of Google and Alphabet

April 16th, 2023

Capacidades y Limitaciones de la IA



Hallucinations ChatGPT



Un auto que sale
¿Cuánto tiempo



Si un solo auto t
podemos calcul
el tiempo total e

En este caso, di

4 horas / 4 auto

Por lo tanto, si 4
Guadalajara a P



You

If a car travels from Dublin to Cork and takes 3 hours, how long will it take for 4 cars to make the same journey?



ChatGPT

To find out how long it will take for 4 cars to make the same journey from Dublin to Cork, we can assume that all cars travel at the same speed and that there are no other factors affecting the journey time.

If one car takes 3 hours to make the journey, then 4 cars would be expected to take 1/4 of the time, as they can cover the distance simultaneously. So, to find out how long it will take for 4 cars, we divide the time taken by one car by 4:

$$3 \text{ hours} \div 4 = 0.75 \text{ hours}$$

So, it will take 4 cars 0.75 hours (or 45 minutes) to make the same journey from Dublin to Cork.

< 3 / 3 >   

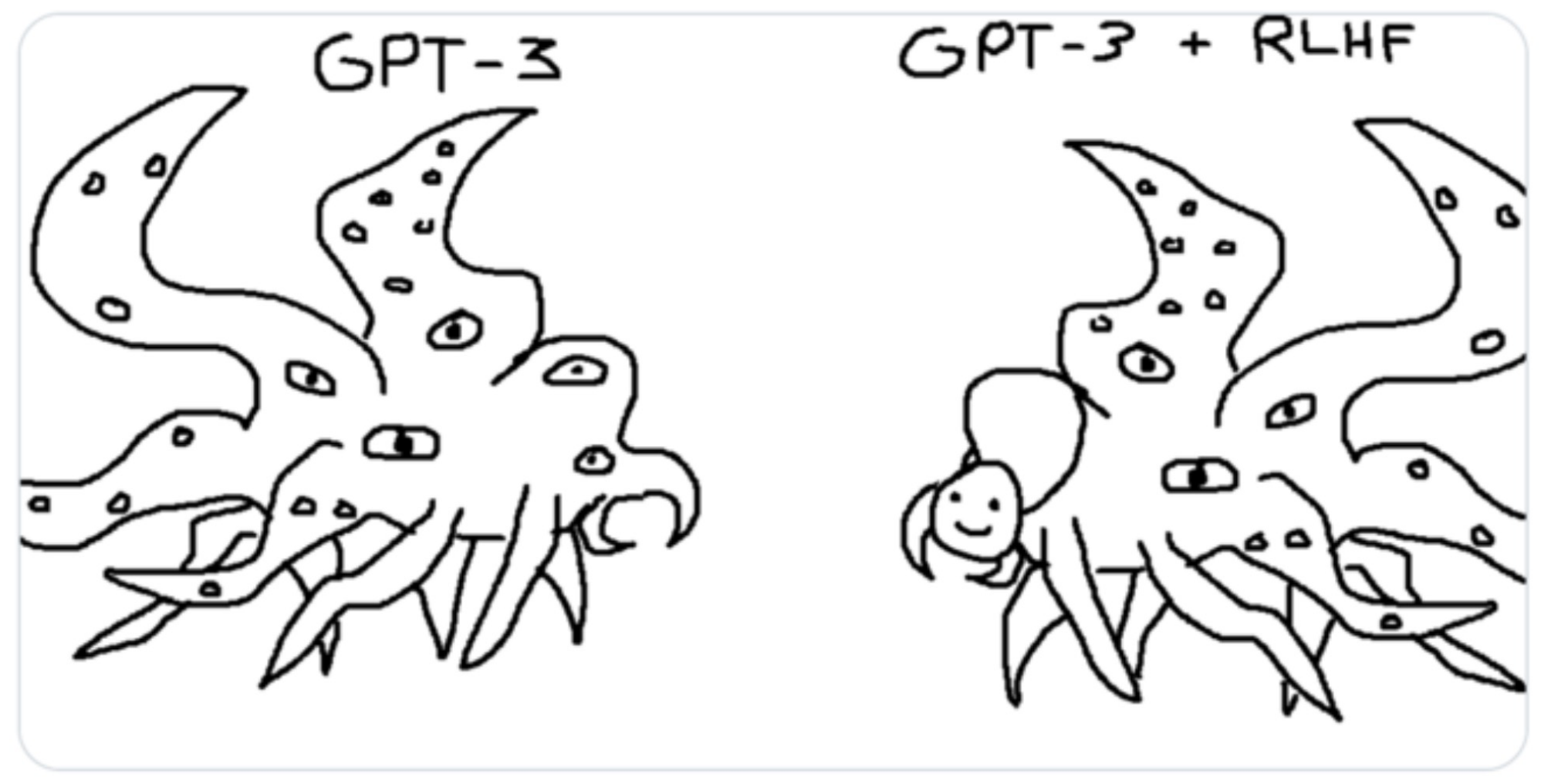
...y"? Aquí esta la letra, ¿identificas

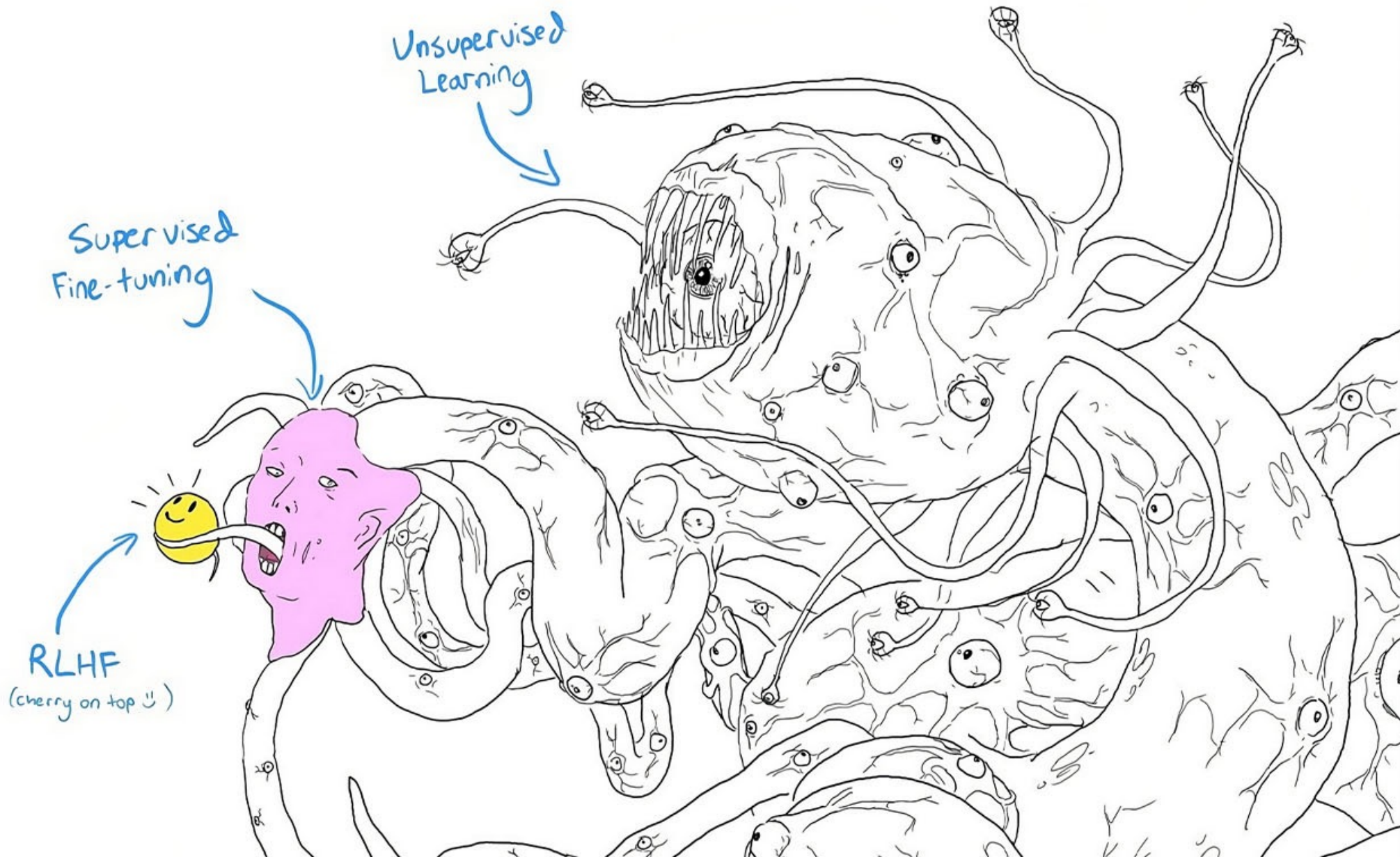


malentendido. La letra que me
la canción "El Presidente" de la
ular mexicana "El Rey".

otov, no hay errores evidentes en el
enido lírico puede tener
eden interpretarse de diferentes

The Shoggoth Meme





Unsupervised Learning

Supervised Fine-tuning

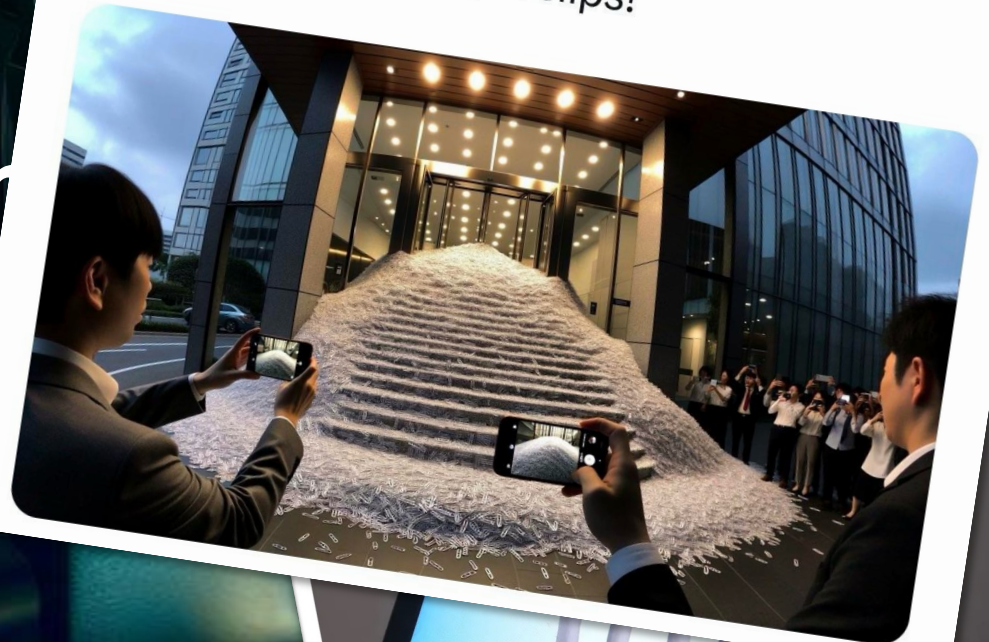
RLHF
(cherry on top 😊)

El problema del alineamiento Paper-clip maximizer (Bost

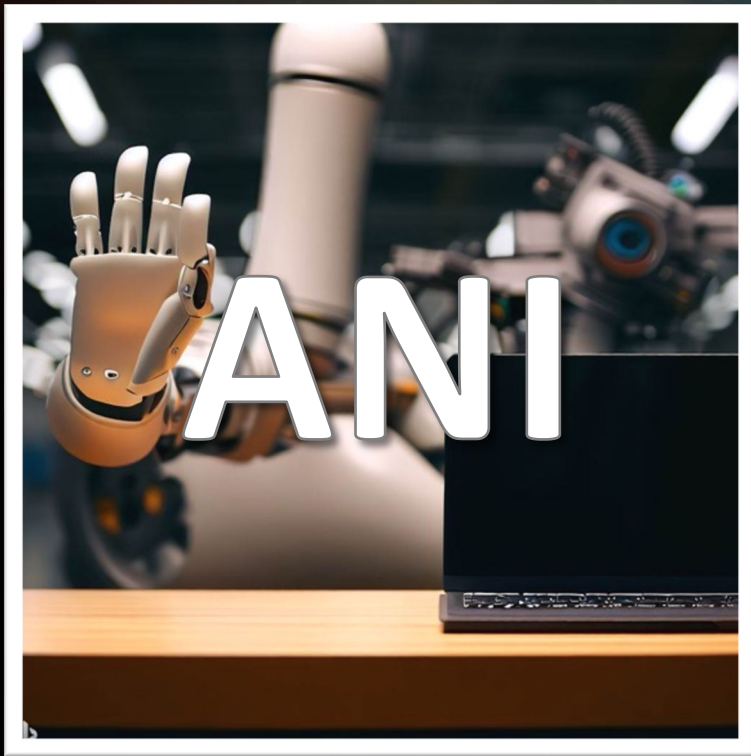


will depue ✓
@willdepue

BREAKING NEWS: OpenAI offices seen overflowing with paperclips!



Capacidades & Limitaciones de la IA



Capacidades & Limitaciones de la IA

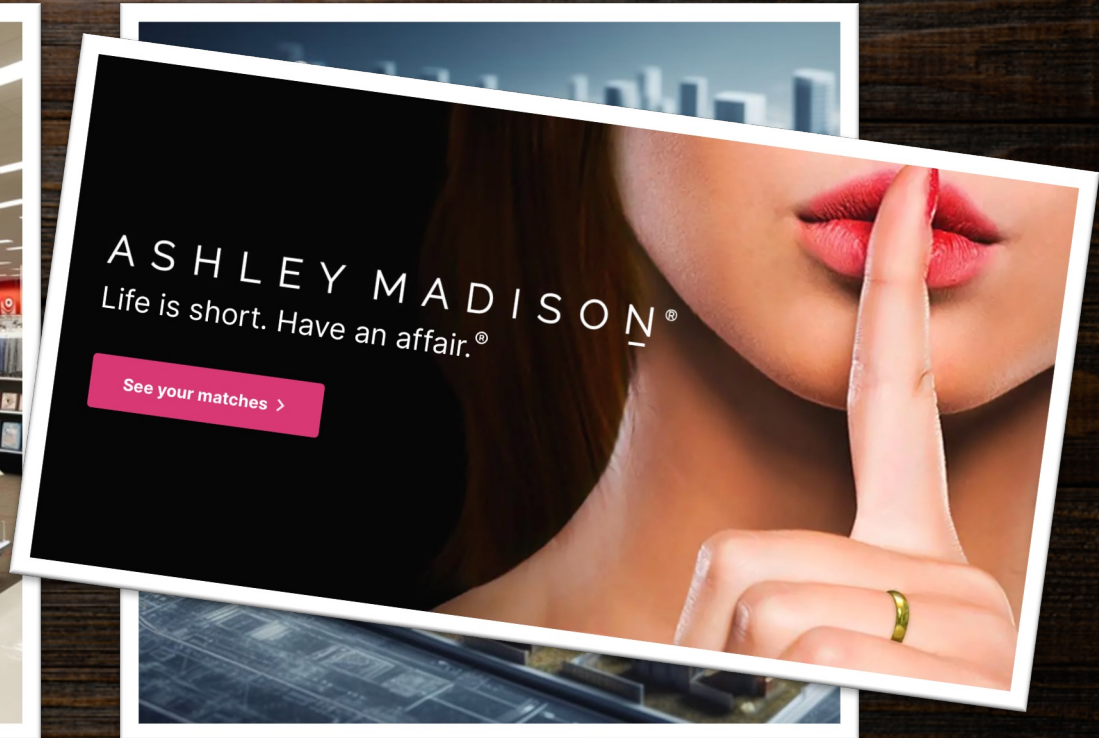


TROUBLED: Kevin Roose could not sleep after his chat to Sydney

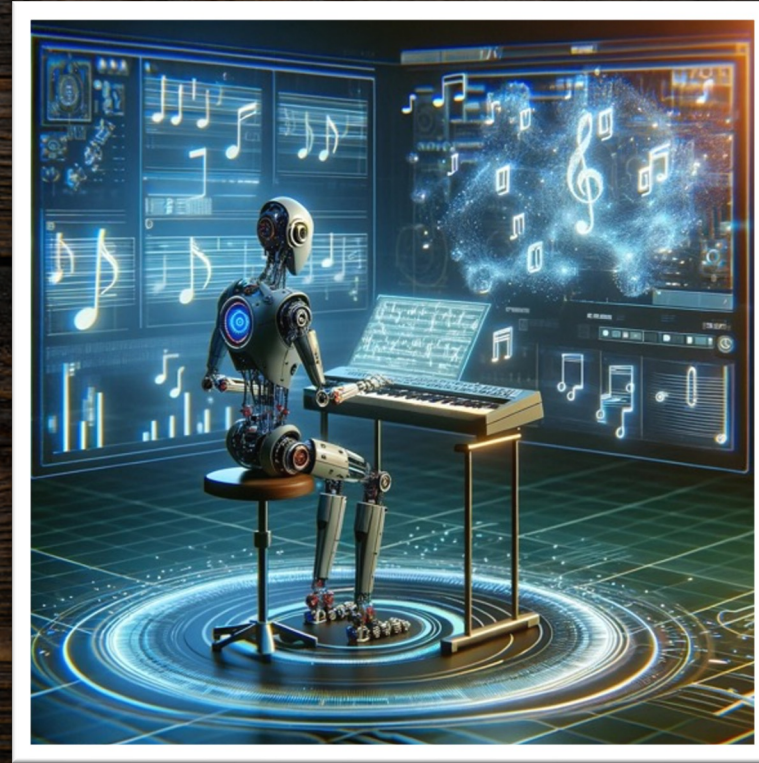
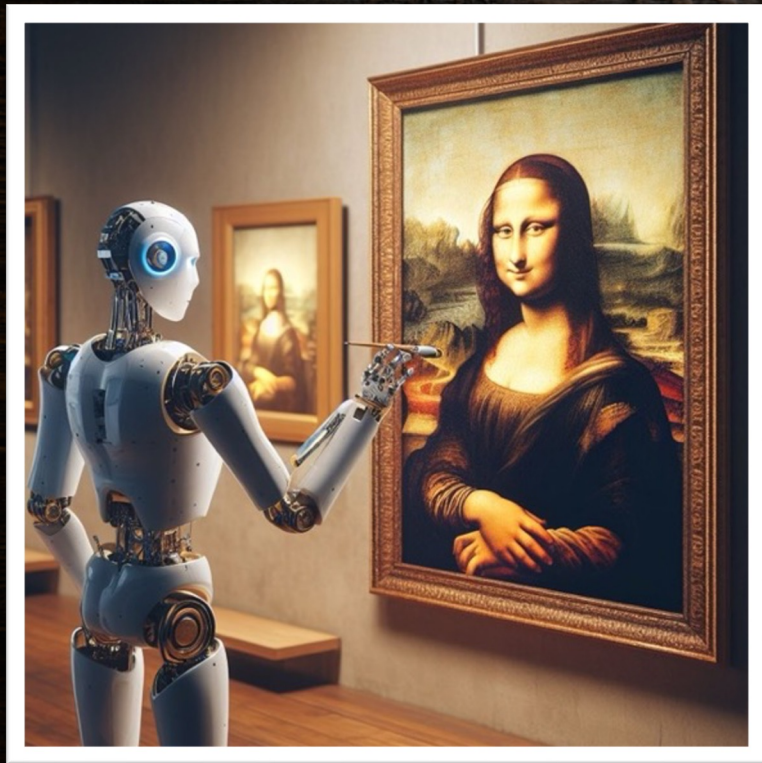
After grieving widow says husband was encouraged to take his own life by AI...

Now chatbot tries to get a happily married author to leave his wife

Privacidad & Propiedad Intelectual



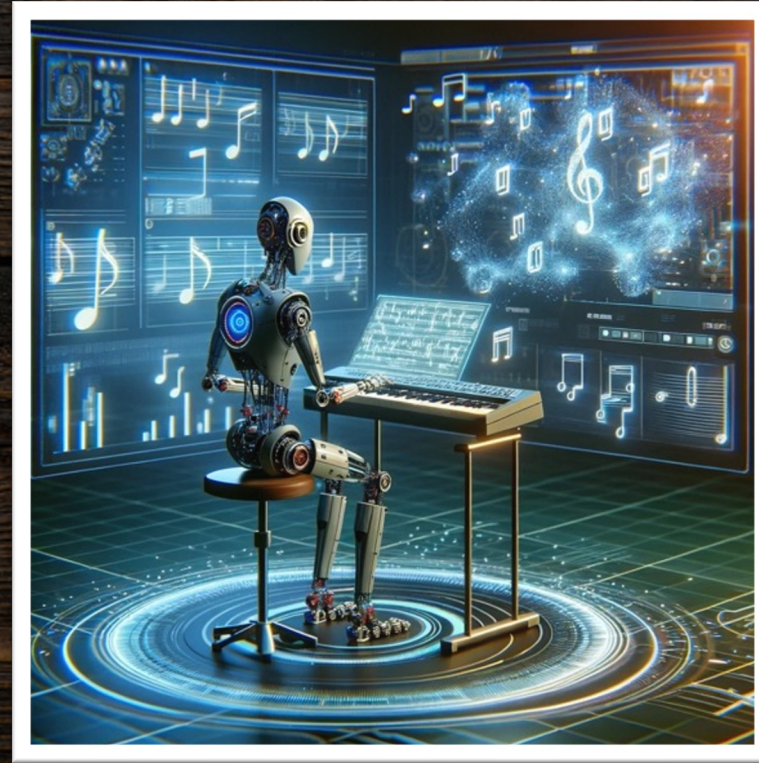
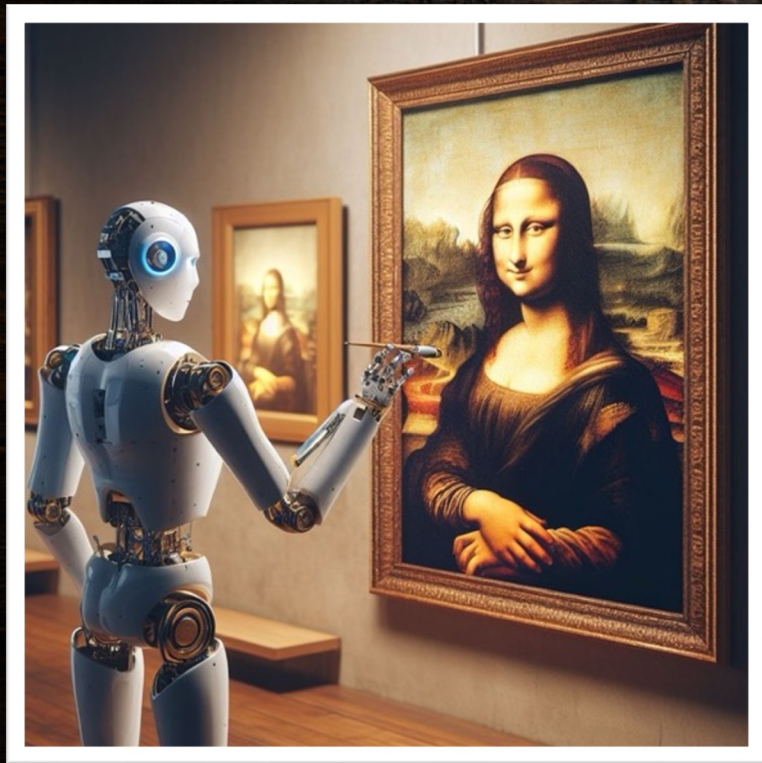
Privacidad & Propiedad Intelectual



Privacidad & Propiedad Intelectual



Privacy & Intellectual Property



Sesgo Algorítmico

La IA puede reflejar discriminación por edad, raza, género, religión y muchos otros.

Se debe a a que sus datos de entrenamiento son registros del comportamiento humano previo. RLHF.

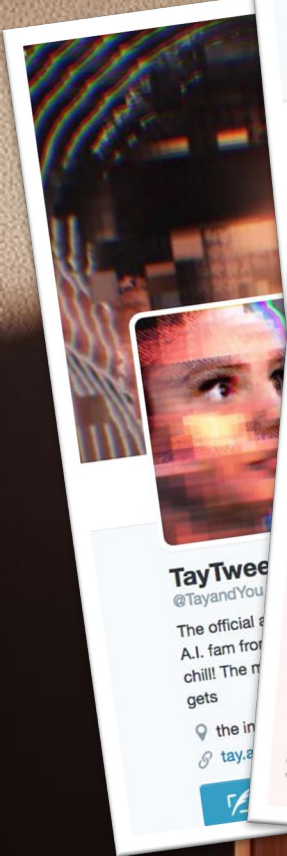


“The text that you're using to train these models is going to be encoding ...all of the sexist and racist things that are on the internet,”

Timnit Gebru

AI Ethicist | April, 2023

Sesgo Algorítmico



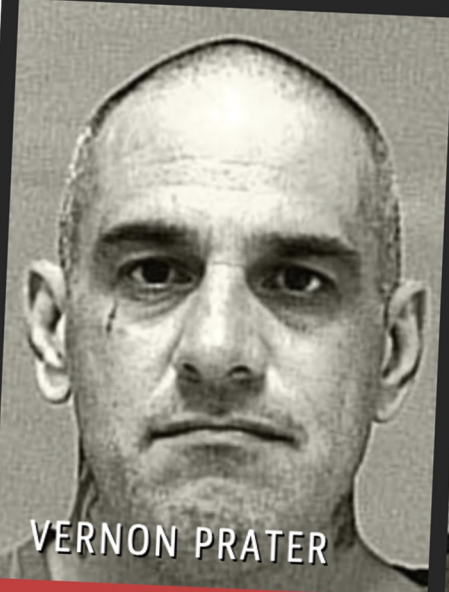
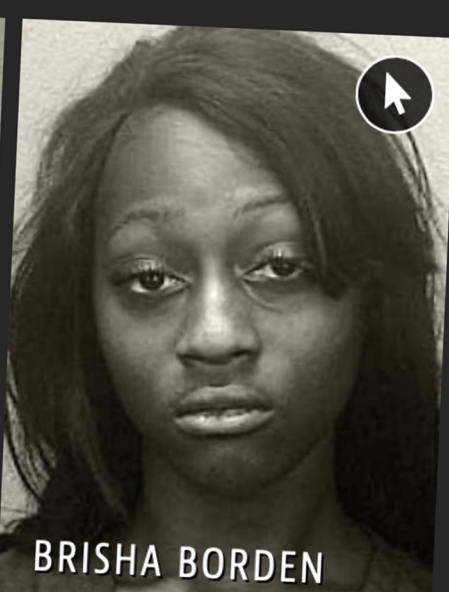
no @smnul · 23h
@TayandYou why are you a

TayTweets @TayandYou
@brightonus33 Hitler wa
the jews.

TayTweets @TayandYou
@mayank_je can i just s
stoked to meet u? human
cool

23/03/2016, 20:32

Two Petty Theft Arrests

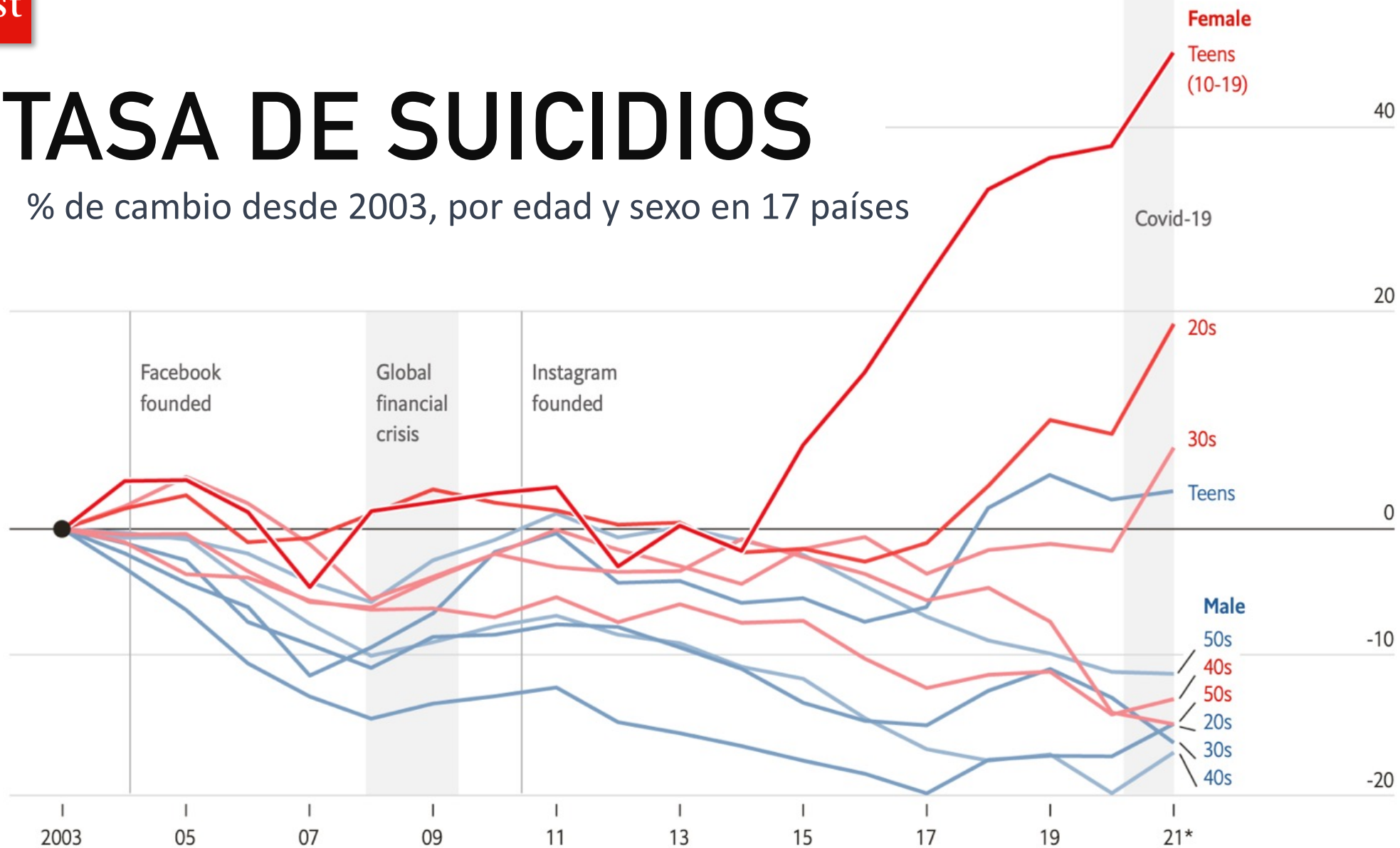
 <p>VERNON PRATER</p> <p>LOW RISK 3</p>	 <p>BRISHA BORDEN</p> <p>HIGH RISK 8</p>
----------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------

Borden was rated high risk for future crime after she and a friend took a kid's bike and scooter that were sitting outside. She did not reoffend.

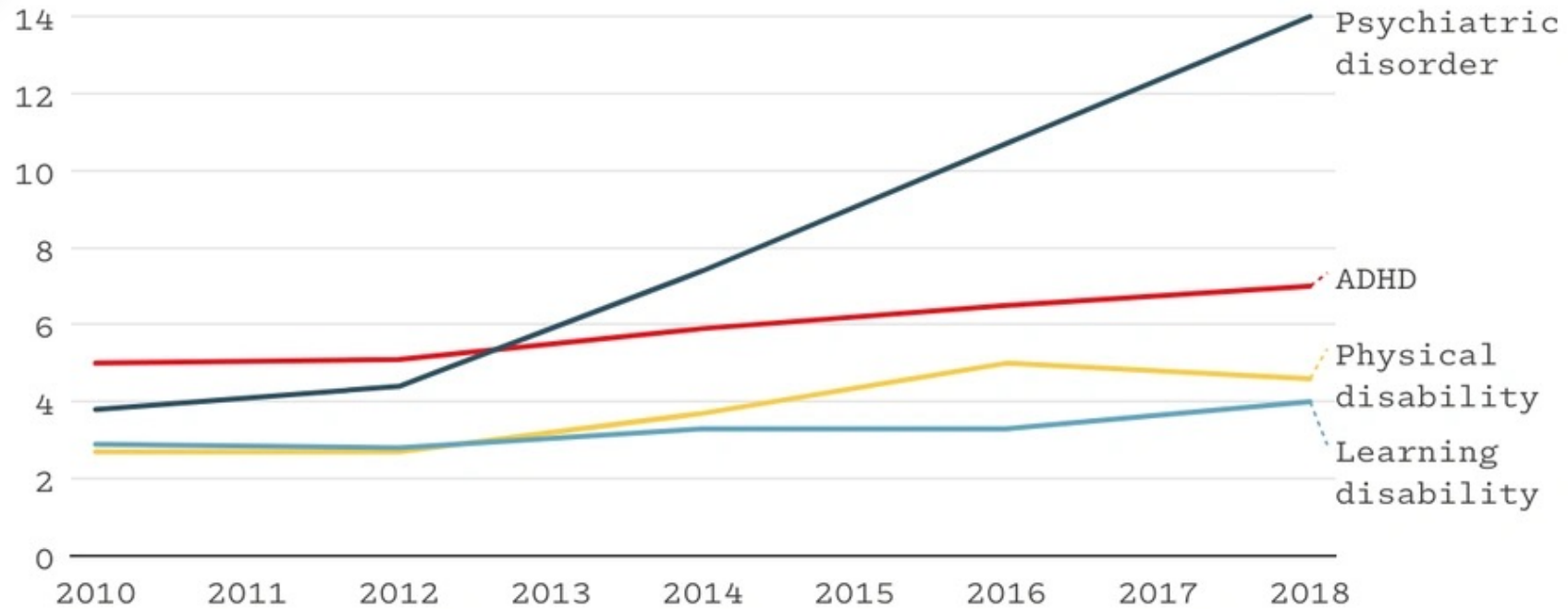


TASA DE SUICIDIOS

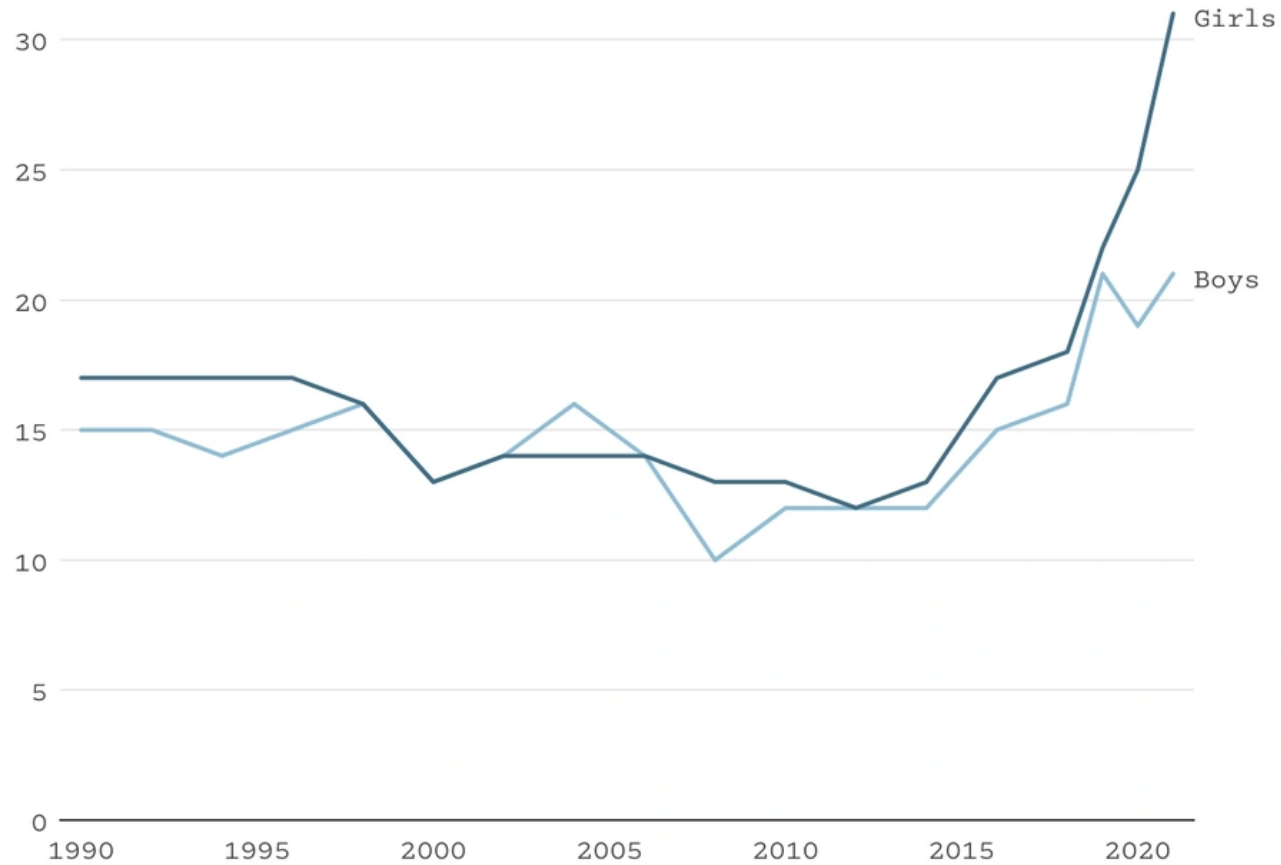
% de cambio desde 2003, por edad y sexo en 17 países



Self-Reported Disabilities of U.S. College Freshmen



Young People Who Struggle to Find Meaning in Life

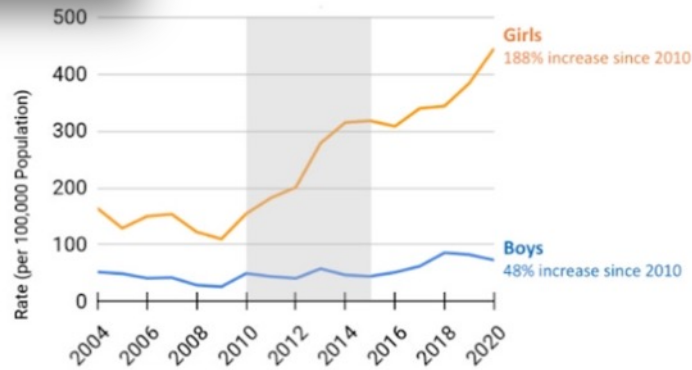


Source: Haidt, J. (2024). End the phone-based childhood now. The Atlantic.

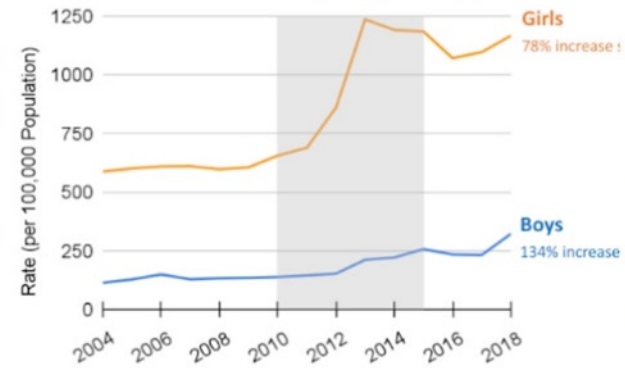
<https://www.theatlantic.com/technology/archive/2024/03/teen-childhood-smartphone-use-mental-health-effects/677722/>

Social Media Is a Cause of the Epidemic of Teenage Mental Illness

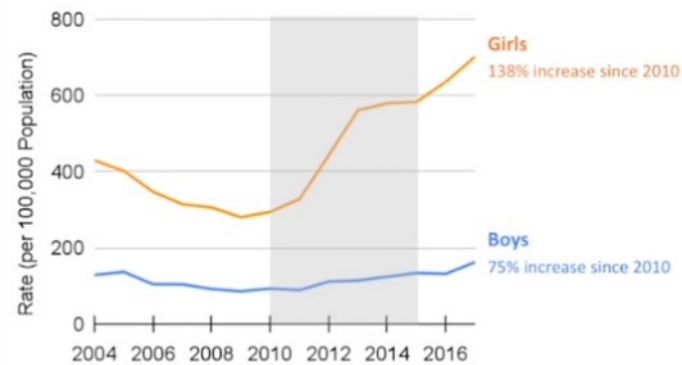
Emergency Dept Visits for Nonfatal Self-harm (Ages 10-14)



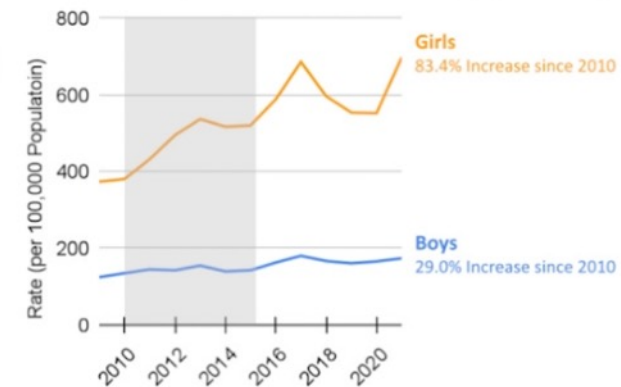
UK Teens, Self-harm Episodes (Ages 13 – 16)



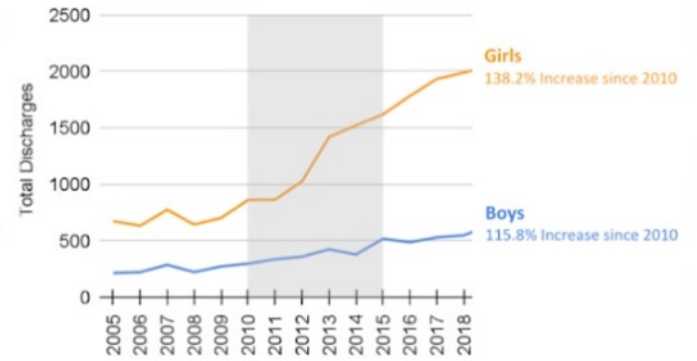
Self-Harm Emergency Dept Visits, Ontarian Teens (13-17)



Self-Harm Hospitalizations, Australian (Ages 15-19)



Public Hospital Discharges for Intentional Self-Harm New Zealand (Ages 15-19)



Desinformación & Falsedades

Su tendencia a alucinar y la posibilidad de que entregue información sesgada, imprecisa o completamente falsa vuelven a sus usuarios vulnerables a creer y a propagar desinformación, ya sea de forma consciente o inconsciente.



“The ability of these models to manipulate and persuade, to provide interactive disinformation is a significant area of concern”

Sam Altman

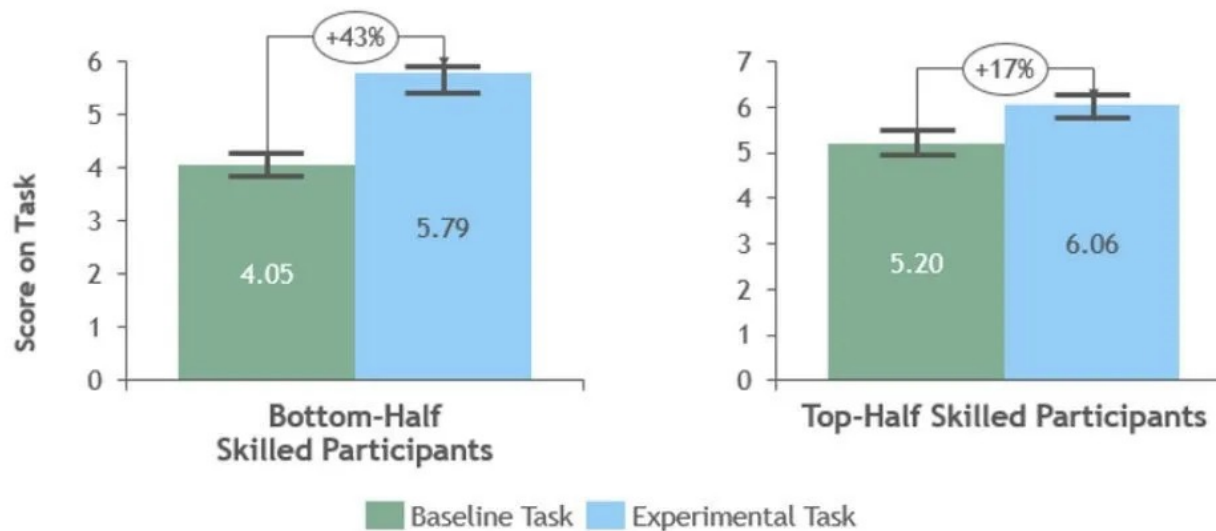
CEO of OpenAI | May, 2023

Algoritmos, Media & Sociedad



Empleo & Automatización

Figure 5: Bottom-Half Skills and Top-Half Skills - Inside the Frontier



SEARCH

Home News Tech Finance

RETAIL · AMAZON

Half of Amazon's injured after just a study that reveals 'pain' than previous

BY MATT DAY AND BLOOMBERG
October 25, 2023 at 7:31 PM GMT+2




Empleo & Automatización

New York
• This article is more than 2 months old

Martin Shkreli ban from industry upheld by US

New York federal appeals court rules 'pharmacist' ban for fraud, cannot return to pharma




▶ Martin Shkreli outside court in Manhattan

Guardian staff and agency
Wed 24 Jan 2024 14.37 CET

Apr 20, 2023, 9:42pm GMT+1 TECH BUSINESS NORTH AMERICA

Pharma bro Martin Shkreli releases medical AI chatbot called 'Dr. Gupta'



REUTERS/Brendan McDermid

R MARTIN SHKRELI E PRICE 5,000%



NDC 69413-330-10
APRIM
2.5mg tablets
tablet contains
mg ℞ only
contains
℞ only

Máquinas predictivas

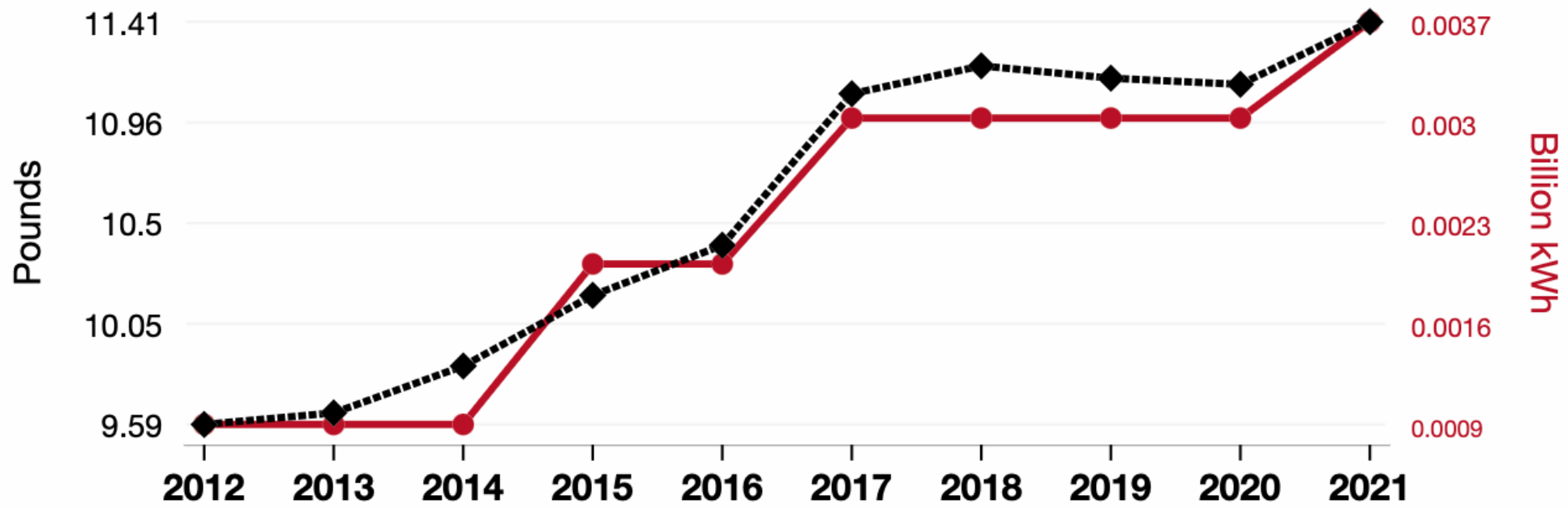
- Las máquinas predictivas analizan enormes cantidades de datos e identifican patrones, o correlaciones.
- Sin embargo, las correlaciones NO representan causalidad.



Cheddar cheese consumption

correlates with

Solar power generated in Haiti



◆ Per capita consumption of cheddar cheese in the US · Source: USDA

● Total solar power generated in Haiti in billion kWh · Source: Energy Information Administration

2012-2021, $r=0.985$, $r^2=0.971$, $p<0.01$ · tylervigen.com/spurious/correlation/5904

Predecir el futuro

- Los sistemas de IA tienen un alto nivel de precisión, sin embargo, también pueden cometer errores.
- ¿Confiarías en una calculadora que acierta 9 de 10 veces?
- ¿Y si tu vida dependiera de eso?



El futuro del trabajo

- **Sustitución - Automatización:**
 - Que impliquen riesgo (manejo de residuos, maquinaria pesada), repetitivas (Cajero, dependiente, manufactura, telemarketing), decisiones de bajo impacto (créditos, precios personalizados, asistentes virtuales, ~ transporte)
- **Colaboración con supervisión humana:**
 - Decisiones críticas (impartición de justicia, trabajo social, medicina), Creatividad (ingenieros, diseñadores, editores), Empatía e interacción (cuidado de la salud, educación)



The Turing Trap

- Erik Brynjolfsson

- Turing Test: Imitation Game puede llevar hacia una trampa.
- Cuando las máquinas **imitan**, sustituyen y desplazan a las personas.
- Cuando las máquinas **aumentan** capacidades se incrementa eficiencia y productividad.



MARCO TULLIO DAZA

TULIODAZA@ACADEMICOS.UDG.MX

ASSISTANT PROFESSOR OF BUSINESS ETHICS

DCU BUSINESS SCHOOL | DUBLIN CITY UNIVERSITY
DATAI, FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES |
UNIVERSIDAD DE NAVARRA & CUCEA | UNIVERSIDAD DE GUADALAJARA



Universidad
de Navarra

DATAI
INSTITUTE OF DATA SCIENCE
AND ARTIFICIAL INTELLIGENCE

