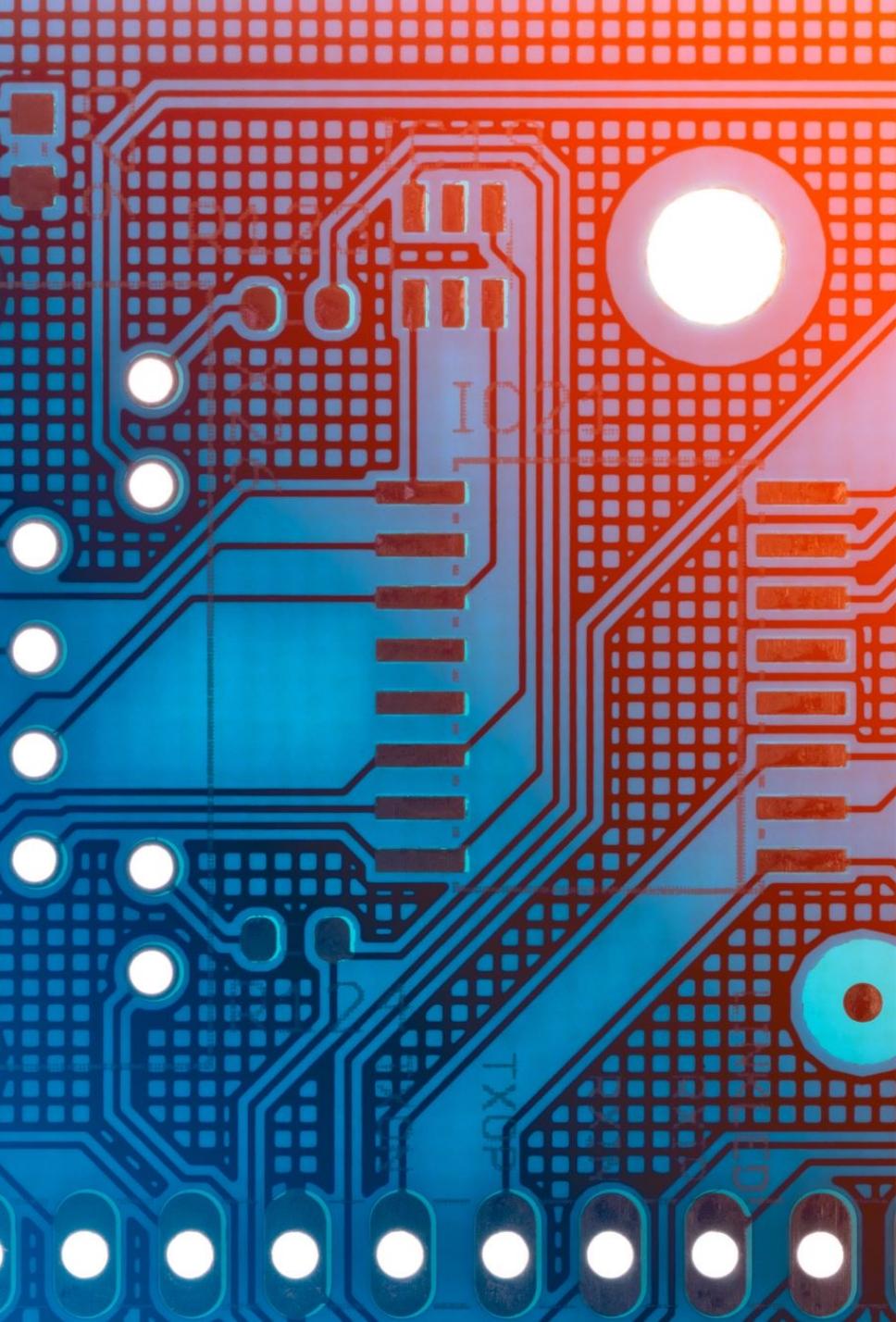


# Gobernar con algoritmos, gobernar los algoritmos

Alejandro Huergo Lora

Catedrático de Derecho Administrativo

Universidad de Oviedo



# Aclarando conceptos

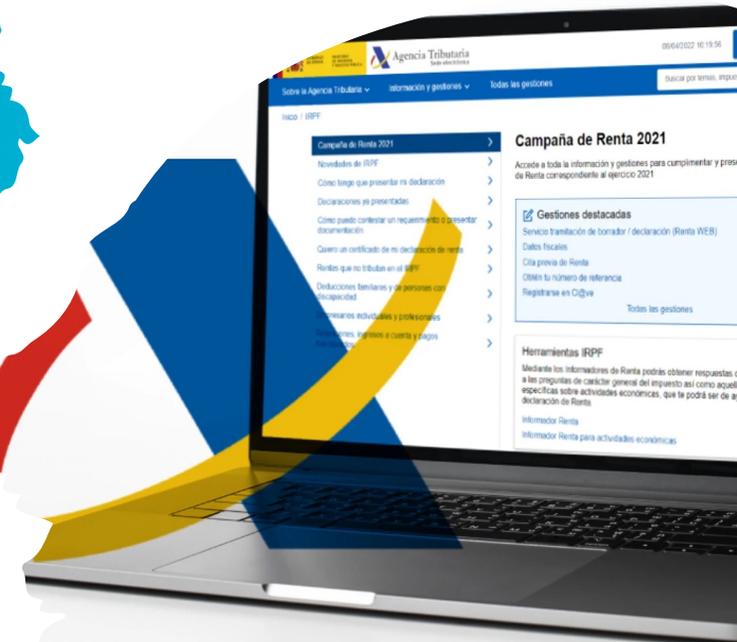
- Administración electrónica
- Actuaciones automatizadas
- Inteligencia artificial

# Clases de actuación automatizada

- Se puede automatizar la decisión o sólo el proceso: utilización de programas de gestión
- Puede afectar a expedientes individuales o a procesos complejos
- Cuando se trata de decisiones individuales, la automatización es irrelevante e “invisible”



# KIT DIGITAL

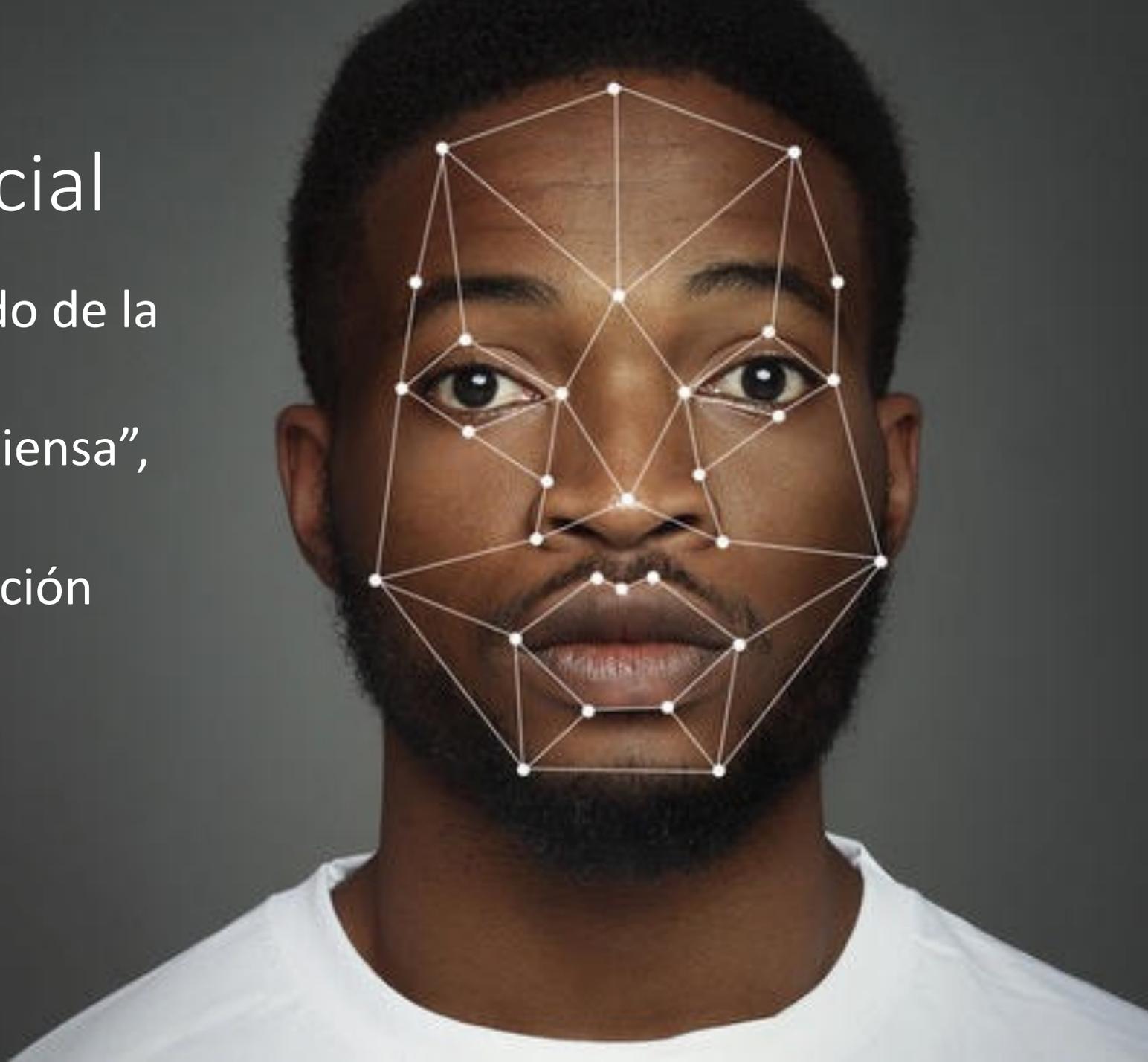


# Algoritmos difíciles de sustituir en la aplicación de la norma

- Ejemplos:
  - Aplicaciones de sorteo.
  - Mecanismos complejos de asignación de plazas en concursos de traslados.
  - “Euphemia”
- Tampoco pueden introducir factores no previstos en la norma.
- No son reglamentos.
- **Pero** es casi imposible saber cómo se habría aplicado la norma sin el algoritmo, por lo que sí hay que controlar su funcionamiento.

# Inteligencia artificial

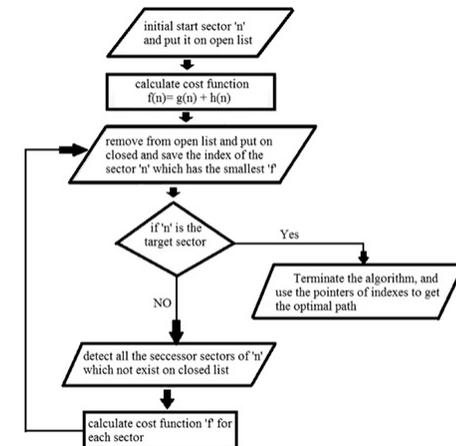
- Sí influye sobre el contenido de la actuación administrativa
- Concepto engañoso: no “piensa”, sino que “predice”
- Datos + algoritmo = predicción



# ¿Qué es un algoritmo?

---

- “Conjunto e instrucciones o reglas, definidas y no ambiguas, ordenadas y finitas, que permite, típicamente, solucionar un problema, realizar o cómputo, procesar datos o llevar a cabo otras tareas o actividades”.
- Concepto muy amplio, que excluye una calificación jurídica única.
- Algoritmos no informáticos: p.ej. una receta de cocina.
- Los ordenadores funcionan **siempre** con algoritmos.



# Predicciones basadas en datos

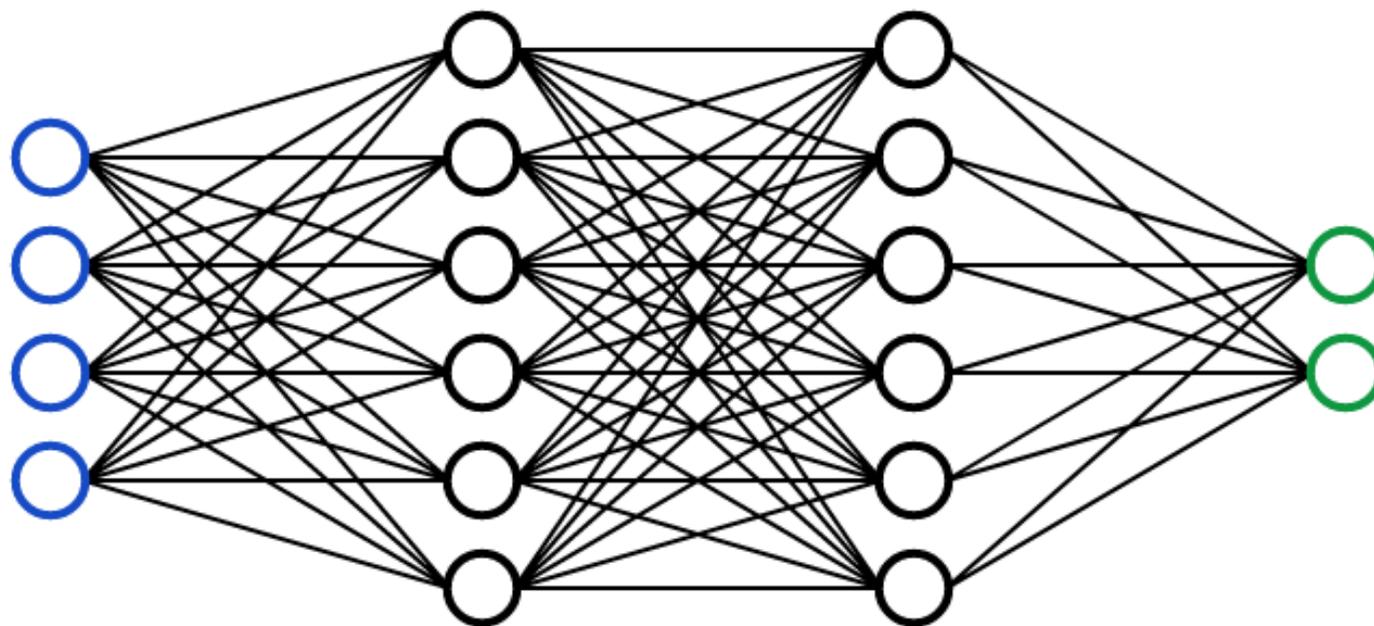
## Automatización (no IA):

- Ejemplo: cálculo de liquidaciones tributarias
- Volcamos la norma jurídica en el programa
- Introducimos los datos del caso (ingresos del sujeto)
- El programa nos da la respuesta (liquidación tributaria)

## Algoritmos predictivos (IA):

- Ejemplo: traductor de idiomas, chatGPT, programa para identificar a posibles infractores, buscador de spam...
- Queremos predecir un hecho en el futuro
- Introducimos datos pasados, tanto de ese hecho como de su contexto
- El modelo nos da la regla (no jurídica) que funcionó en el pasado y que intenta predecir el futuro

Funcionamiento práctico: redes neuronales



$y = g(x)$

Secant Lines

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

$$f(x) = \lim_{h \rightarrow 0} \frac{(x+h)^2 - x^2}{h}$$

$$= \lim_{h \rightarrow 0} \frac{x^2 + 2xh + h^2 - x^2}{h}$$

$$= \lim_{h \rightarrow 0} \frac{2xh + h^2}{h}$$

$$= \lim_{h \rightarrow 0} h(2x + h)$$

# Características

---

- El algoritmo halla **correlaciones** y formula predicciones
- La “ecuación” funciona para el pasado
- El algoritmo no busca causas, sólo sigue tendencias
- Aprende a base de observar conductas, sin “prejuicios”
- Es un sistema “caro”, que existe gracias a la actual capacidad de computación y almacenamiento
- Sustituye la decisión subjetiva por un **baremo**... establecido por el algoritmo, no por el programador

# Ejemplos de usos

---

- Identificación de situaciones que necesitan más atención (medicina, servicios sociales)
- Identificación de posibles infractores (RISCANVI, VIOGEN, inspección Trabajo)
- A quiénes orientar la publicidad
- A quiénes es necesario ofrecer una rebaja: condiciones contractuales personalizadas
- Qué riesgo tiene cada cliente que solicita un seguro
- Predicción de averías
- Saber quién tiene COVID sin hacerle un test
- Vehículo autónomo
- Identificación de superbacterias (investigación *in silico*)

# Ventajas

---

Eficiencia en el uso de los recursos

---

Aprovechar el conocimiento que dan los datos

---

Superar los límites del juicio humano

---

Evitar perder el tiempo con tareas sencillas

# Problemas

---

Algoritmos “mal hechos”: utilización de datos erróneos, anticuados o que producen un resultado incorrecto (sesgos)

---

El modelo “predice” el futuro basándose en datos del pasado: incompatibilidad con el 9.2 CE

---

Security through obscurity (problemas de control). *Every metric is gameable.*

---

*Distorsión de la Realidad por el algoritmo: profecía autocumplida*

# Algunas ideas

---

La garantía de los ciudadanos nunca ha sido la intervención humana, sino el sometimiento a reglas

---

Los humanos tienen muchos más sesgos, que son indetectables

---

No podemos exigir a la decisión “algorítmica” lo que no exigimos a la decisión humana (ej.: iniciación de procedimiento)

---

Las exigencias jurídicas (habilitación, transparencia, etc.) dependen del contexto y de los efectos del “algoritmo”

---

Es un tratamiento de datos, algunos de ellos anonimizados

# Encaje jurídico de la IA

---

- Si es un instrumento para tomar una determinada decisión, hay que respetar el marco jurídico de esa decisión
- Pero la IA tiene sus propios “peligros” que exigen una regulación
- El algoritmo no es un vehículo para crear normas jurídicas; autolimita un espacio decisonal otorgado por una norma
- Allí donde el algoritmo produce efectos jurídicos, es necesario controlar su funcionamiento y acceder a él como a cualquier otro elemento del expediente administrativo



	Sistemas que <u>no</u> influyen en el contenido de la decisión	Sistemas que <u>sí</u> influyen en el contenido de la decisión
Resoluciones regladas	Campo de aplicación típico. Irrelevantes	
Resoluciones discrecionales	No son útiles	Pueden servir para concretar algunos de sus elementos
Resoluciones con conceptos legales indeterminados	Pueden servir para aplicar el concepto (limitativo)	Puede servir para concretar algunos elementos
Iniciación procedimientos	No son útiles porque no son decisiones regladas sometidas a la verificación de unos requisitos establecidos en normas	Campo típico de aplicación. Son decisiones que están sometidas a un escaso control incluso cuando se toman sin IA
Objetivos de inspección		
Decisiones organizativas		
Actividad prestacional		

# Algunos avances de regulación

- Carta de Derechos Digitales (2021)
- Proyecto de Reglamento de Inteligencia Artificial (IA Act)
- Reforma Estatuto de los Trabajadores (2021): “Ser informado por la empresa de los parámetros, reglas e instrucciones en los que se basan los algoritmos o sistemas de inteligencia artificial que afectan a la toma de decisiones que pueden incidir en las condiciones de trabajo, el acceso y mantenimiento del empleo, incluida la elaboración de perfiles”
- Anteproyecto Ley Medidas de Eficiencia Digital Justicia
- Sentencia BVerfG 16-2-2023 (Leyes de Hesse y Hamburgo)

# Ley 15/2022, integral para la igualdad de trato

## **Artículo 23. Inteligencia Artificial y mecanismos de toma de decisión automatizados.**

1. En el marco de la Estrategia Nacional de Inteligencia Artificial, de la Carta de Derechos Digitales y de las iniciativas europeas en torno a la Inteligencia Artificial, las administraciones públicas **favorecerán** la puesta en marcha de mecanismos para que los algoritmos involucrados en la toma de decisiones **que se utilicen en las administraciones públicas** tengan en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas, siempre que sea factible técnicamente. En estos mecanismos se incluirán su diseño y datos de entrenamiento, y abordarán su potencial impacto discriminatorio. Para lograr este fin, se **promoverá** la realización de evaluaciones de impacto que determinen el posible sesgo discriminatorio.
2. Las administraciones públicas, en el marco de sus competencias en el ámbito de los **algoritmos involucrados en procesos de toma de decisiones**, priorizarán la transparencia en el diseño y la implementación y la capacidad de interpretación de las decisiones adoptadas por los mismos.
3. Las administraciones públicas y las empresas **promoverán** el uso de una Inteligencia Artificial ética, confiable y respetuosa con los derechos fundamentales, siguiendo especialmente las recomendaciones de la Unión Europea en este sentido.
4. Se promoverá un sello de calidad de los algoritmos.



# Proyecto de Reglamento Europeo

# Concepto

- Concepto: generar “información de salida” como predicciones, recomendaciones o decisiones que influyan en los entornos con los que interactúa.
- Funciona con elementos de autonomía
- No está claro si se establecerá un Anexo con técnicas concretas

# Prácticas prohibidas

- Técnicas subliminales que alteren sustancialmente el comportamiento de una persona, de un modo que provoque o sea probable que provoque perjuicios físicos o psicológicos a esa persona o a otra.
- Que, aunque no sea subliminal, se aprovecha de alguna de las vulnerabilidades de un grupo específico de personas debido a su edad o a su discapacidad para lograr lo mismo que en el apartado anterior.
- Sistemas de “recompensa” cuando son utilizados por autoridades públicas (o no sólo). Cuando el trato negativo sea desproporcionado o se utilice en contextos que no guarden relación
- Sistemas de identificación biométrica remota en tiempo real en espacios de acceso público con fines de aplicación de la ley, salvo... Además de proporcionalidad, se exige autorización previa judicial o administrativa.

# Prácticas que deben advertirse

- Chatbot: advertir
- Sistemas de reconocimiento de emociones: advertir salvo uso para detección e investigación de infracciones penales
- Deepfake: advertir, salvo libertad de expresión



# Sistemas de alto riesgo

- Sistemas unidos a productos regulados (Anexo II)
- Anexo III: los que se utilizan con efectos jurídicos, o perjudiciales, etc.
- Infraestructuras críticas, biometría, educación, empleo, acceso a servicios esenciales, aplicación de la ley, control de fronteras, justicia. *No prohíbe la policía predictiva*

# Requisitos

- Sistema de gestión de riesgos
- Calidad de datos
- Documentación técnica
- Registros
- Transparencia: “prospecto”
- Vigilancia humana
- Precisión, solidez y ciberseguridad
- *Requisitos ambiciosos pero definidos de forma indeterminada. El nivel concreto deberá determinarse en cada caso. No incluye acceso al “código fuente”*

## ¿Cómo se vigila?

- Se pueden aprobar normas de armonización (varios niveles)
  - Productos sometidos a autorización: dentro de ese procedimiento
  - Mayoría de sistemas de alto riesgo: control interno (no es obligatorio contar con un organismo certificado)
  - Códigos de conducta para sistemas que no son de alto riesgo
  - Multas de hasta el 6% del volumen de negocio total anual mundial de la empresa infractora
- 

# Conclusiones

- Es una normativa preventiva, que no garantiza que el sistema no produzca daños
- Intenta evitar daños a la salud, la seguridad y los derechos fundamentales
- Las autoridades nacionales pueden retirar un sistema del mercado
- Aunque es armonizadora, no impide la aplicación de normas propias del sector en que se use el sistema de IA

## ¿El futuro?

- Riesgo de frenar la innovación
- Posible liderazgo europeo en la regulación
- ¿Generará confianza esta regulación?
- ¿Es necesaria la colaboración de hackers o informantes internos?
- ¿Se cumplirán las expectativas generadas por la IA en el sector público?
- Amplias necesidades de formación

AI for  
Social Good

