



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



PRINCIPIOS ÉTICOS PARA LA IA: SENTIDO Y RETOS

Francisco Lara



ugr

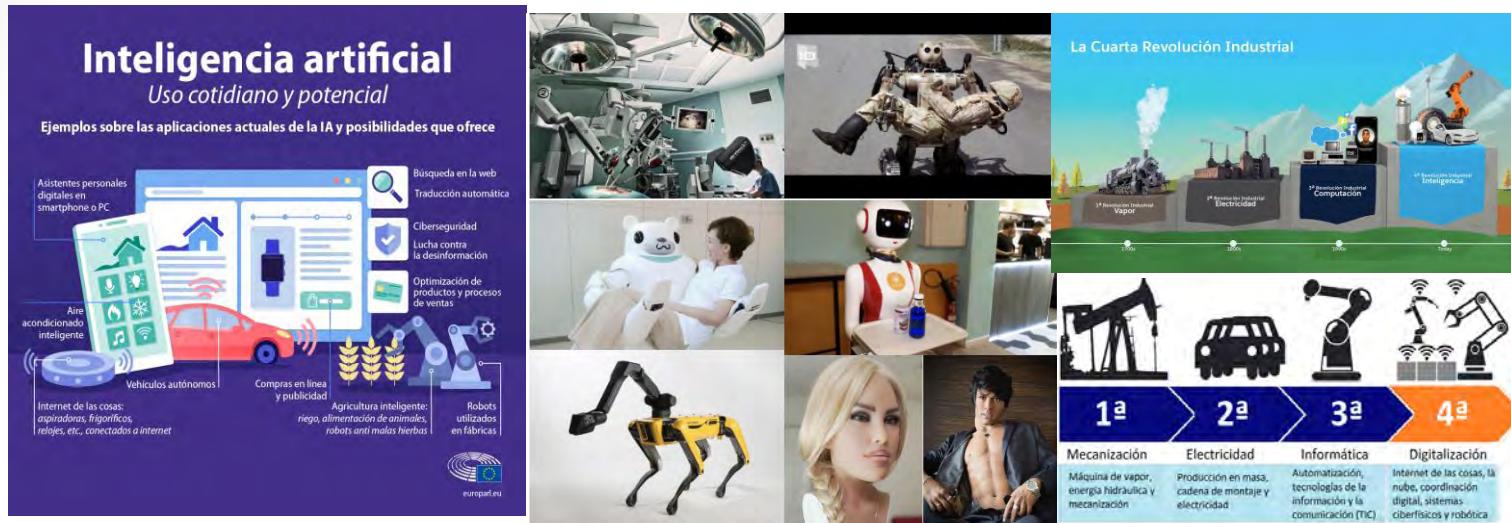
Universidad
de Granada



GetTEC

Grupo de Investigación
en Ética Tecnológica

TECNOLOGÍA DISRUPTIVA EN LA REVOLUCIÓN DE LA INTELIGENCIA



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

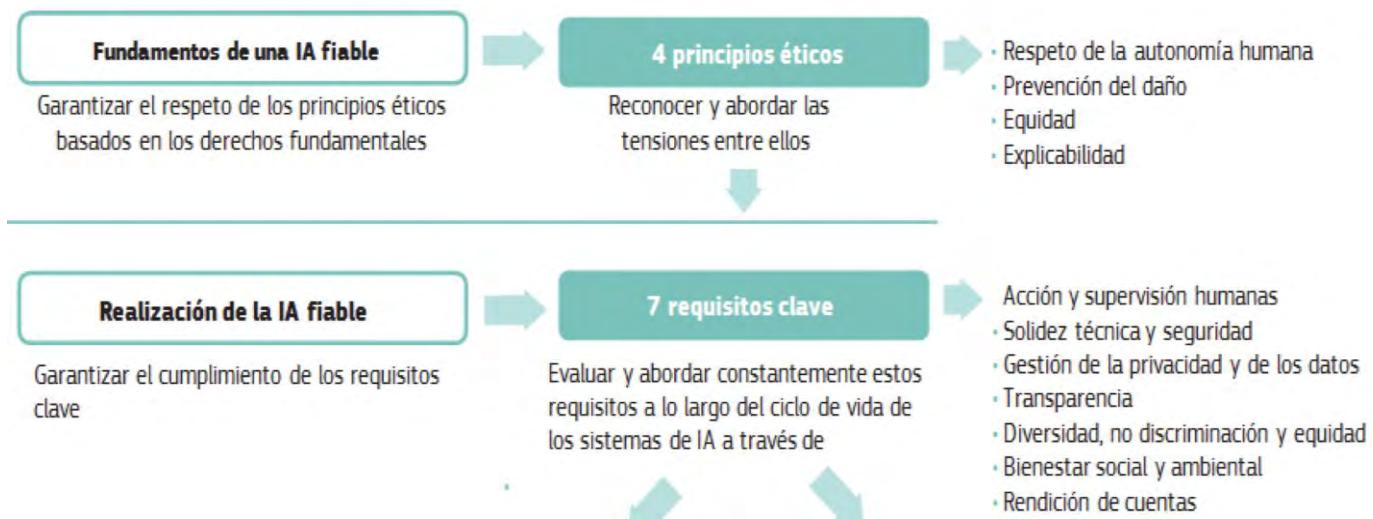
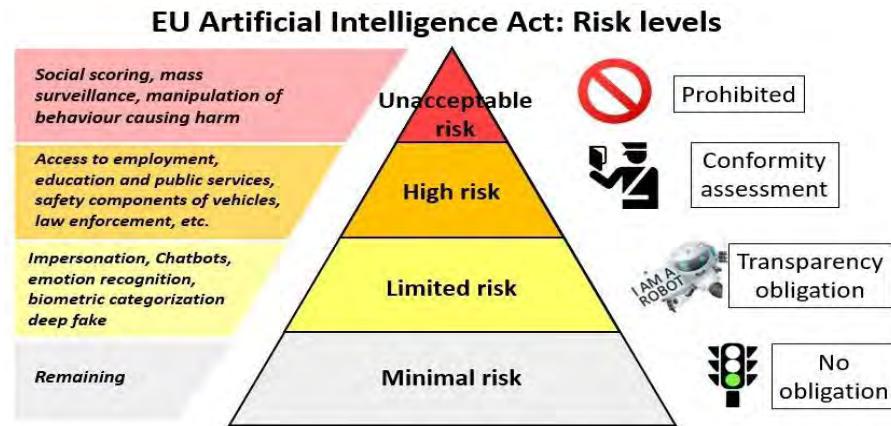
INSTITUCIONES INTERNACIONALES

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



EL CASO EUROPEO: IA FIABLE

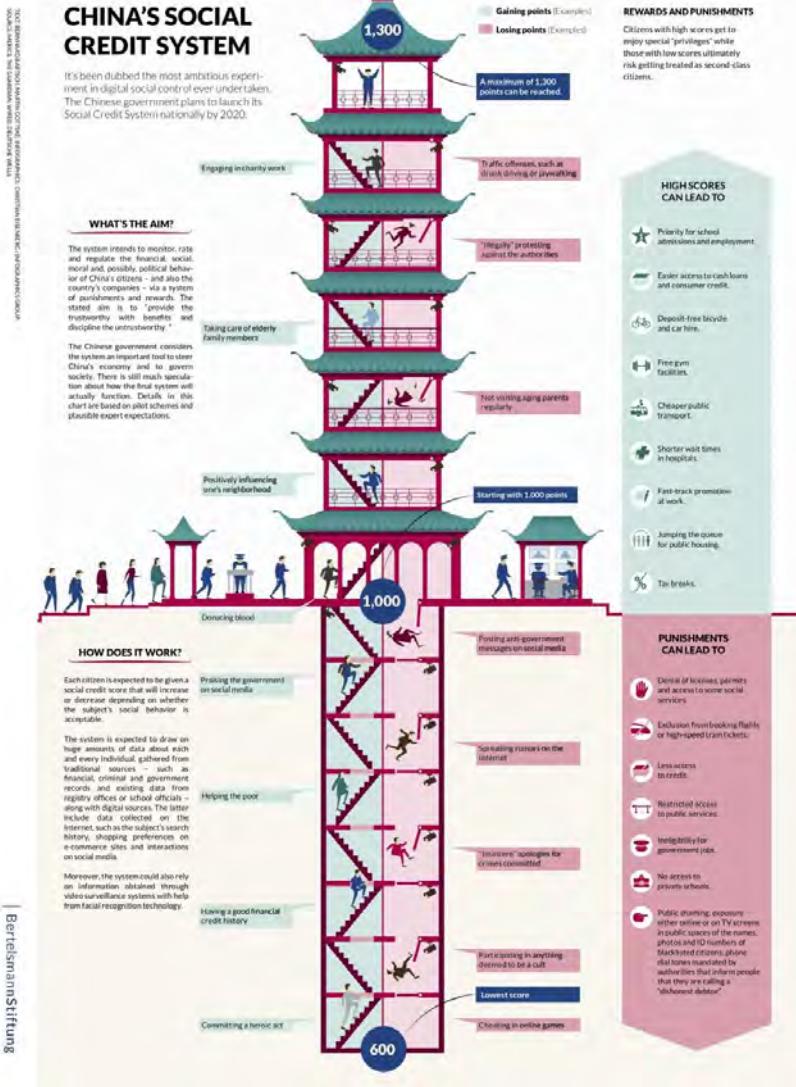
THE AI ACT





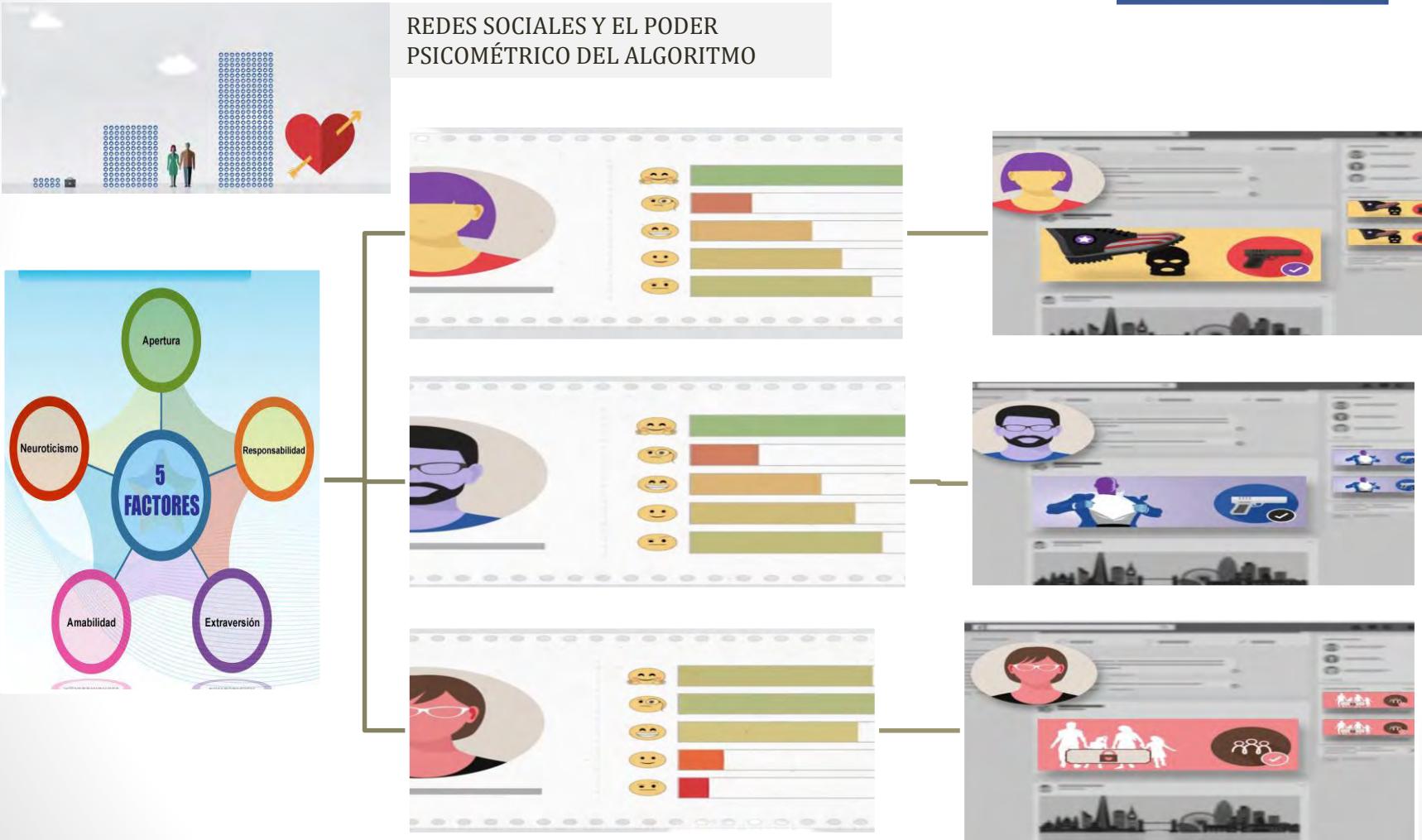
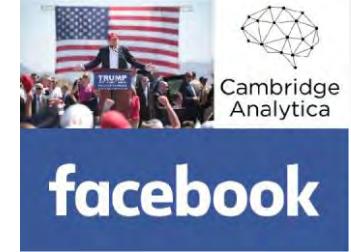
CHINA'S SOCIAL CREDIT SYSTEM

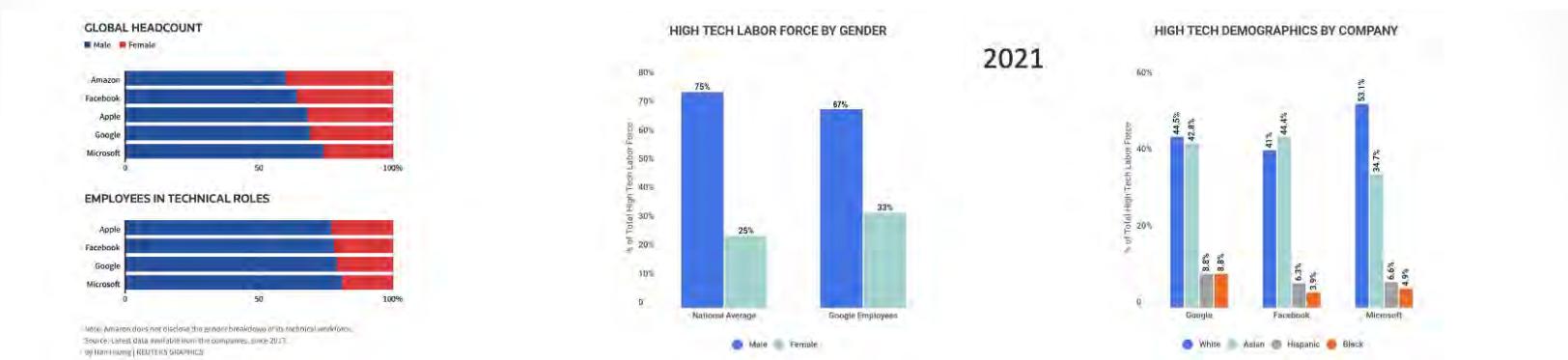
It's been dubbed the most ambitious experiment in digital social control ever undertaken. The Chinese government plans to launch its Social Credit System nationally by 2020.



COERCIÓN POLÍTICA (CONTRA LA AUTONOMÍA HUMANA)

OTRA AMENAZA A LA AUTONOMÍA HUMANA: LA MANIPULACIÓN



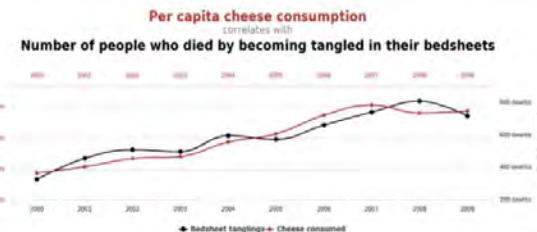
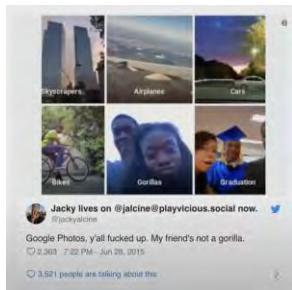


CONTRA EL PRINCIPIO DE EQUIDAD ¿Y QUÉ SE PUEDE HACER?





DATOS SESGADOS + CORRELACIONES ESPÚREAS + MACHINE LEARNING



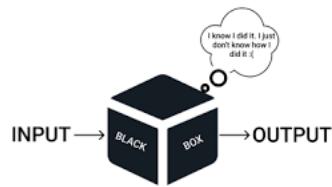
¿ALGORITMOS CIEGOS?



¿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
INCOMPRENSIBLES 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

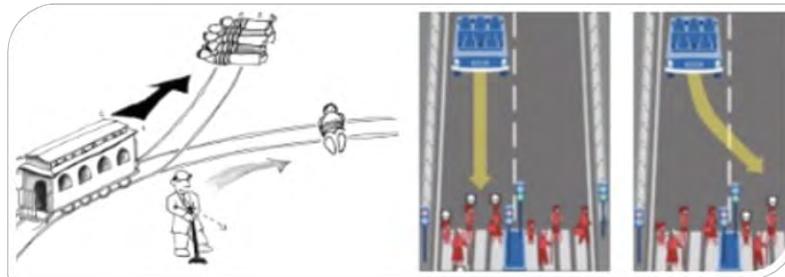
- Explicando lo que se necesitaría para dar una respuesta distinta

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...

