



Fei-Fei Li

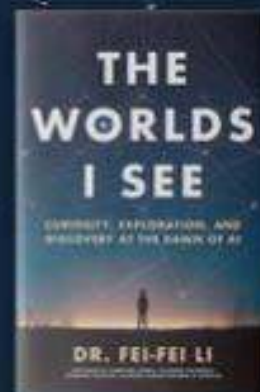
2,531 posts



THE WORLDS I SEE

CURIOSITY, EXPLORATION, AND
DISCOVERY AT THE DAWN OF AI

NOVEMBER 7TH



Following

Fei-Fei Li

@drfeifei

Prof (CS @Stanford), Co-Director @StanfordHAI, CoFounder/Chair @ai4allorg,
Researcher #AI #computervision #ML AI+healthcare

📍 Stanford, CA, U.S.A. 🔗 svl.stanford.edu 📅 Joined April 2010

1,071 Following 452.1K Followers



Followed by Salim Azak, Peter Stone, and 16 others you follow



Chelsea Finn

532 posts



Following

Chelsea Finn

@chelseabfinn

Asst Prof of CS & EE @Stanford.

PhD from @Berkeley_EECS, EECS BS from @MIT

📍 Palo Alto, CA 🔗 ai.stanford.edu/~cbfinn 📅 Joined June 2014

381 Following 68.2K Followers



Followed by RL_Conference, Peter Stone, and 15 others you follow



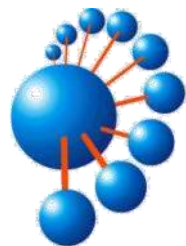
I imagine a future where robots are so integrated in the fabric of human life that they become as common as smartphones are today.





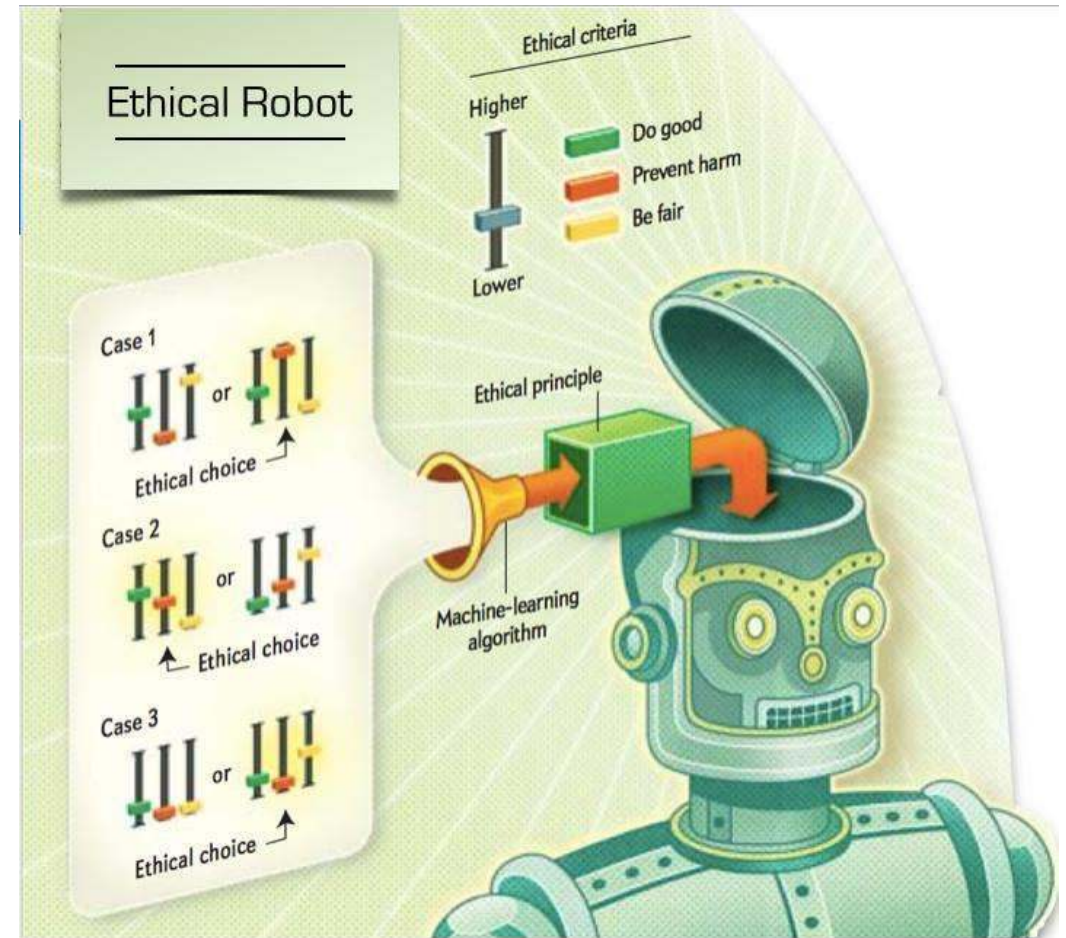
Ética en Inteligencia Artificial

Julio Godoy
DIICC



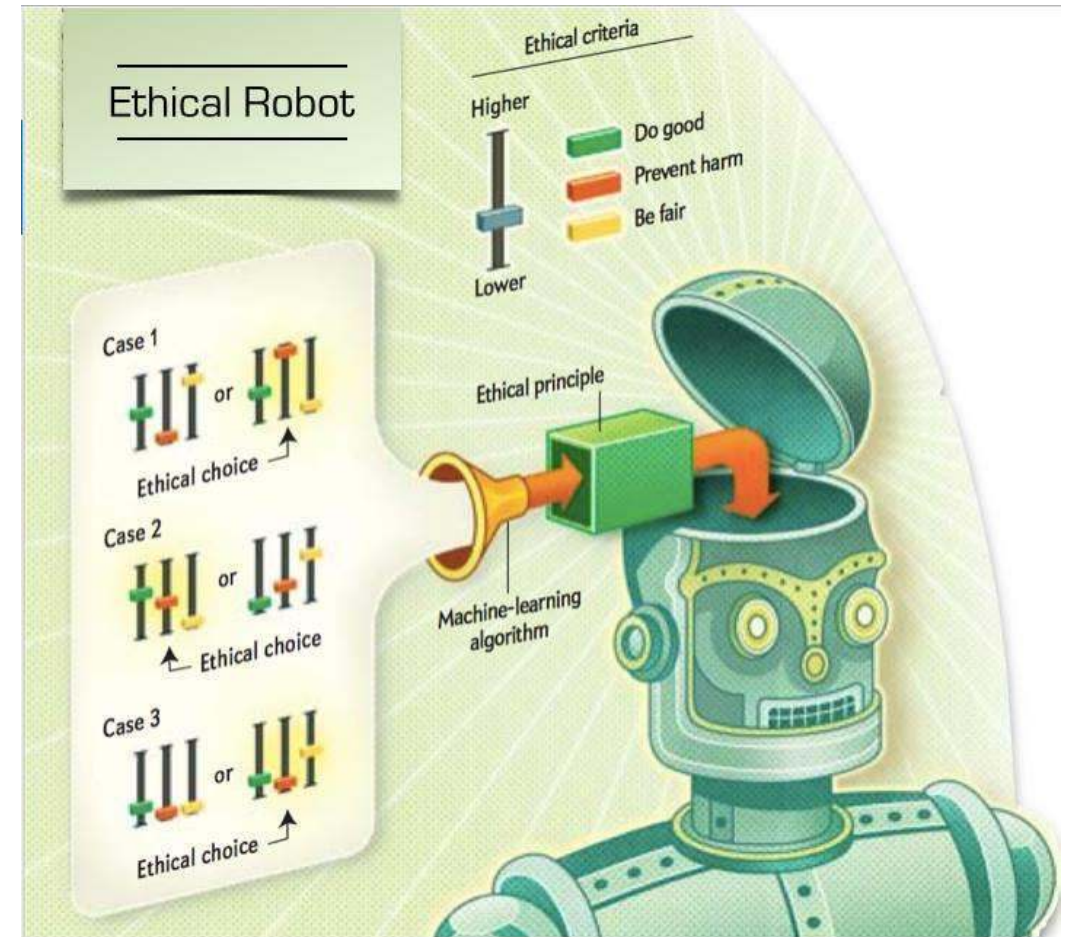
¿Por qué es importante hablar de esto?

- Desarrollo vertiginoso de métodos de inteligencia artificial
- Inteligencia artificial cada vez en más áreas, a cargo de muchos recursos
- Autonomía en toma de decisiones
- “Machine Ethics”



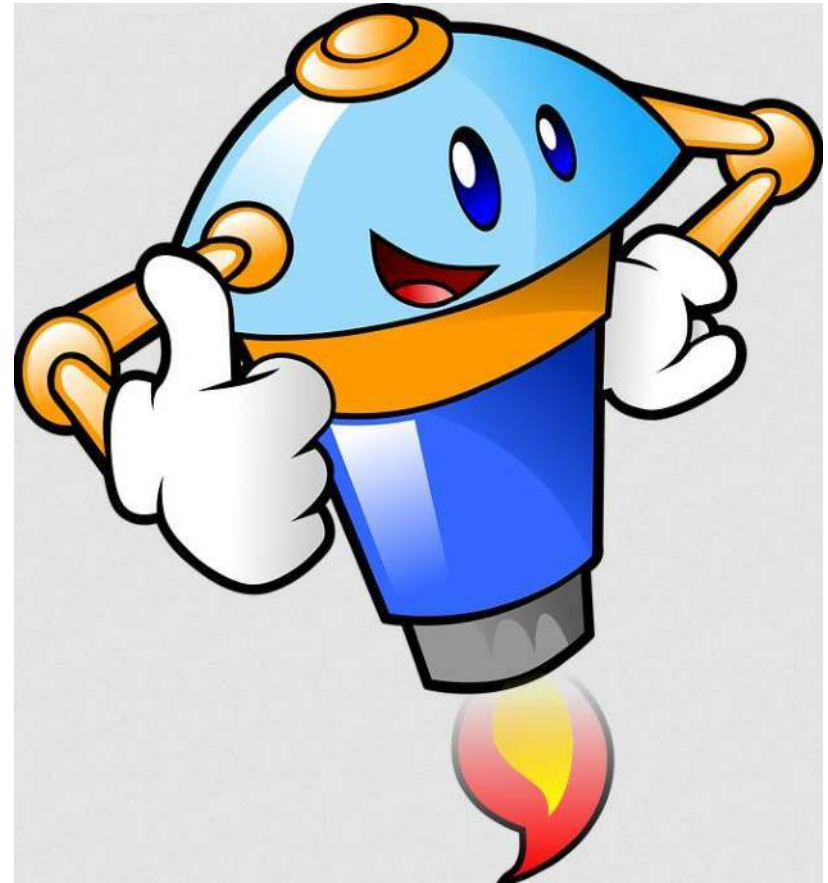
Ética de la Inteligencia Artificial

- Dada la capacidad de la IA, tenemos una obligación moral de usarla bien
 - promover sus aspectos positivos
 - evitar o mitigar los negativos



Ética de la Inteligencia Artificial

- Lo bueno:
 - IA puede salvar vidas
 - Diagnósticos médicos
 - Predicción de eventos climáticos extremos
 - Asistencia en la conducción y eventualmente vehículos autónomos seguros



Ética de la Inteligencia Artificial

- Lo bueno:
 - IA puede mejorar la calidad de vida
 - Microsoft “IA para acción humanitaria”
 - Google “IA para el bien social”, etc..
 - Monitoreo de contaminación, mediciones de emisiones de combustibles fósiles
 - Prevención de suicidio



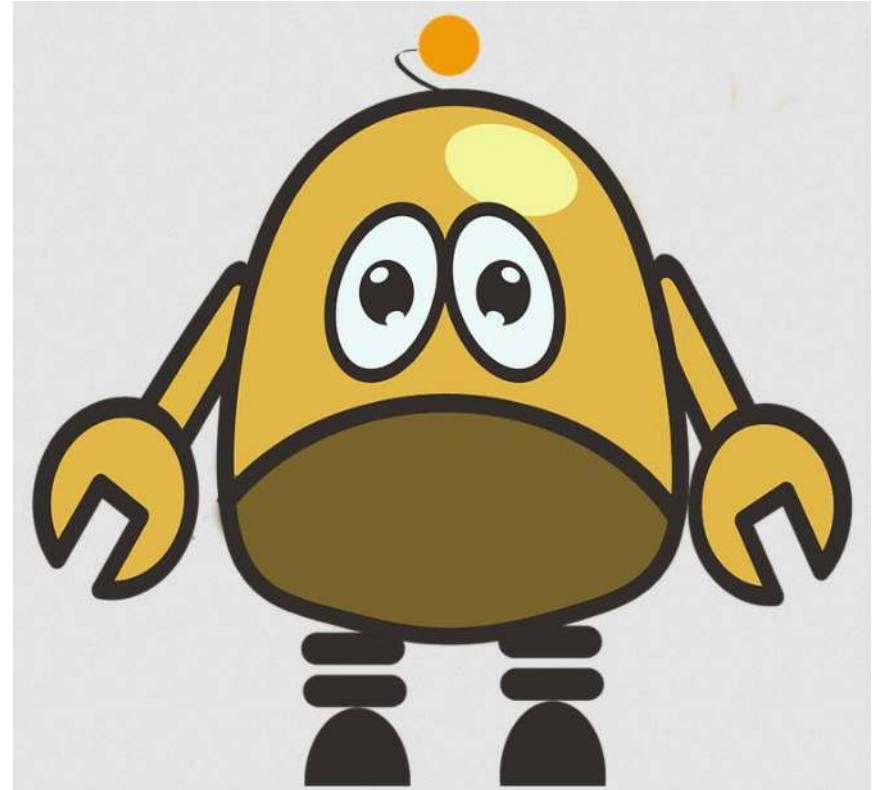
Ética de la Inteligencia Artificial

- Lo bueno:
 - IA puede mejorar la calidad de vida
 - Recuperación post-desastre
 - Manejo de cultivos y producción de comida ayudará a alimentar el mundo
 - Automatización puede reemplazar tareas tediosas y/o peligrosas
 - Personas con necesidades especiales se beneficiarán de la IA para poder ver, escuchar, y movilizarse.



Ética de la Inteligencia Artificial

- Lo malo:
 - Puede tener efectos secundarios negativos
- Desarrolladores de IA se enfrentan a consideraciones éticas:
 - ¿Qué proyectos deberían tomar?
 - ¿Cómo asegurarse de que su ejecución es segura y beneficiosa?



Ética de la Inteligencia Artificial

- Varias organizaciones han creado conjuntos de principios a los que su quehacer se adscribe:
 - Asegurar la seguridad
 - Asegurar la justicia
 - Respetar la privacidad
 - Establecer responsabilidades
 - Proveer transparencia
 - Limitar usos dañinos de la IA

Clase 2

AGENDA

UNIDAD	TEMA
01	Introducción: ¿Por qué hablar de ética en la IA?
02	Armas autónomas - Vigilancia y privacidad
03	Sesgos y justicia
04	El futuro del trabajo
05	Noticias recientes

Armas autónomas

- Localizan, seleccionan y ejecutan objetivos sin supervisión humana
 - “Tercera revolución en armamento de guerra”
- Discusiones en torno a la prohibición no han tenido apoyo unánime
 - Se oponen: EEUU, Israel, Rusia, Corea del Sur, entre otros



Harop

Armas autónomas

- Aspectos legales: diferenciar entre combatiente y no combatiente
- Aspectos éticos: ¿es moralmente aceptable delegar la decisión de matar humanos a una máquina?
 - “Campaña para Detener a los Robots Asesinos”
 - Pueden reducir la incidencia de bajas civiles
- Aspectos prácticos: son armas de destrucción masiva escalable



Harop

Armas autónomas

- Stanislav Petrov y los Minuteman



Stanislav Petrov

Armas autónomas

- Aspectos legales: diferenciar entre combatiente y no combatiente
- Aspectos éticos: ¿es moralmente aceptable delegar la decisión de matar humanos a una máquina?
 - “Campaña para Detener a los Robots Asesinos”
 - Pueden reducir la incidencia de bajas civiles
- Aspectos prácticos: son armas de destrucción masiva escalable



Harop

Vigilancia y privacidad

- Balancear el derecho a la privacidad del individuo con el valor que genera para la sociedad el poder compartir los datos.
 - Curar enfermedades manteniendo el historial médico privado
 - Prevenir ataques terroristas
- De-identificación!
 - Eliminar información que puede identificar a un individuo
 - Existen formas de re-identificación
 - Ej. Netflix data



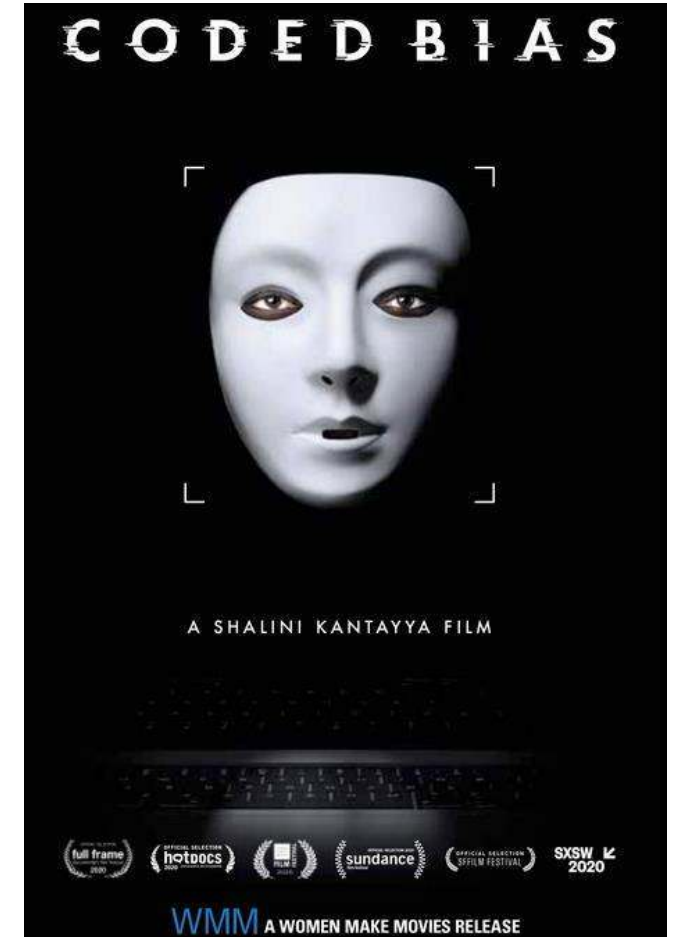
Vigilancia y privacidad

- Avance de tecnología de reconocimiento facial
- Gran parte de imágenes usadas para métodos de reconocimiento facial son sacadas de la web
 - ¿Consentimiento?



Vigilancia y privacidad

- Avance de tecnología de reconocimiento facial
- Gran parte de imágenes usadas para métodos de reconocimiento facial son sacadas de la web
 - ¿Consentimiento?



Clase 2

AGENDA

UNIDAD	TEMA
01	Introducción: ¿Por qué hablar de ética en la IA?
02	Armas autónomas - Vigilancia y privacidad
03	Sesgos y justicia
04	El futuro del trabajo
05	Noticias recientes

Sesgos y justicia



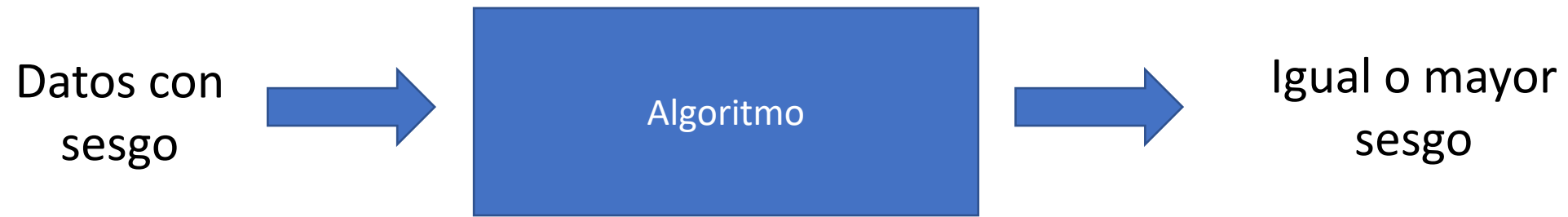
Algoritmo

Sesgos y justicia

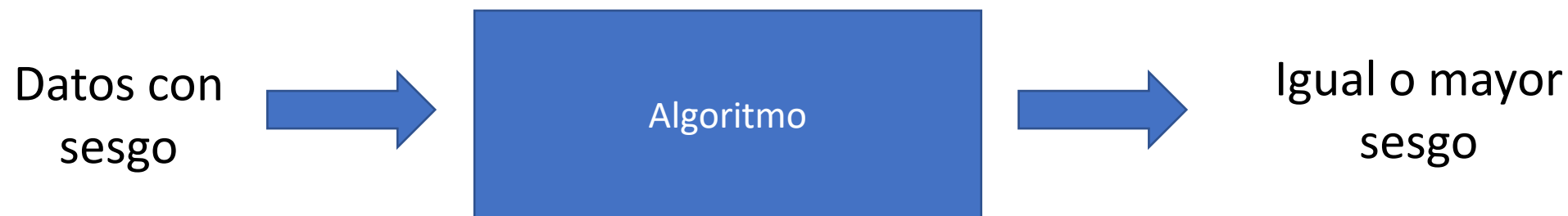
Datos con
sesgo



Sesgos y justicia



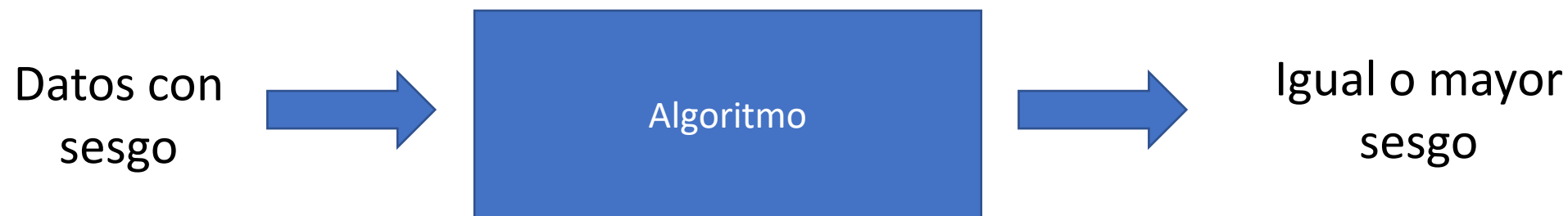
Sesgos y justicia



Neutral?
Justo?

Es importante en la medida
en que tiene consecuencias
negativas en las personas

Sesgos y justicia



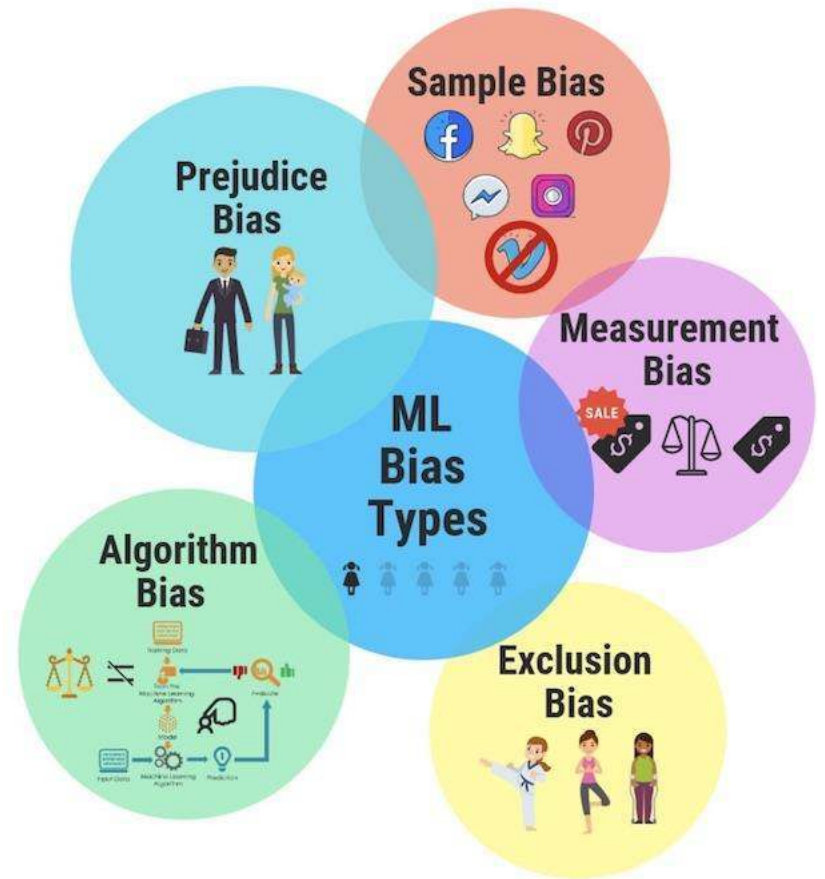
Neutral?
Justo?

Sistemas de IA no necesitan ser perfectos,
sólo mejores que nosotros

Es importante en la medida
en que tiene consecuencias
negativas en las personas

Sesgos y justicia

- Métodos de machine learning pueden perpetuar sesgos de la sociedad
- ¿Cómo evitarlo?
 - Comprender los límites de los datos que se usen
 - Des-sesgar los datos: sobre muestrear en clases minoritarias
 - Crear nuevos modelos y algoritmos de machine learning más resistentes a los sesgos



Sesgos y justicia

Bloomberg **Opinion**

you stay in the know. **Enable Notifications.**

Subscriber Only

Technology & Ideas

Amazon's Gender-Biased Algorithm Is Not Alone

They're everywhere, but nobody wants to know about it.

By Cathy O'Neil

16 de octubre de 2018 10:00 GMT-3



LIVE ON BLOOMBERG
[Watch Live TV >](#)
[Listen to Live Radio >](#)



Sesgos y justicia

Facial recognition systems show rampant racial bias, government study finds

By Brian Fung, CNN Business

Updated 6:37 PM EST, Thu December 19, 2019

(CNN Business) — Federal researchers have found widespread evidence of racial bias in nearly 200 facial recognition algorithms in an extensive government study, highlighting the technology's shortcomings and potential for misuse.

Racial minorities were far more likely than whites to be misidentified in the US government's testing, the study found, raising fresh concerns about the software's impartiality even as more government agencies at the city, state and federal level clamor to use it.

Sesgos y justicia

Facial recognition systems show rampant racial bias, government study finds

By Brian Fung, CNN Business

Updated 6:37 PM EST, Thu December 19, 2019

(CNN Business) — Federal researchers have found widespread evidence of racial bias in nearly 200 facial recognition algorithms in an extensive government study, highlighting the technology's shortcomings and potential for misuse.

Racial minorities were far more likely than whites to be misidentified in the US government's testing, the study found, raising fresh concerns about the software's impartiality even as more government agencies at the city, state and federal level clamor to use it.

Sesgos y justicia

BUSINESS
INSIDER

ONDERNEMEN

TECH

FINANCE

CARRIÈRE

AUTO

PREMIUM

ALLES

STRATEGIE

BETTER CAPITALISM

[HOME](#) » [INTERNATIONAAL](#) » OUTRAGE OVER POLICE BRUTALITY HAS FINALLY CONVINCED AMAZON, MICROSOFT, AND IB

Outrage over police brutality has finally convinced Amazon, Microsoft, and IBM to rule out selling facial recognition tech to law enforcement. Here's what's going on.

Business Insider US

🕒 13 jun 2020

Sesgos y justicia

MOTHERBOARD
TECH BY VICE

Faulty Facial Recognition Led to His Arrest— Now He's Suing

Michael Oliver is the second Black man found to be wrongfully arrested by Detroit police because of the technology—and his lawyers suspect there are many more.



By [Natalie O'Neill](#)

September 4, 2020, 9:39am



[Share](#)



[Tweet](#)



[Snap](#)

Sesgos y justicia



Sesgos y justicia

- Perfilamiento criminal: Gotham (Palantir)
 - Usado en ciudades de EEUU y Europa
- Un error tiene consecuencias desastrosas

on Sport Culture Lifestyle More ▾

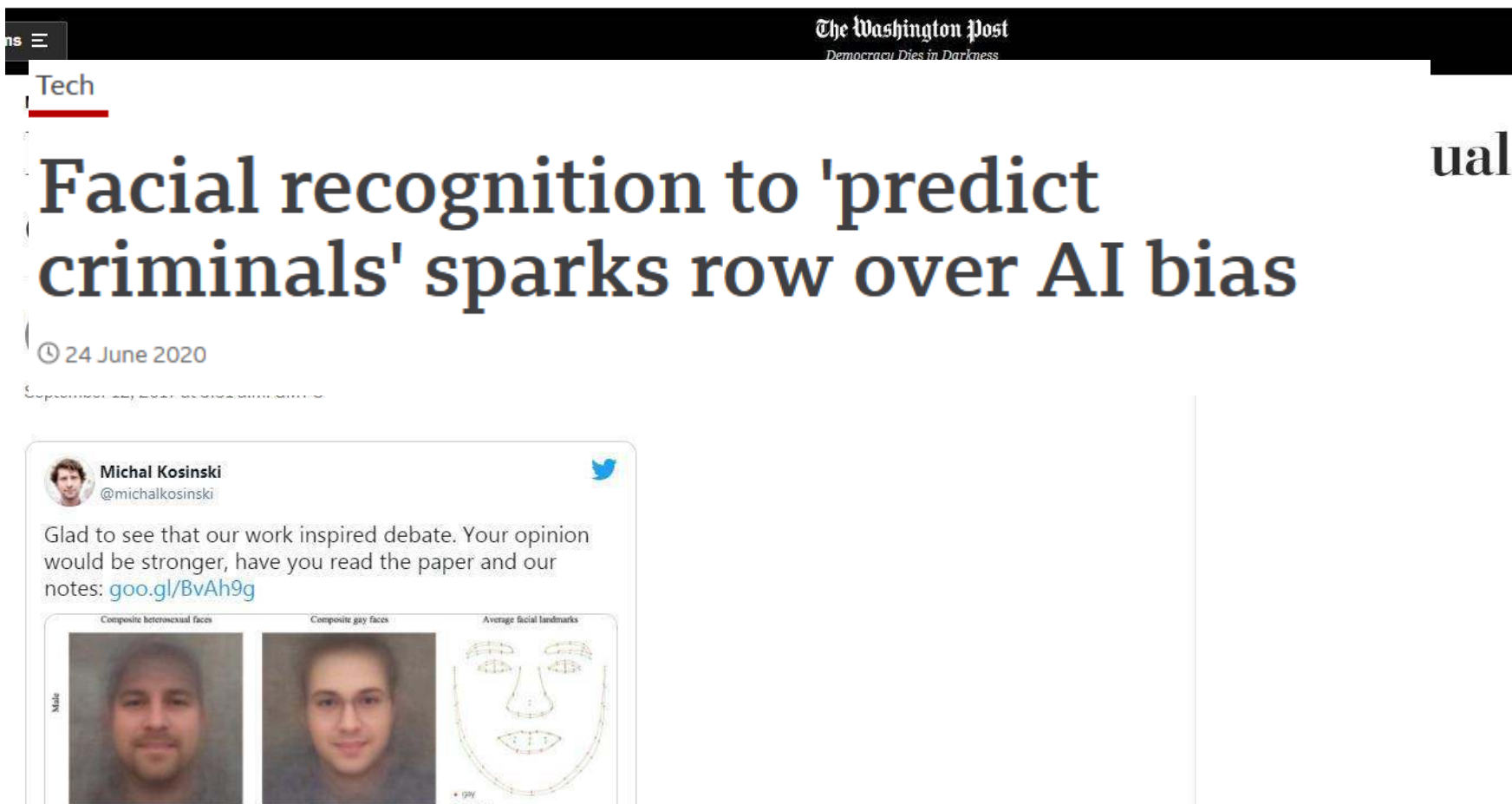
Asia Australia Middle East Africa Inequality Global development

Palantir: the 'special ops' tech giant that wields as much real-world power as Google

Peter Thiel's CIA-backed, data-mining firm honed its 'crime predicting' techniques against insurgents in Iraq. The same methods are now being sold to police departments. Will they inflame already tense relations between the public and the police?



Sesgos y justicia



Sesgos y justicia



scientific reports

Mor

R

[Explore content](#) ▾

[Journal information](#) ▾

[Publish with us](#) ▾

O



Sep

[nature](#) > [scientific reports](#) > [articles](#) > [article](#)

Article | [Open Access](#) | [Published: 11 January 2021](#)

Facial recognition technology can expose political orientation from naturalistic facial images

Michal Kosinski

Scientific Reports **11**, Article number: 100 (2021) | [Cite this article](#)

ias

Sesgos y justicia

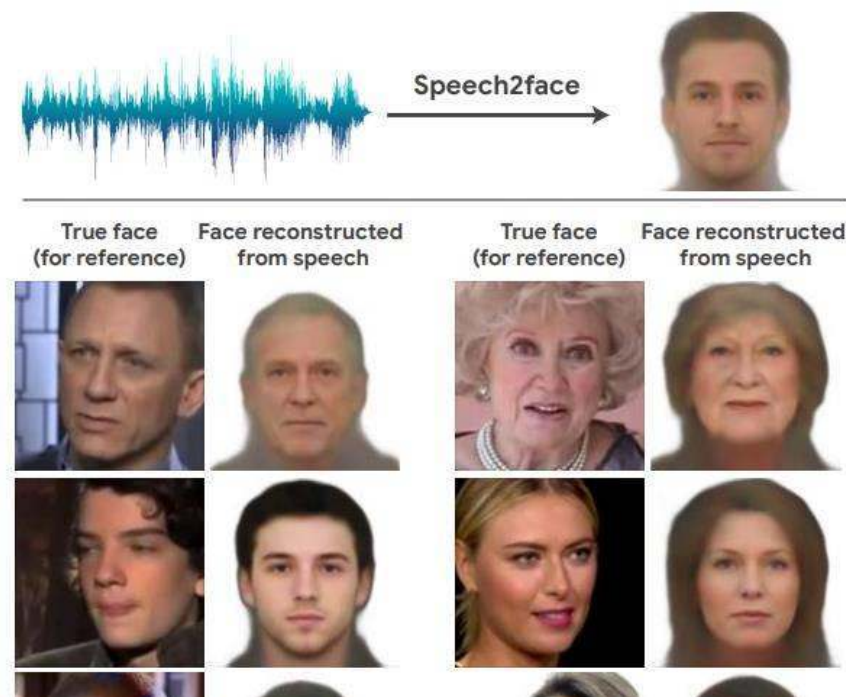
Speech2Face: Learning the Face Behind a Voice

Tae-Hyun Oh^{*†} Tali Dekel^{*} Changil Kim^{*†} Inbar Mosseri
William T. Freeman[†] Michael Rubinstein Wojciech Matusik[†]

[†]MIT CSAIL

Abstract

How much can we infer about a person's looks from the way they speak? In this paper, we study the task of reconstructing a facial image of a person from a short audio recording of that person speaking. We design and train a deep neural network to perform this task using millions of natural Internet/YouTube videos of people speaking. During training, our model learns voice-face correlations that allow it to produce images that capture various physical attributes of the speakers such as age, gender and ethnicity. This is done in a self-supervised manner, by utilizing the natural co-occurrence of faces and speech in Internet videos, without the need to model attributes explicitly. We evaluate and numerically quantify how—and in what manner—our Speech2Face reconstructions, obtained directly from audio, resemble the true face images of the speakers.



Sesgos y justicia

- Entender las limitaciones
 - No se puede aprender lo que no está en los datos
 - Más que la exactitud del modelo, hay que tener en cuenta el impacto de los errores
 - Examen COVID

Todos los modelos están equivocados
pero algunos son útiles

George E.P. Box



Ciencia y Tecnología

Médico estadounidense asegura que existen brechas para implementar la Inteligencia Artificial en salud

Publicado por [Francisco Iturra](#)

La información es de [Comunicado de Prensa](#)

Lunes 11 marzo de 2024 | 09:32

[Leer más tarde](#)



Ciencia y Tecnología

Médico estadounidense asegura que existen brechas para implementar la Inteligencia Artificial en salud

Publicado por [Francisco Iturra](#)

La hepatóloga también detalló cuáles son los puntos que se deben abordar para transformar la IA en una herramienta infaltable en el área de la salud de nuestro país. “Existe una gran brecha en cuanto al aspecto de los datos y su calidad, situación que no ocurre en otras áreas de diversas industrias en donde la IA es bastante bien aplicada”, señaló la doctora Strauss.





health

Life, But Better

Fitness

Food

Sleep

Mindfulness

Relationships

ChatGPT struggles to answer medical questions, new research finds

By Giri Viswanathan, CNN

🕒 5 minute read · Updated 8:49 AM EST, Sun December 10, 2023

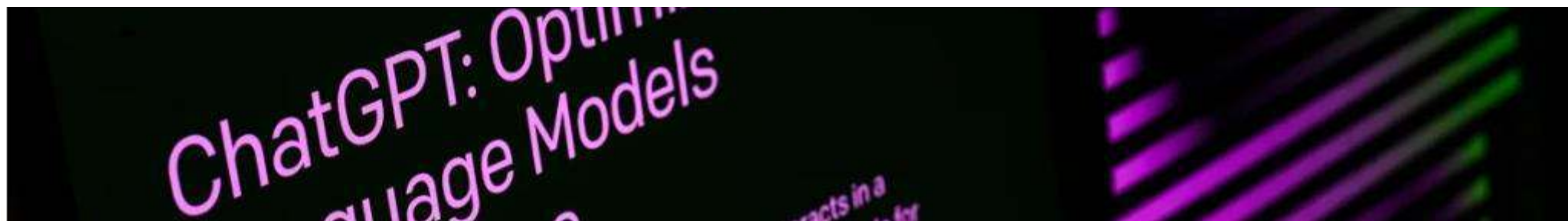


ChatGPT struggles to answer medical questions, new research finds

By Giri Viswanathan, CNN

🕒 5 minute read · Updated 8:49 AM EST, Sun December 10, 2023

The spokesperson pointed to ChatGPT's usage policies, which indicate that "OpenAI's models are not fine-tuned to provide medical information." The policy also states that the models should never be used to provide "diagnostic or treatment services for serious medical conditions."



Confianza y transparencia

- ¿Cómo lograrlo?
 - Uso de certificaciones
 - Métodos capaz de explicar sus decisiones:
“Explainable AI” (XAI)



Clase 2

AGENDA

UNIDAD	TEMA
01	Introducción: ¿Por qué hablar de ética en la IA?
02	Armas autónomas - Vigilancia y privacidad
03	Sesgos y justicia
04	El futuro del trabajo
05	Noticias recientes

El futuro del trabajo

- Desde la primera revolución de la agricultura (aprox. 10.000 AC), las nuevas tecnologías han cambiado la forma en que la humanidad vive y trabaja.

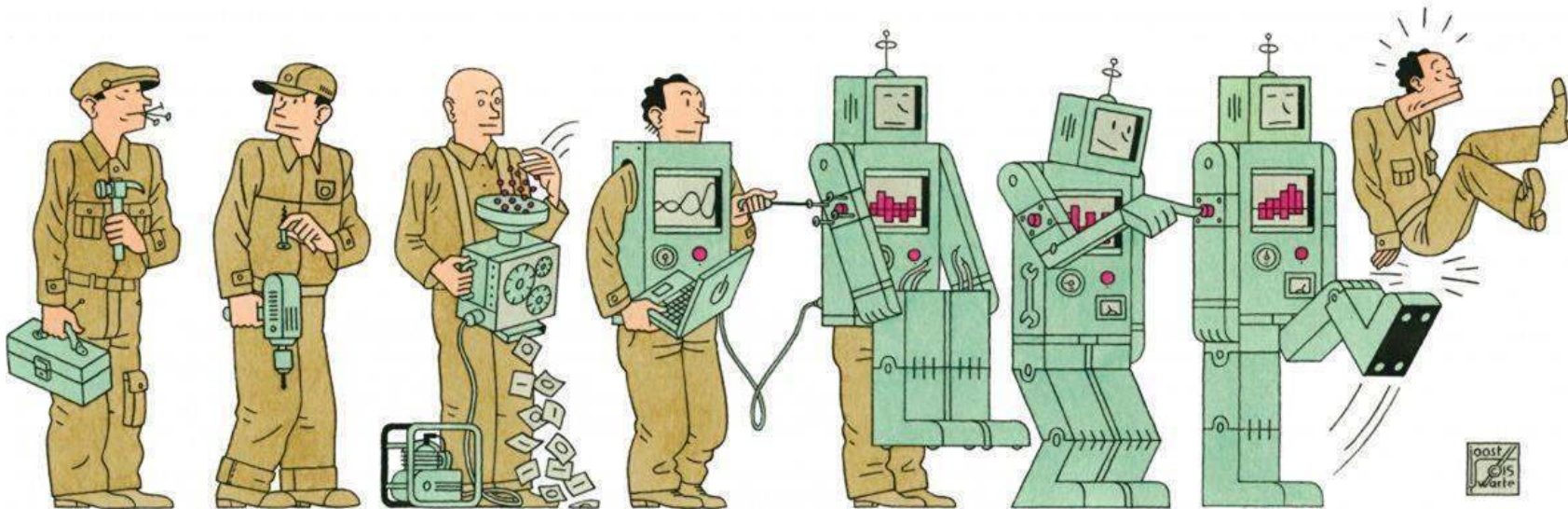


Image Source: MIT Technology Review

El futuro del trabajo

- Una de las principales preocupaciones del avance de la IA es que el trabajo humano quede obsoleto.
- <https://willrobotstakemyjob.com/>

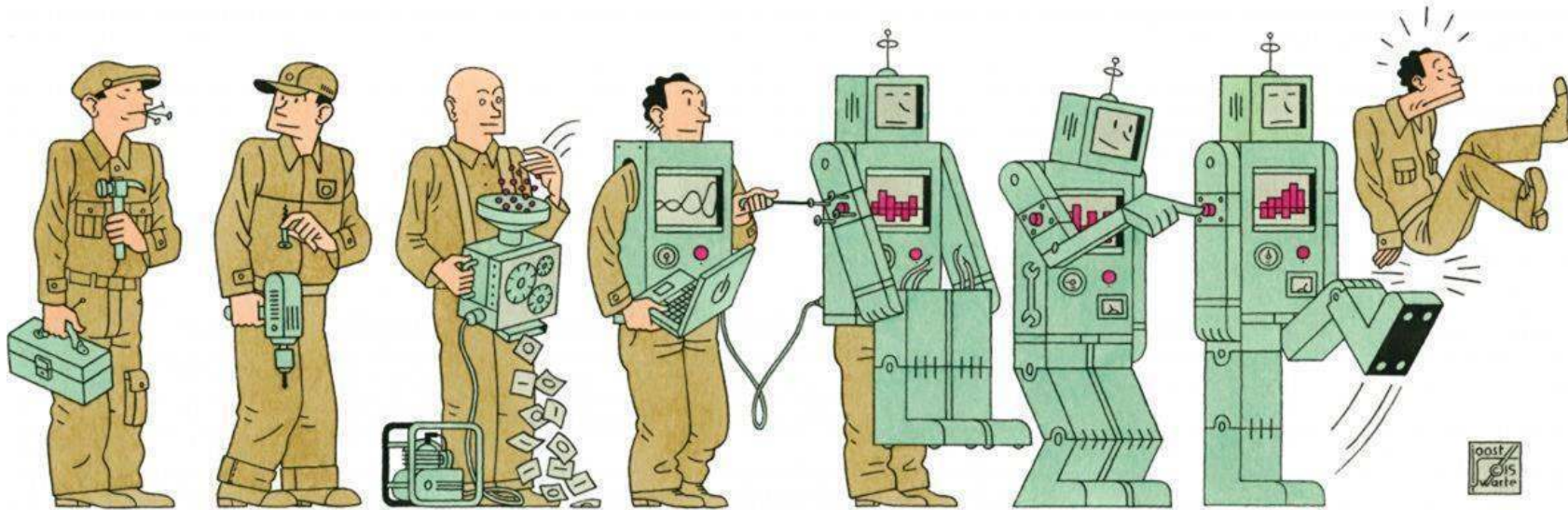


Image Source: MIT Technology Review

El futuro del trabajo

- Objetivo es “aumentar” a los humanos, no necesariamente reemplazarlos.
 - Ejemplo: IA en radiología

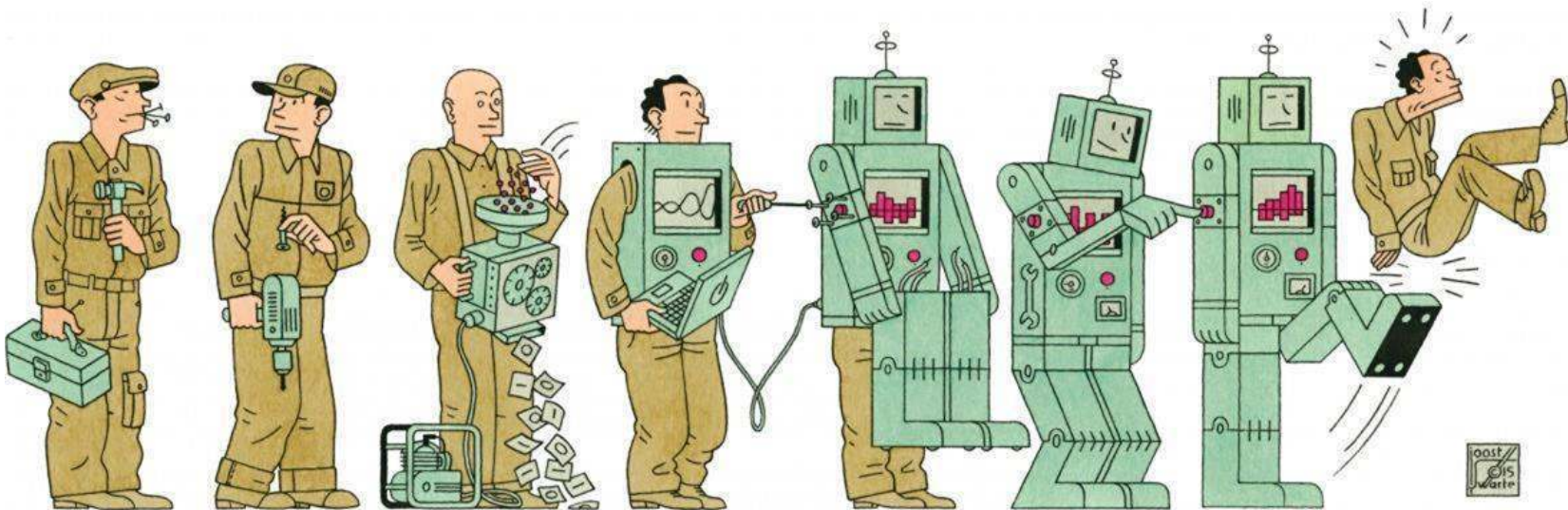


Image Source: MIT Technology Review

El futuro del trabajo

- Costos vs calidad
 - ¿Nos enfocamos en reducir costos? → pérdida de trabajos se ve como algo positivo
 - ¿Nos enfocamos en mejorar la calidad? → mejorando la calidad de vida para el trabajador y el cliente



El futuro del trabajo

- El ritmo del cambio es importante
 - Antes, los cambios eran más lentos
 - Ahora, el cambio puede ser más brusco, implicando que mucha gente deba re-entrenarse para nuevos trabajos.
 - Para esto, la ayuda de tener una educación “por toda la vida” es valiosa



El futuro del trabajo

- Trabajo cumple varios propósitos:
 - Producción de bienes y servicios
 - Medios para la supervivencia
 - Sentido de propósito e integración social
- En el futuro, estos tres propósitos pueden desagregarse:
 - La automatización proveerá, en parte, las necesidades de la sociedad
 - Personas tendrán que buscar sentido y propósito en otras actividades



Clase 2

AGENDA

UNIDAD	TEMA
01	Introducción: ¿Por qué hablar de ética en la IA?
02	Armas autónomas - Vigilancia y privacidad
03	Sesgos y justicia
04	El futuro del trabajo
05	Responsabilidad
06	Noticias recientes

Responsabilidad



Responsabilidad

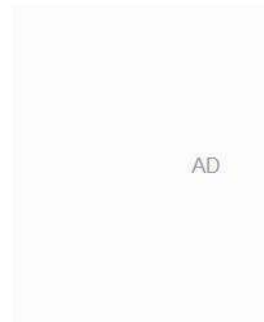
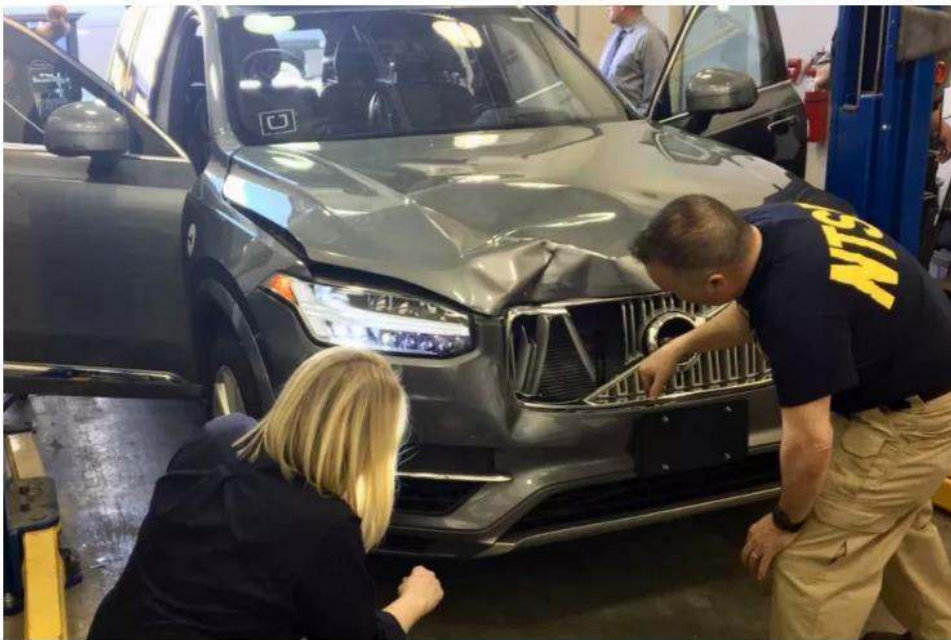
REPORT \ TRANSPORTATION \ UBER \

The world's first robot car death was the result of human error — and it can happen again

The damage from the Uber crash will have far-reaching consequences

By Andrew J. Hawkins | @andyjayhawk | Nov 20, 2019, 2:23pm EST

f   SHARE



VERGE DEALS



Imagen: theverge.com

Responsabilidad

REPORT \ TRANSPORTATION \ UBER \

The world's first robot car death was the result of human error — and it can happen again

The damage from the Uber crash will have far-reaching consequences

By Andrew J. Hawkins | @andyjayhawk | Nov 20, 2019, 2:23pm EST

f t SHARE



“In my opinion they’ve put technology advancement here before saving lives,” NTSB board member Jennifer Homendy said of the National Highway Traffic Safety Administration (NHTSA), which is responsible for regulating vehicle safety standards.

AD

VERGE DEALS



Imagen: theverge.com

Clase 2

AGENDA

UNIDAD	TEMA
01	Introducción: ¿Por qué hablar de ética en la IA?
02	Armas autónomas - Vigilancia y privacidad
03	Sesgos y justicia
04	El futuro del trabajo
05	Responsabilidad
06	Noticias recientes

CBP wants to use AI to scan for fentanyl at the border



/ Customs and Border Protection are installing “high-tech drug detection machines” that are actually just very, very strong X-rays.

By [Gaby Del Valle](#), a policy reporter. Her past work has focused on immigration politics, border surveillance technologies, and the rise of the New Right.

Mar 14, 2024, 2:52 PM GMT-3

[Link](#) [Facebook](#) [Twitter](#) | 15 Comments (15 New)

In his State of the Union address last week, President Joe Biden urged Congress to pass a bipartisan immigration bill “with the toughest set of border security reforms we’ve ever seen.” The bill would have, among other things, funded “100 more high-tech drug detection machines to significantly increase the ability to screen and stop vehicles from smuggling fentanyl into America.”



Ciencia y Tecnología

OpenAI pondrá GPT en los cerebros de robots con forma humana

Por Matías Vega

Viernes 01 marzo de 2024 | 09:45

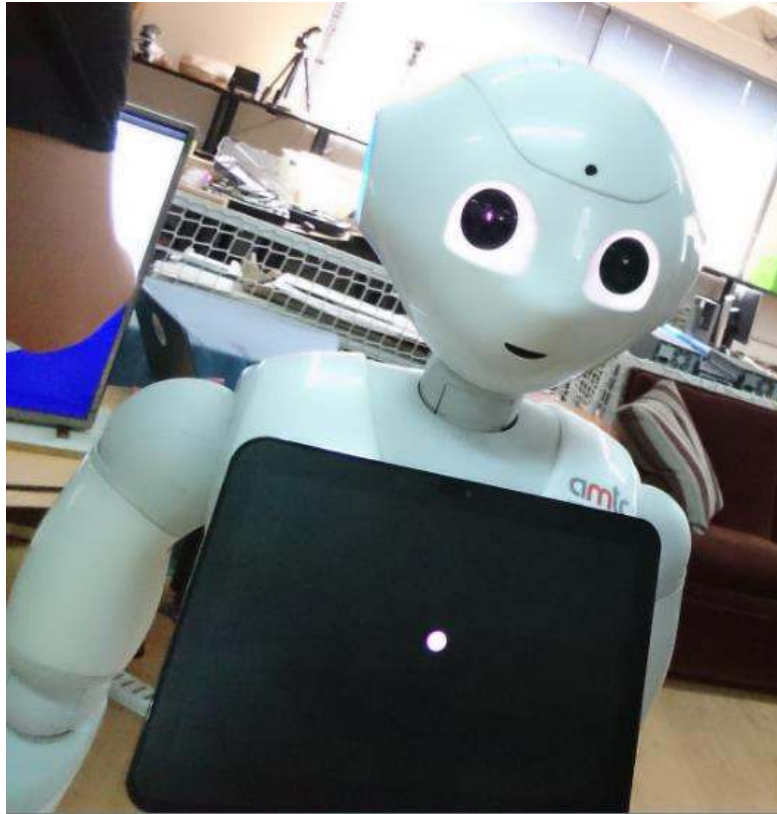
Leer más tarde



¿Podemos asegurar el comportamiento éticamente “correcto” en la IA?



Foto: Natalia Soto



La Unión Europea ratificó la primera ley de Inteligencia Artificial (IA) del mundo: ¿en qué consiste?

Publicado por Sara Jerez

La información es de Deutsche Welle

Jueves 14 marzo de 2024 | 08:30

Leer más tarde

En alianza con Deutsche Welle



Relacionados

- Médico estadounidense asegura que existe para implementar Inteligencia Artificial en salud



¿Podemos asegurar el comportamiento éticamente “correcto” en la IA?

- En resumen:
 - *“with great power comes great responsibility”*
- Desarrollo de sistemas avanzados de inteligencia artificial debe ser robusto a posibles usos maliciosos del mismo
- Eficiencia y uso ético de sistemas de IA pueden entrar en conflicto

