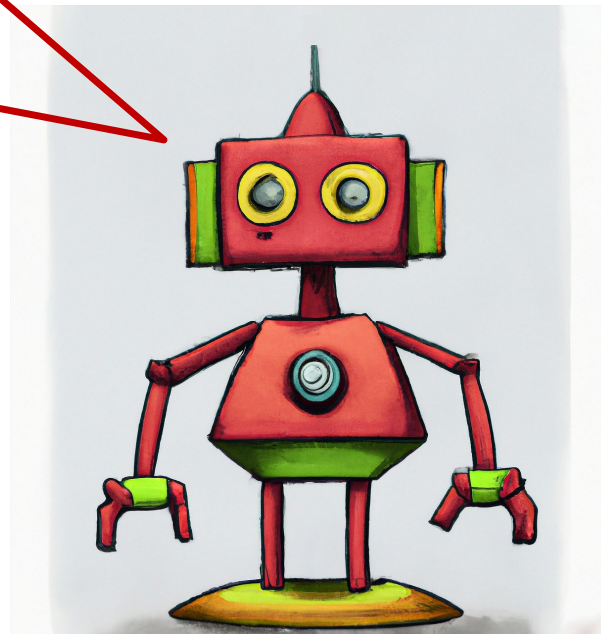


Instituto de Ciencia de los Datos e Inteligencia Artificial (DATAI)  
Universidad de Navarra  
30 de octubre de 2024



## Ética y Tecnología: explicabilidad y responsabilidad en el uso de herramientas de inteligencia artificial

Gonzalo Génova  
[ggenova@inf.uc3m.es](mailto:ggenova@inf.uc3m.es)  
Departamento de Informática  
Universidad Carlos III de Madrid



## Una historia de uso irresponsable (¿inocente?)



Elizabeth Laraki  
@elizlaraki

A week ago, I came across an altered, more suggestive photo of myself online.

It turned out to be **an innocent use of AI** with unintended consequences.

I'm talking at a conference later this year (on UX+AI).

**Someone edited my photo** to unbutton my blouse and reveal a made-up hint of a bra or something else underneath. 🙄

Analysis and proposal:

1. Consider if AI is a good tool for the task.
- 2. Critically evaluate the output.**
3. If relevant, ask for consent.



<https://x.com/elizlaraki/status/1848779238708760726>

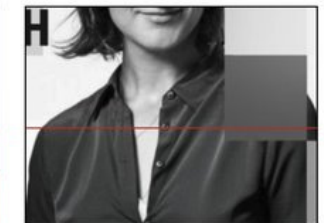


Original photo



Square cropped photo

Image cut off at second button



Gen-AI filled photo

Lower part of image generated from second button down

**¿Cuáles fueron los motivos?  
¿Hubo motivos?**

## Un futuro... ¿no tan lejano?



Illustrated by Virgil Finlay

Isaac Asimov  
*La sensación de poder*  
1958

*Graphitics was a startlingly new idea!*

*So revolutionary, in fact, it rocked the top army brass.*

*Imagine computing—without a computer!*



**¿Qué tipo de responsabilidad puede asumir un ingeniero que utiliza herramientas de inteligencia artificial para desarrollar su trabajo?**



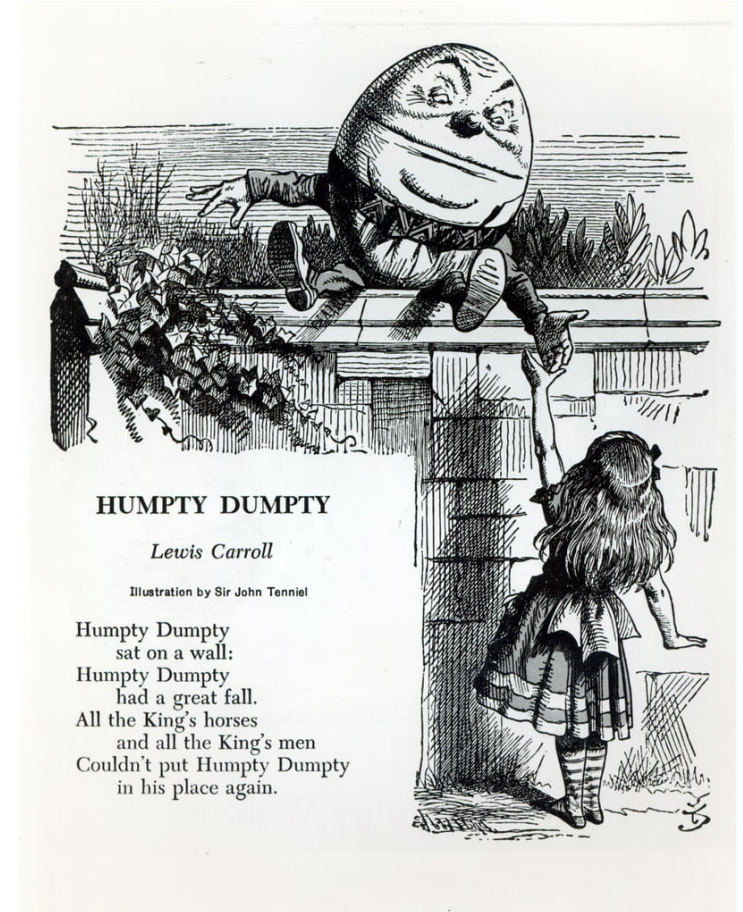
## Una respuesta muy “razonable”

—¿Por qué está usted sentado aquí fuera tan solo? — dijo Alicia, que no quería meterse en discusiones.

— ¡Hombre! Pues **porque no hay nadie que esté conmigo** —exclamó Humpty Dumpty—. ¿Te creíste acaso que no iba a saber responder a eso? Pregunta otra cosa.



Bill Watterson  
Calvin & Hobbes



### HUMPTY DUMPTY

Lewis Carroll

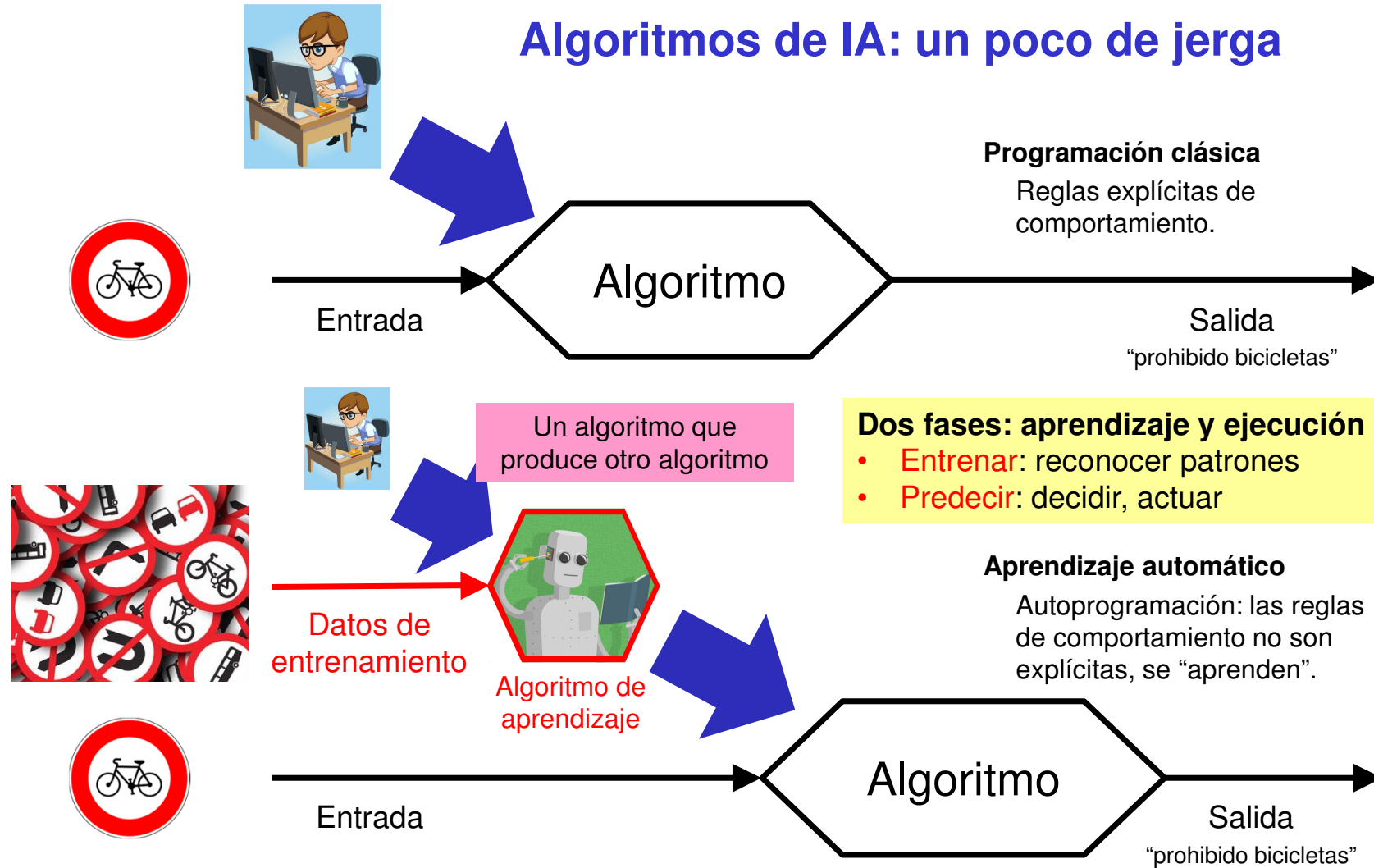
Illustration by Sir John Tenniel

Humpty Dumpty  
sat on a wall:  
Humpty Dumpty  
had a great fall.  
All the King's horses  
and all the King's men  
Couldn't put Humpty Dumpty  
in his place again.

Lewis Carroll

A través del espejo y lo que Alicia encontró allí (1871)

## Algoritmos de IA: un poco de jerga



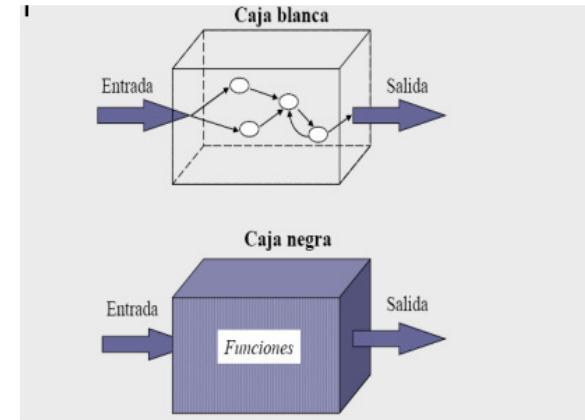
## Diversas técnicas de inteligencia artificial

- **IA simbólica (deductiva, *top-down*)**
  - Razonamiento automático
  - Sistemas expertos

*reglas a priori*
  
- **IA subsimbólica (inductiva, *bottom-up*)**
  - Redes neuronales
  - Algoritmos genéticos

*reglas a posteriori*
  
- **Sistemas mixtos**
  - Traductor automático

*gramática y estadística*



<http://www.lsi.us.es/docencia/get.php?id=361>



## ¿Qué fiabilidad ofrece cada uno?

### *Reglas a priori*

- Basadas en la autoridad de expertos.
- No implica fácil **explicabilidad**.

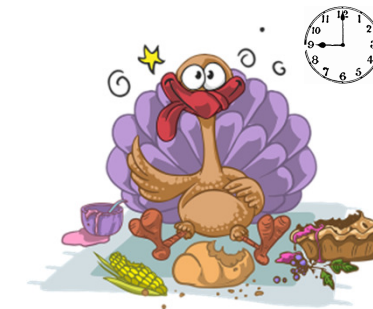


37% (B) - 35% (E) - 28% (N)



### *Reglas a posteriori*

- Regularización: “siempre ha sido así”.
- El problema de la inducción.



El pavo inductivista  
Bertrand Russell



## Explicabilidad: ¿Causa o Razón?

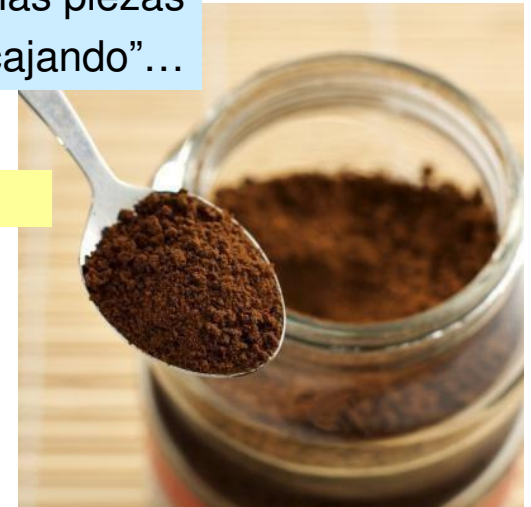
¿Por qué este tornillo está aquí?  
Diseño racional: **alguien sabe la respuesta.**



¡Fui yo!

En cambio, cuando las piezas van “encajando”...

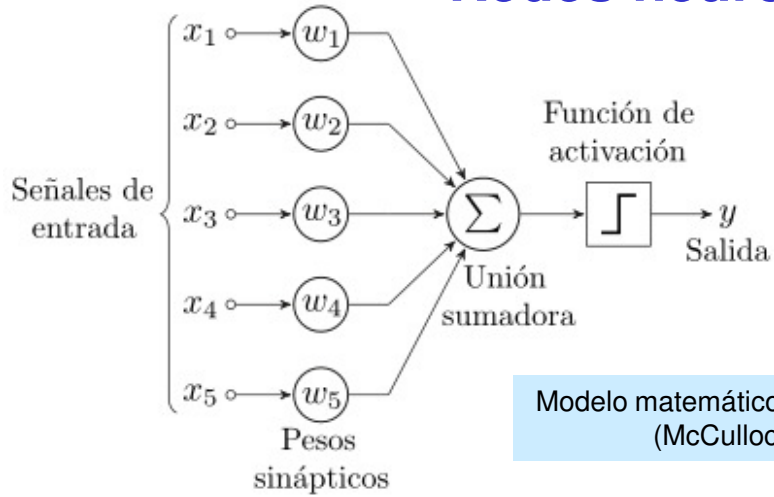
Granularidad



—¿Por qué está usted tan solo?  
— Pues porque no hay nadie que esté conmigo.

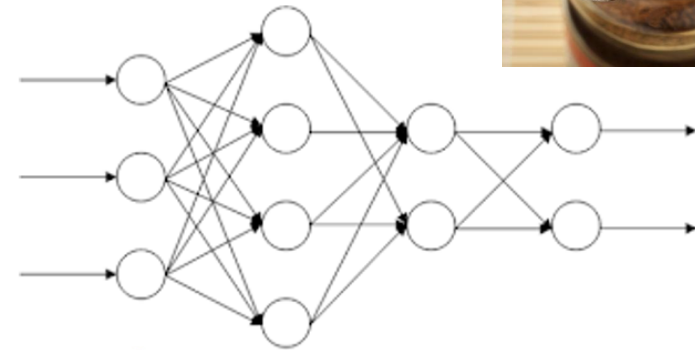


# Redes neuronales: encaje de piezas evolutivo



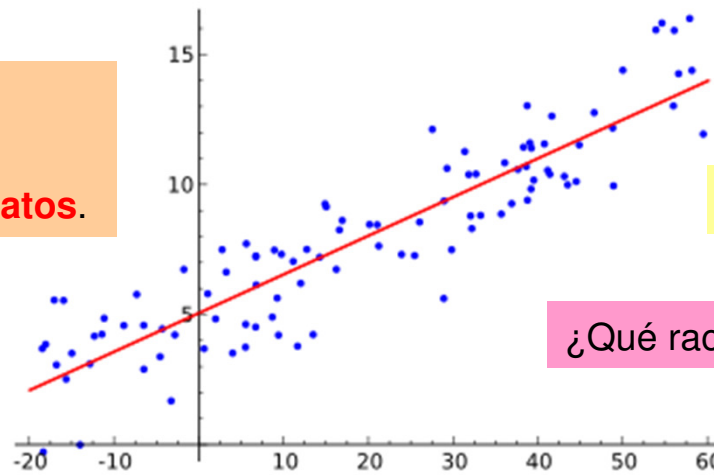
Modelo matemático de una neurona artificial (McCulloch & Pitts, 1943)

Red de neuronas multicapa



Entrenamiento, aprendizaje.

- Un problema de optimización.
- Racionalidad del **ajuste a los datos**.



$$y = a \cdot x + b$$

¿Qué racionalidad hay en una regresión lineal?

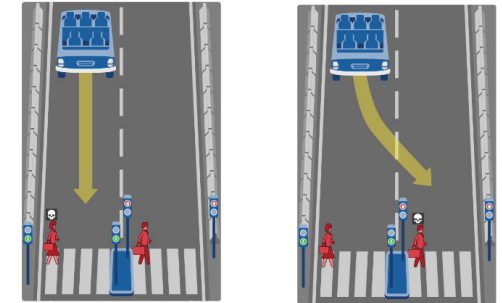
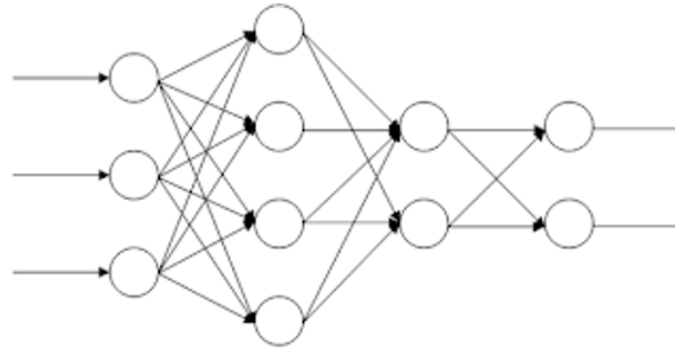
## Aprendizaje automático: racionalidad imitativa

El resultado del aprendizaje automático es una fórmula, un algoritmo, **una relación funcional**.

¿Por qué funciona la fórmula?

¿Cuál es su **justificación**?

- Porcentaje de éxito.
- Ajuste con los datos de entrenamiento.
- Poder predictivo.



Vehículos autónomos:  
¿a quién atropello?  
(The Moral Machine)



Recomendaciones **razonadas**:

- médicos,
- inspectores de hacienda,
- profesores...



**No es aceptable tomar una decisión moral con base en el resultado de una caja negra que ha proporcionado un número no verificable.**

## Dos tipos muy distintos de inexplicabilidad

### ¿Es responsable la imitación?

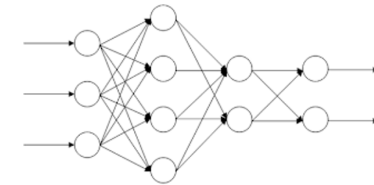
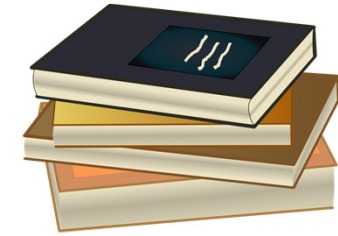
La imitación puede ser razonable...  
si tengo a quién preguntar,  
y me puede dar una explicación.

#### Inexplicabilidad **práctica y relativa**:

- Diseño racional complejo.
- Pero alguien sabe la respuesta.

#### Inexplicabilidad **esencial**:

- No hay diseño racional.
- No hay nadie a quien preguntar.



¡Queremos saber cuál es el diseño!  
¡Cuáles son los motivos!

*Nam causarum finalium inquisitio sterilis est, et,  
tanquam virgo Deo consecrata,  
nihil parit.*

La investigación de las causas finales es una cosa estéril,  
no parirá nada, igual que una virgen consagrada a Dios.

(De Augmentis Scientiarum, III, 5)

El mero reconocimiento de un **patrón** en  
un conjunto de fenómenos no permite  
asegurar que haya una **intención** de  
diseño que dé origen a esas regularidades.



Francis Bacon  
(1561-1626)



## Responsabilidad profesional

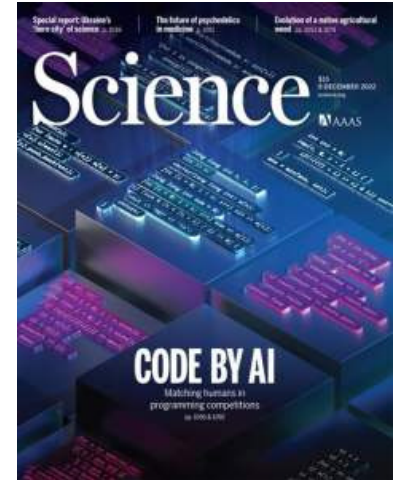


Todo sistema automático  
tiene consecuencias éticas.

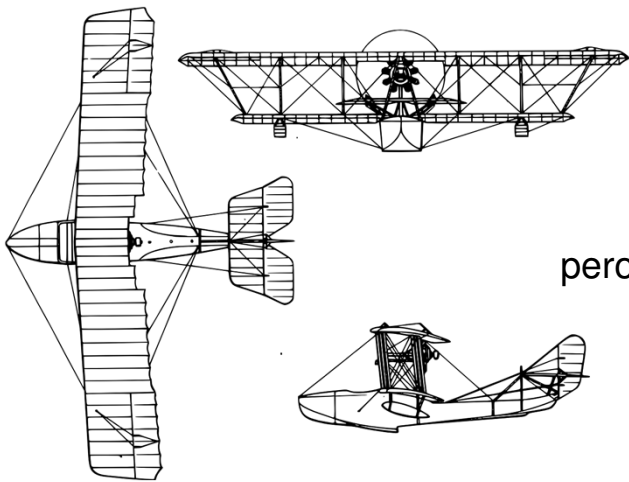
Industria de caja negra



**¿Cómo voy a hacerme responsable de mi trabajo si no sé explicarlo, ni a los demás ni a mí mismo?**



Diciembre 2022




Si se cae el puente porque fue diseñado **como dijo la máquina**, pero nadie entiende verdaderamente ese diseño (nadie sabe **explicarlo**), ¿quién puede hacerse **responsable**?



## ¿A dónde vamos?

# El truco definitivo para saber si una gasolinera está al lado de la autovía: solo es necesario fijarse bien en el cartel

ANTONIO PÉREZ SAZ / NOTICIA / 26.09.2023 - 20:31H 

El vídeo que evidencia la clave de la distancia en el cartel ubicado en la vía de alta capacidad tiene ya más de un millón de visitas y **son muchos los usuarios que han comprobado esta tesis**, por lo que es un truco que funciona, además el que te hemos comentado acerca del color lleva años demostrándose.

¿Un curiosísimo fenómeno de la naturaleza?



Señal de desvío a una gasolinera en la Autovía Mudéjar. / Wikipedia

<https://www.20minutos.es/motor/movilidad/truquillo-saber-gasolinera-cerca-autovia-cartel-5172350/>

## De máquinas e intenciones

Reflexiones sobre la tecnología, la ciencia y la sociedad

<https://demaquinaseintenciones.wordpress.com/>



**¡GRACIAS!**