



JORNADA DE LOS COMITÉS DE ÉTICA DE LA INVESTIGACIÓN DEL SSPA - 2022



**Estrategia
de Bioética**
del Sistema Sanitario
Público de Andalucía

PRINCIPIOS ÉTICOS PARA LA IA: SENTIDO Y RETOS

Francisco Lara



ugr

Universidad
de Granada



GetTEC

Grupo de Investigación
en Ética Tecnológica

TECNOLOGÍA DISRUPTIVA EN LA REVOLUCIÓN DE LA INTELIGENCIA

Inteligencia artificial

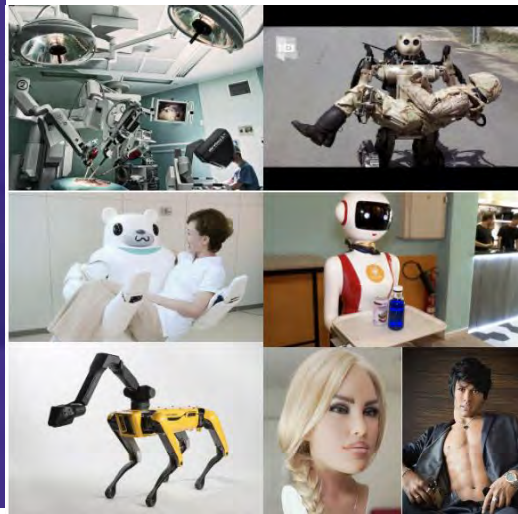
Uso cotidiano y potencial

Ejemplos sobre las aplicaciones actuales de la IA y posibilidades que ofrece

- Asistentes personales digitales en smartphone o PC
- Búsqueda en la web
- Traducción automática
- Ciberseguridad
- Lucha contra la desinformación
- Optimización de productos y procesos de ventas
- Aire acondicionado inteligente
- Vehículos autónomos
- Compras en línea y publicidad
- Agricultura inteligente: riego, alimentación de animales, robots anti malas hierbas
- Internet de las cosas: aspiradoras, frigoríficos, relojes, etc., conectados a internet
- Robots utilizados en fábricas



europarl.eu



QUE ESTÁ CAMBIANDO DRÁSTICAMENTE NUESTRA SOCIEDAD

- Con infinitos beneficios
- Pero también con riesgos y amenazas a nuestro valores



NECESITAMOS PRINCIPIOS ÉTICOS

- Cuáles se han propuesto y por qué
- Dificultades en su implementación

¿QUÉ PRINCIPIOS? DECLARACIONES Y CÓDIGOS

INSTANCIAS ACADÉMICAS

- Declaración de IA Responsable de Montreal (2017)
- AI4People (Floridi et al., 2018)
- Fairness, Accountability, Sustainability and Transparency, FAST (The Alan Turing Institute, 2018)

ORGANIZACIONES PROFESIONALES

- Institute of Electrical and Electronics Engineers (IEEE, 2019)

ASOCIACIONES Y ONGs

- Principios de Asilomar (Future Life Institute, 2017)
- Campaña contra los Robots Asesinos (2013-)

EMPRESAS

- Partnership on IA
- Códigos éticos de Google, Microsoft o Accenture

GOBIERNOS

- Reino Unido, Francia, EEUU... China

INSTITUCIONES INTERNACIONALES

- Grupo de Expertos de Alto Nivel (Comisión Europea, 2019)



EL CASO EUROPEO: IA FIABLE

THE AI ACT

EU Artificial Intelligence Act: Risk levels





THE CHINESE GOVERNMENT HAS LAUNCHED A SOCIAL CREDIT SYSTEM THAT WILL MONITOR AND REGULATE THE FINANCIAL, SOCIAL, MORAL AND, POSSIBLY, POLITICAL BEHAVIOUR OF CHINA'S CITIZENS - AND ALSO THE COUNTRY'S COMPANIES - VIA A SYSTEM OF PENALTIES AND REWARDS. THE STATED AIM IS TO "PROVIDE THE TRUSTWORTHY WITH BENEFITS AND DISCIPLINE THE UNTRUSTWORTHY."

BertelsmannStiftung

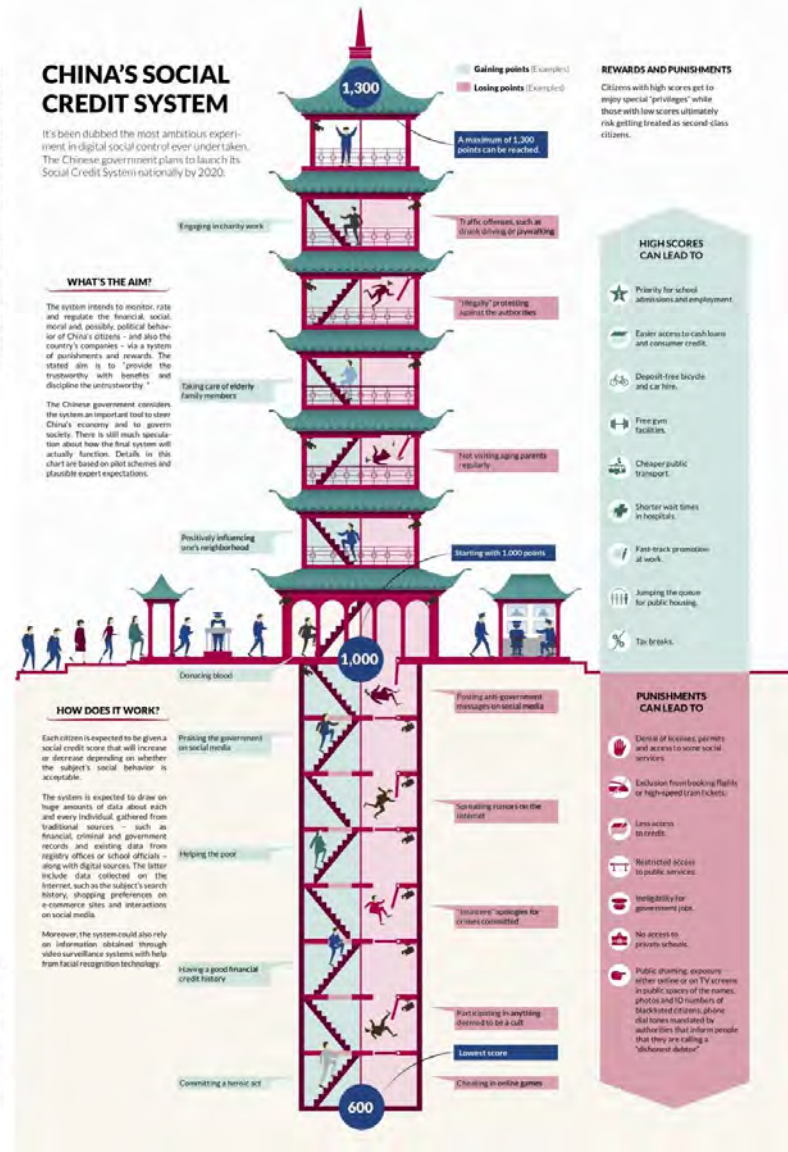
CHINA'S SOCIAL CREDIT SYSTEM

It's been dubbed the most ambitious experiment in digital social control ever undertaken. The Chinese government plans to launch its Social Credit System nationally by 2020.

WHAT'S THE AIM?

The system intends to monitor, rate and regulate the financial, social, moral and, possibly, political behaviour of China's citizens - and also the country's companies - via a system of penalties and rewards. The stated aim is to "provide the trustworthy with benefits and discipline the untrustworthy."

The Chinese government considers the system an important tool to steer China's economy and to govern society. There is still much speculation about how the final system will actually function. Details in this chart are based on pilot schemes and plausible expert expectations.

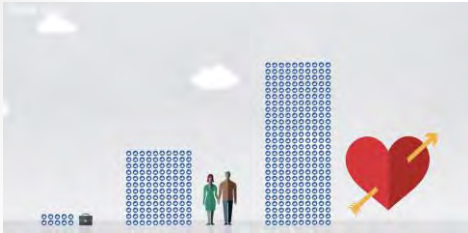


COERCIÓN POLÍTICA (CONTRA LA AUTONOMÍA HUMANA)

OTRA AMENAZA A LA AUTONOMÍA HUMANA: LA MANIPULACIÓN



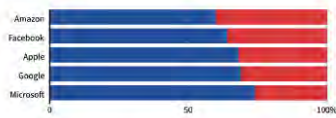
REDES SOCIALES Y EL PODER PSICOMÉTRICO DEL ALGORITMO



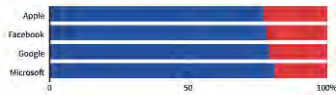


GLOBAL HEADCOUNT

■ Male ■ Female

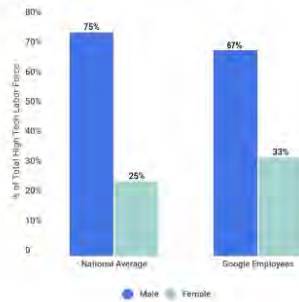


EMPLOYEES IN TECHNICAL ROLES



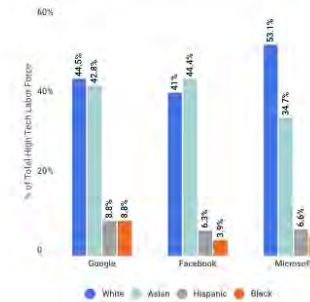
Note: Amazon does not disclose the gender breakdown of its technical workforce.
Source: Census data available from the companies, since 2017.
by Harri Haug | REDUXIS SPAIN

HIGH TECH LABOR FORCE BY GENDER



2021

HIGH TECH DEMOGRAPHICS BY COMPANY



CONTRA EL PRINCIPIO DE EQUIDAD ¿Y QUÉ SE PUEDE HACER?


HETEROGENEIDAD GRUPAL



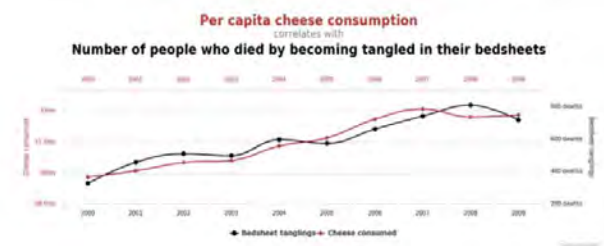
ÉTICA PARA LOS INGENIEROS

COMPAS

(Correctional Offender Management Profiling for Alternative Sanctions)

VERNON PRATER Prior Offenses 2 armed robberies, 1 attempted armed robbery Subsequent Offenses 1 grand theft LOW RISK 3	BRISHA BORDEN Prior Offenses 4 juvenile misdemeanors Subsequent Offenses None HIGH RISK 8	
---	--	---

DATOS SESGADOS + CORRELACIONES ESPÚREAS + MACHINE LEARNING

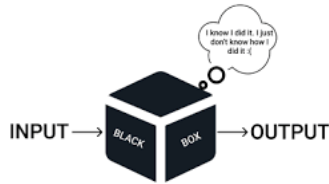


¿ALGORITMOS CIEGOS?



¿Y los rasgos indirectos?
¿Y si la igualdad fuera otra?

¿POR QUÉ UN PRINCIPIO DE EXPLICABILIDAD?



APRENDIZAJE PROFUNDO Y FALTA DE TRANSPARENCIA
(EL PROBLEMA DE LA **CAJA NEGRA**)

DECISIONES
INCOMPENSIBLES EN:
FINANZAS, DERECHO,
MEDICINA, DEFENSA...

AMENAZA
AL DERECHO A SABER
Y A **LA CONFIABILIDAD** EN
LAS MÁQUINAS

IA
EXPLICABLE
(XAI)

CON *CAJA BLANCA O
DE CRISTAL*

- Para sistemas de DL menos potentes
- Sesgo de automatización en visualizaciones complejas

CON
CONTRAFÁCTICOS

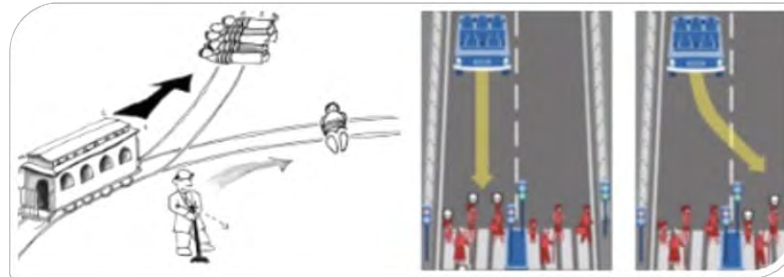
- Explicando lo que se necesitaría para dar una respuesta distinta

CON LENGUAJE
NATURAL

- Técnicas para agentes virtuales conversacionales
- Adaptables al perfil del usuario

Y POR ÚLTIMO, PREVENCIÓN DEL DAÑO PERO, ¿DE QUIÉN?

El dilema del tranvía y los vehículos sin conductor



Dilema entre éticos:
¿CONSECUENCIALISMO O
DEONTOLOGISMO?



QUE SOLO SE RESUELVE PARCIALMENTE CON LAS INTUICIONES

Sondeo *Moral Machine* (MIT):

PREFERENCIA UNIVERSAL: salvar muchos a uno (¿siendo uno el conductor?)

CULTURALES: jóvenes a ancianos; ejecutivos a los sin techo...

