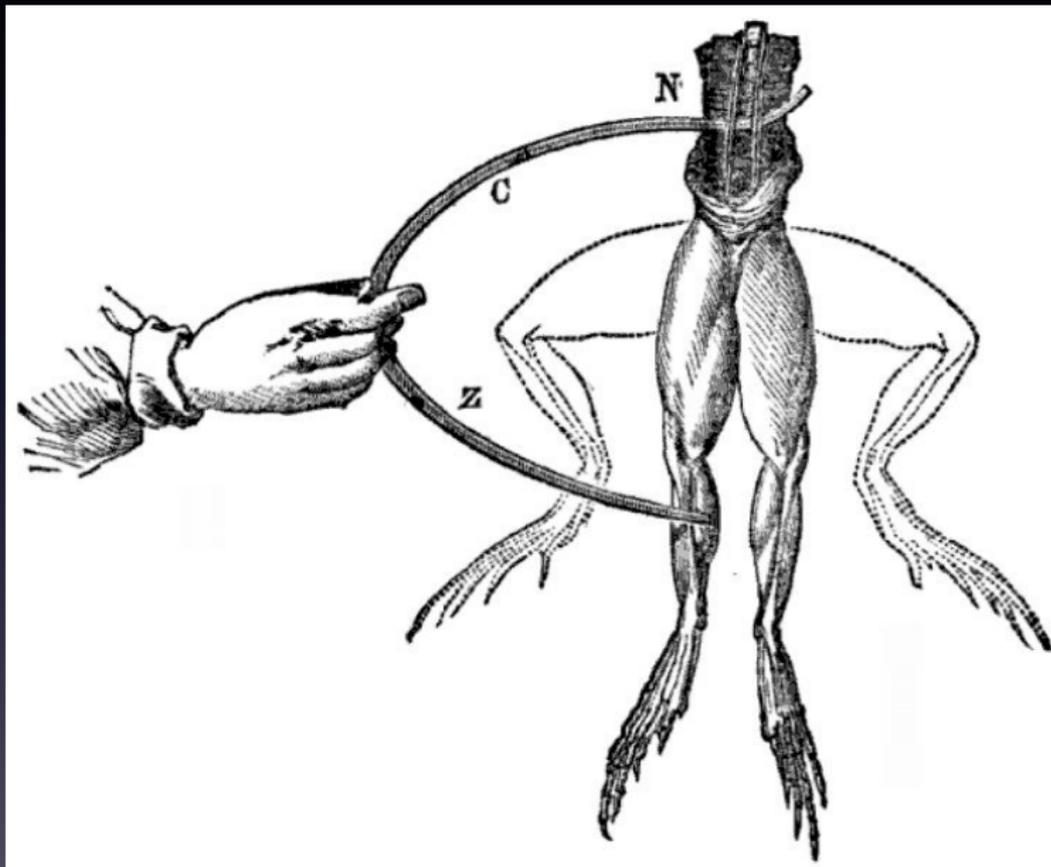


Inteligencia artificial, mecánica estadística y ficción

Miguel Hoyuelos



Ciencia y ciencia ficción





Luigi Galvani



Alessandro Volta



Giovanni Aldini (1762-1834)

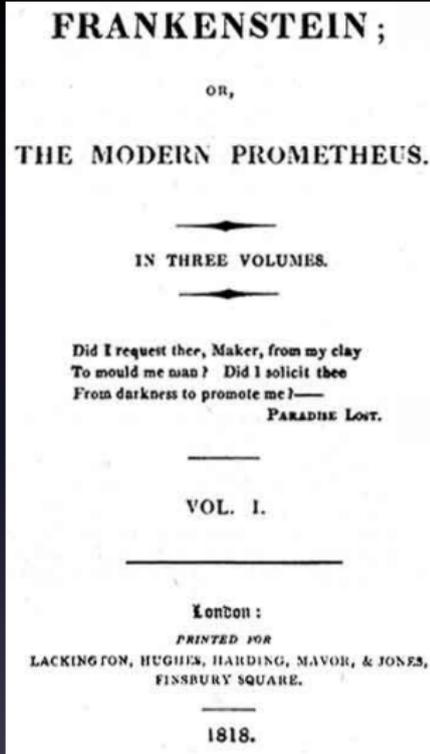




A GALVANISED CORPSE



Mary Shelley



Inteligencia artificial

Inteligencia artificial débil: aplicable solo a un tipo de problemas
(omnipresente)

Inteligencia artificial fuerte: tendría la versatilidad y la capacidad de adaptación de una inteligencia humana, auto-consciente
(no existe)

Ginoides y androides de ficción

*Villiers
de l'Isle-Adam*

L'Eve future



GF-Flammarion

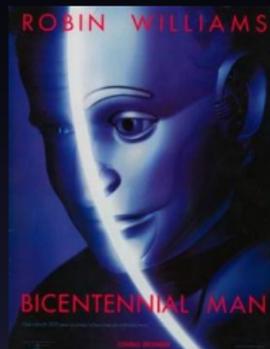
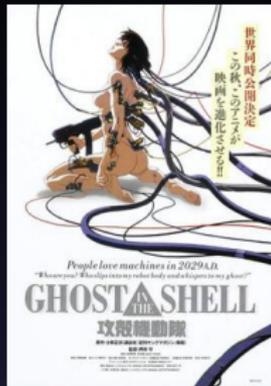
Metropolis (Lang, 1927)



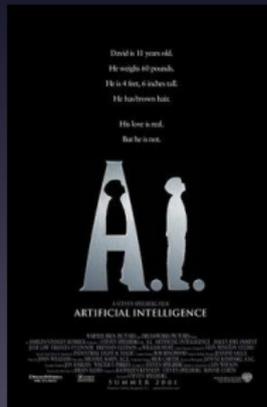
Conflicto: la nueva inteligencia queda fuera de nuestro control, se torna una amenaza.

Otros conflictos

Fusión
humano-
máquina en
futuro ciber-
punk



Niño androide



Últimos años



2014



2013



2015



2016



2015



2013



1977

DOMMONALL GLEESON ALICIA VIKANDER and OSCAR ISAAC

ex machina

WHAT HAPPENS TO ME IF I FAIL YOUR TEST?



UNIVERSAL PICTURES IN ASSOCIATION WITH A24 PRESENTS "EX MACHINA" A FILM BY ALEX GARLAND CASTING BY JESSICA ROBERTSON COSTUME DESIGNER JAMES HARRIS EXECUTIVE PRODUCERS JAMES HARRIS AND JAMES HARRIS PRODUCED BY JAMES HARRIS AND JAMES HARRIS WRITTEN BY ALEX GARLAND AND ALEX GARLAND DIRECTED BY ALEX GARLAND CASTING BY JESSICA ROBERTSON COSTUME DESIGNER JAMES HARRIS EXECUTIVE PRODUCERS JAMES HARRIS AND JAMES HARRIS PRODUCED BY JAMES HARRIS AND JAMES HARRIS WRITTEN BY ALEX GARLAND AND ALEX GARLAND DIRECTED BY ALEX GARLAND

R

MPAA

FILM

MPAA

coming soon

© 2015

UNIVERSAL PICTURES

IN ASSOCIATION WITH

A24

ALL RIGHTS RESERVED

UNIVERSAL PICTURES

IN ASSOCIATION WITH

A24

ALL RIGHTS RESERVED

2015

1968



HAL

2013



Samantha

¿Qué dice la ciencia?

¿Qué dice la ciencia?

¿Qué demonios es la
conciencia?

What in the world is consciousness?, Zeman (2005)

Conciencia

Vigilia: un individuo es consciente si tiene los ojos abiertos o es capaz de mantener una conversación

Discernimiento: estar consciente de algo en particular y conocer su significado; va más allá de la vigilia; más difícil de establecer.

Auto-conciencia

Propia: capacidad para reconocernos no solo como un cuerpo sino también como una mente, como un sujeto con experiencia, con deseos, creencias e intereses propios.

Discernimiento del discernimiento de los otros: reconocer que los demás tienen la misma capacidad, que *adivinamos* con limitaciones similares a las propias.

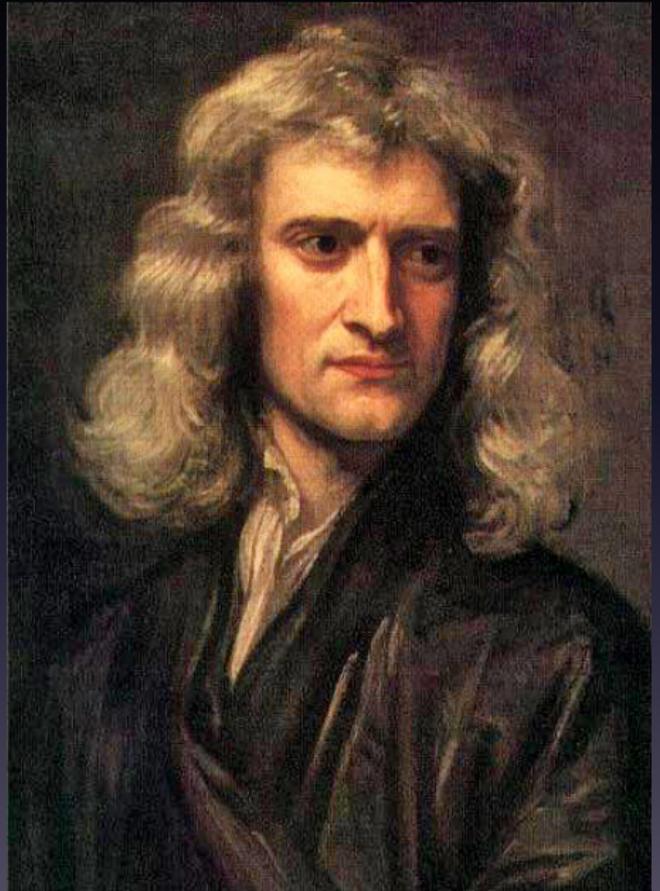
Problema

La conciencia es inobservable: La conciencia es un proceso interno, oculto, privado, solo accesible al que lo experimenta e inaccesible para el resto.

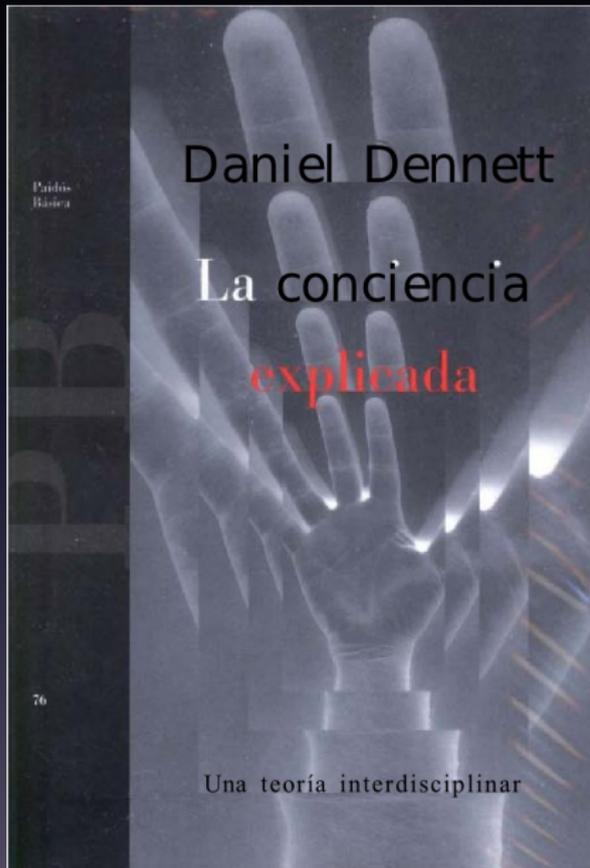
“El problema difícil de la conciencia”: Entre la observación de la actividad cerebral y la experiencia interna de la conciencia parece haber una brecha imposible de superar (Chalmers 1995).

Quale: una experiencia consciente

No es tan fácil determinar con qué acciones o de qué modo la luz produce el fantasma del color en nuestras mentes (Newton, 1672).



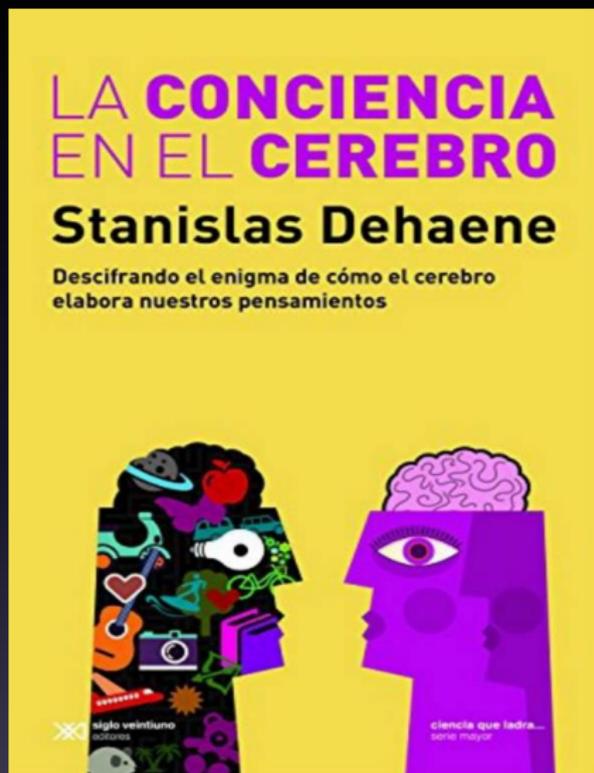
Dennett “La conciencia es una ilusión”



1991

Dehaene, 2014

- quale \rightarrow vitalismo
- fácil \leftrightarrow difícil
- espacio de trabajo global



Global workspace theory (GWT, B. Baars 1988, 2002)

Integrated information theory (IIT, G. Tononi 2004)

Objeciones

- Por lo que respecta a la conciencia, la apariencia es la realidad; el hecho de que esa apariencia exista la hace real (Searle 1997)
- Si la conciencia es una ilusión, qué o quién experimenta esa ilusión.

La conciencia artificial

Máquina consciente — Inteligencia artificial fuerte

Paradigmas

Funcionales: basados en el comportamiento de la máquina sin importar cómo está constituida.

Anatómicos: basados en lo que se conoce de la anatomía y el funcionamiento del cerebro.

Antes de 2020 se creía que en 20 o 30 años alguno de estos modelos resolvería el problema difícil de la conciencia y daría lugar a una máquina consciente. Ahora se supone que será antes.

¿Es posible la máquina consciente?

Las computadoras carecen de espontaneidad o libre albedrío: están a merced de sus usuarios. Los humanos tienen la capacidad de innovar, desobedecer y hacer trampa. Tales normas están motivadas y representadas por emociones sociales, como la empatía, la simpatía, la compasión, la vergüenza, el orgullo, la confianza y la desconfianza, que están más allá del alcance de las máquinas. (Bunge 2010, *Matter and Mind*).

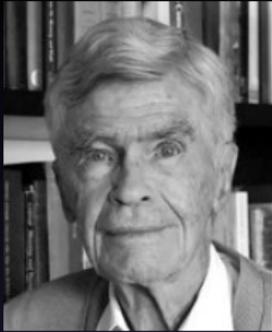
¡No!



“Sí” para un materialista eliminativo

- Conciencia y libre albedrío: conceptos mal definidos o ficticiales.
- Nosotros aparentamos tener libre albedrío, algo que una máquina también podría hacer.
- Por ejemplo, una máquina es determinista, pero puede tener un comportamiento muy parecido a uno aleatorio.

La respuesta depende de qué entendemos por conciencia o libre albedrío.



“En general, para imitar cualquier cosa, comience por aprender sobre el artículo genuino”

Red de neuronas vivientes.

Patrones:

- oscilaciones
- sincronización
- ondas
- **fluctuaciones críticas**

Fluctuaciones críticas

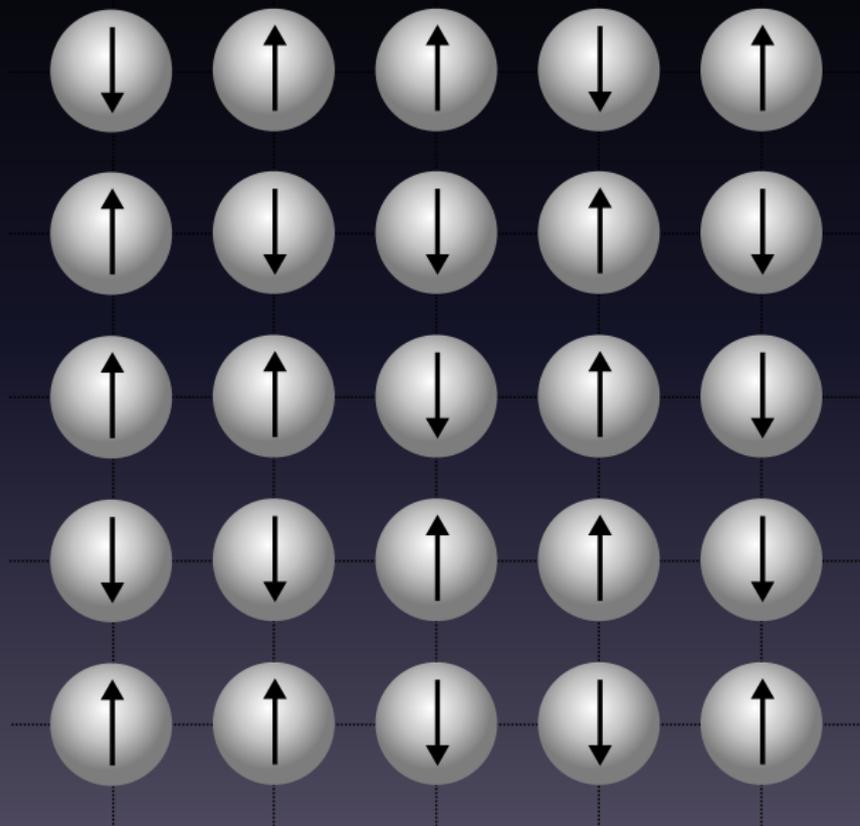
Joaquín Marro
Dante R. Chialvo

La mente es crítica

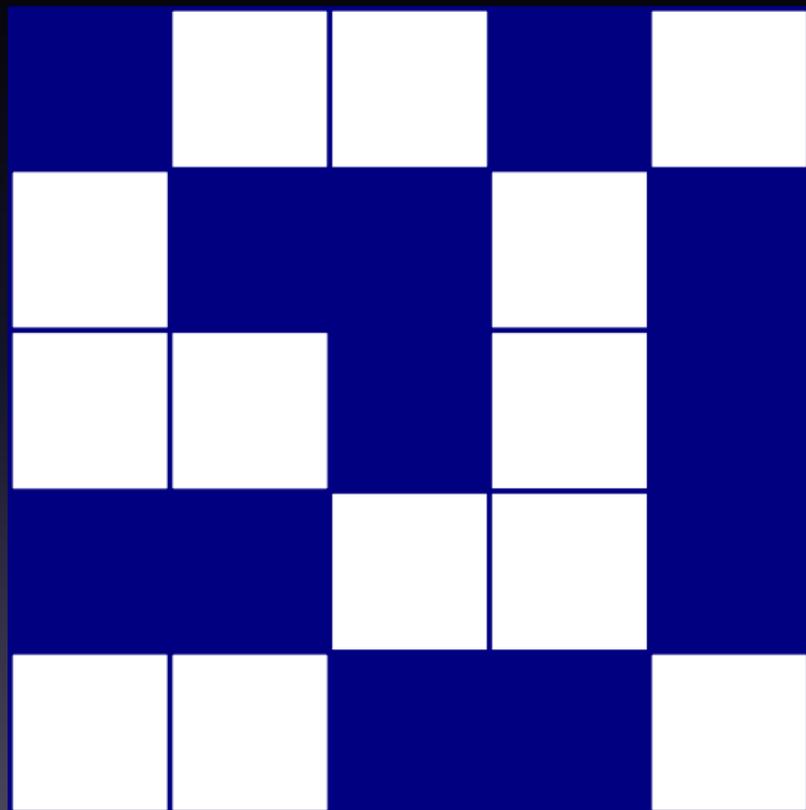
Descubriendo la admirable
complejidad del cerebro



Fluctuaciones críticas



Fluctuaciones críticas



Modelo Ising

Invariancia

Modelo de Ising

$T < T_c$



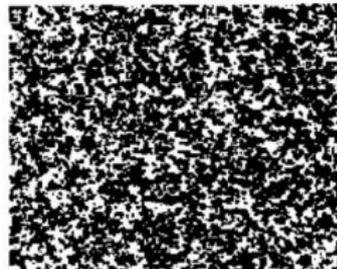
Subcritical

$T \sim T_c$



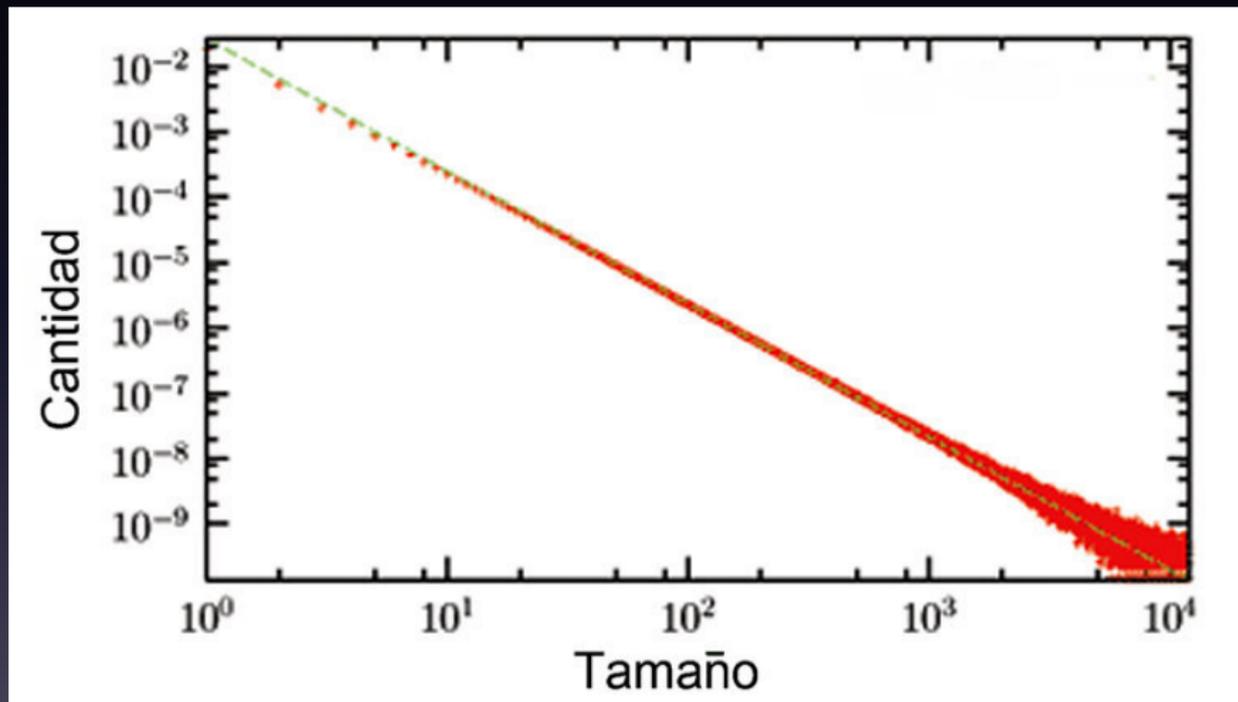
Critical

$T > T_c$



Supercritical

Cantidad vs. tamaño n , Ising a T_c



Criticalidad auto-organizada

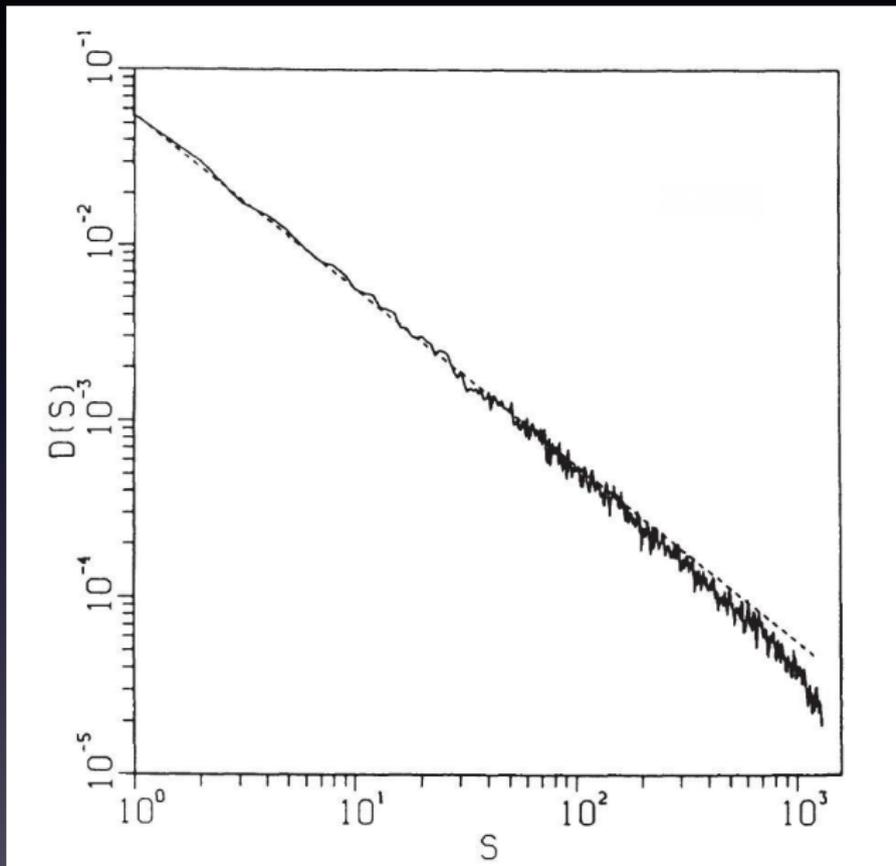


Se agregan
granos
(aumenta
pendiente)

Pendiente
crítica

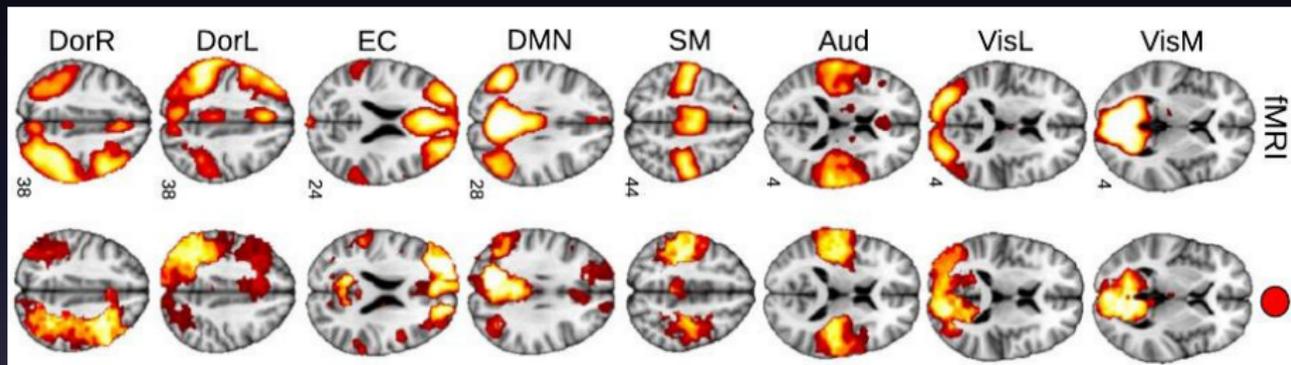
Avalancha
(reduce
pendiente)

Cantidad vs. tamaño de avalancha



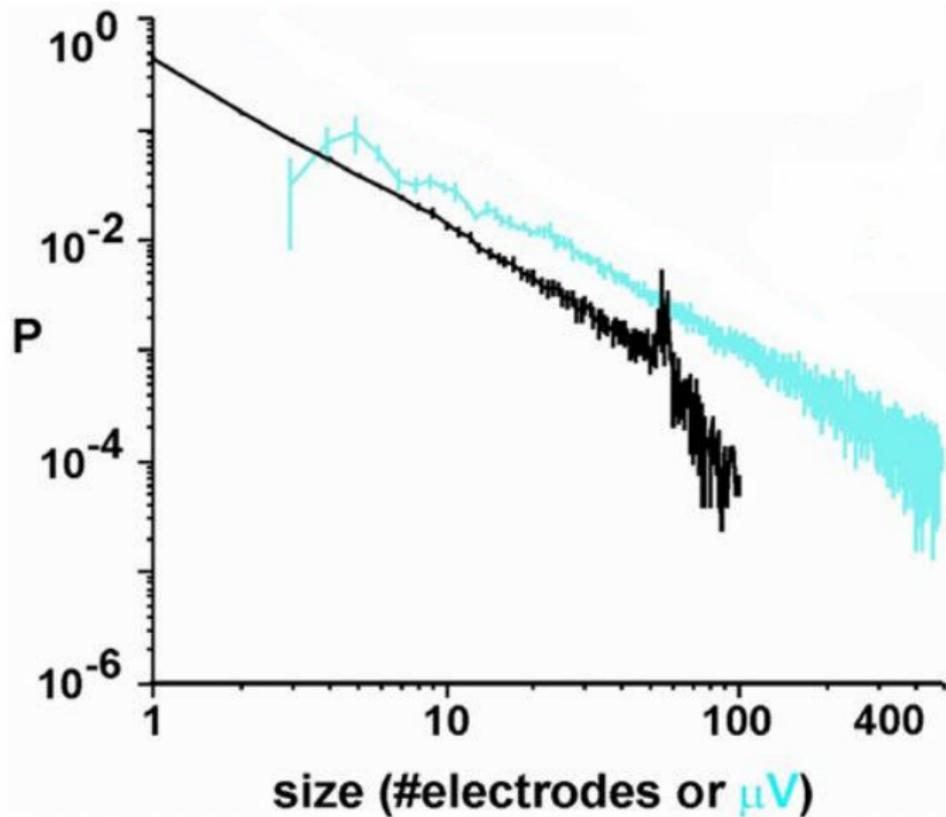
Bak, Tang, Wiesenfeld, PRA 38, 364 (1988)

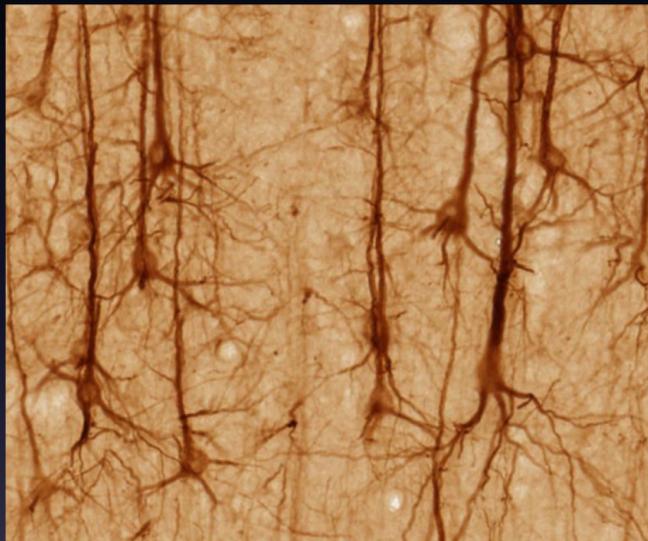
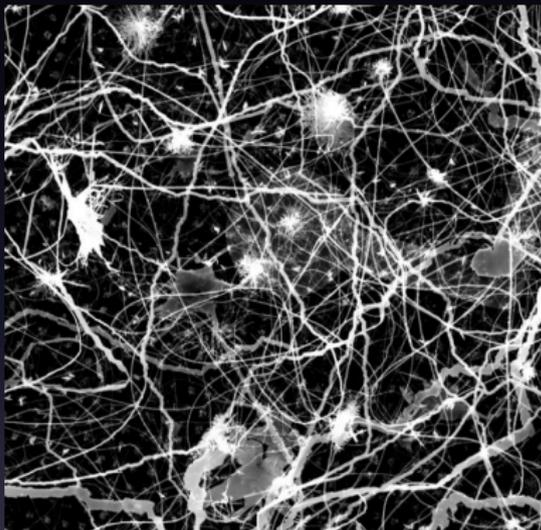
Fluctuaciones en cerebro en reposo



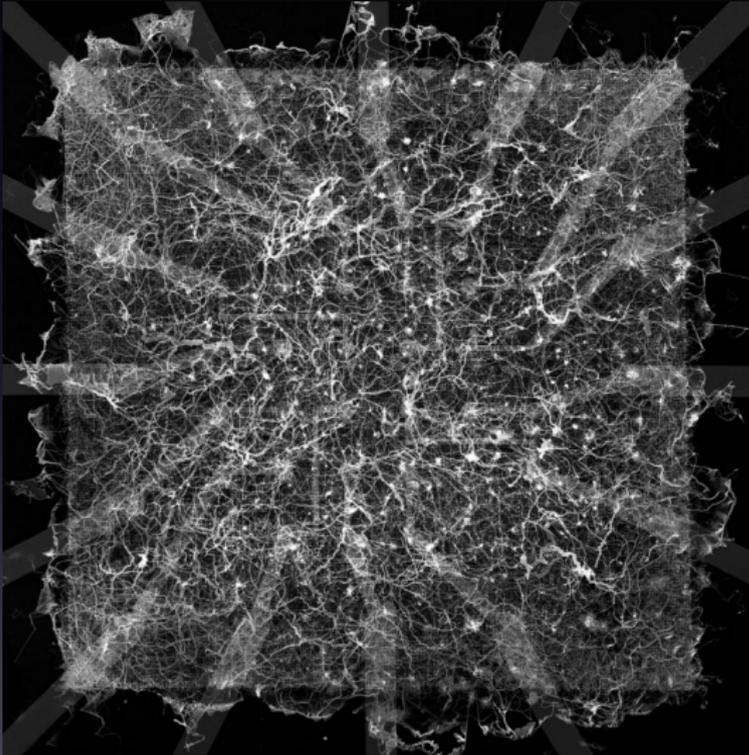
Haimovici, Tagliazucchi, Balenzuela, Chialvo, PRL 110, 178101 (2013)

Prob. avalancha neuronal vs. tamaño





Red de nanoalambres de plata



Tiene fluctuaciones críticas

Neuromorphic computing

Atomic switch networks as complex adaptive systems, Scharnhorst et al., Japanese Journal of Applied Physics 57, 03ED02 (2018)

In materia reservoir computing with a fully memristive architecture based on self-organizing nanowire networks, Milano, G. et al., Nat. Mater. 2, 1 (2021).

El manejo del lenguaje

NLP: Procesamiento de lenguaje natural

GPT: Generative Pre-trained Transformer

- Ver 3, mayo 2020: Red neuronal con 175.000.000.000 parámetros. Entrenada con textos, libros, páginas, o artículos de acceso libre (toda la Wikipedia es el 3 %); “asombroso”, “espeluznante”, “humillante”, “más que un poco aterrador”
- Ver 4, marzo 2023: más confiable y creativo; no se reveló información técnica. Entrenado con textos e imágenes. Supera sistemas expertos para medicina, aprueba exámenes universitarios. Tendencia a alucinar o confabular.
- Otros: Gemini, Claude, Llama...



El colectivo negro

por **Miguel Hoyuelos** 15 de julio de 2022



Cada mañana me levantaba temprano para tomar el colectivo. En la parada me encontraba, a veces, con un viejo de pelo blanco bien peinado con gomina, ojos grises y manos gruesas. Era alto y parecía conservar la fuerza ganada con décadas de trabajo que, por el aspecto de su ropa, no habían dado frutos abundantes. Nuestros esporádicos cruces en la parada se repitieron hasta un punto en que ya no podíamos considerarnos extraños.



Yuval Noah Harari

Historiador y filósofo

ARTÍCULO SOBRE LA IA (THE ECONOMIST)

“Hablamos del posible fin de la historia humana”

¿Qué pasará cuando una inteligencia no humana sea mejor que el ser humano medio para contar historias, componer melodías, dibujar imágenes y redactar leyes y escrituras?...

Las religiones han sostenido a lo largo de la historia que sus libros sagrados procedían de una fuente no humana. Eso podría ser pronto una realidad...

Acabamos de darnos de bruces con una xenointeligencia, aquí en la Tierra. No sabemos mucho sobre ella, salvo que podría destruir nuestra civilización. (...) la primera regulación que sugiero es que sea obligatorio que una inteligencia artificial revele que es una inteligencia artificial. Si mantengo una conversación con alguien y no puedo saber si es un humano o una inteligencia artificial, se acabó la democracia.

Este texto ha sido generado por un humano.

¿O quizás no?

La Vanguardia, mayo 2023.

¿Peligro existencial?

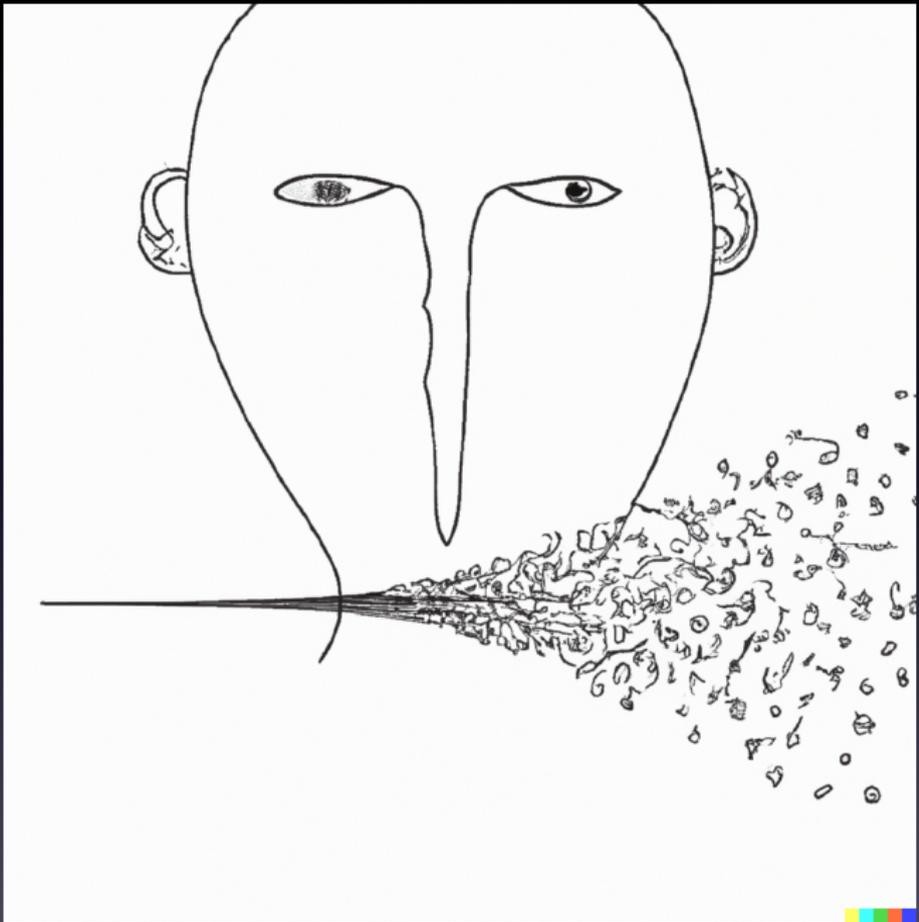
- Marzo 2023. Carta abierta de Future of Life Institute, firmada por 27500 personas (S. Wozniak, E. Musk, Y. Harari, etc.), pide pausar todo entrenamiento de IA que supere a GPT-4

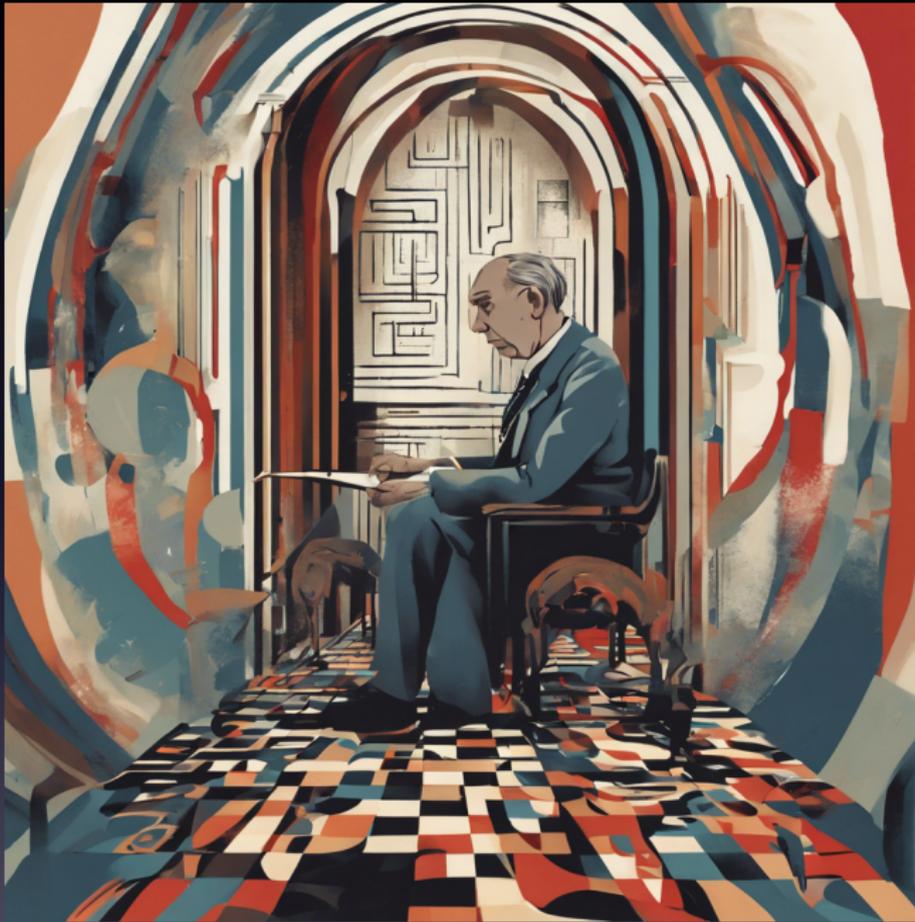
¿Peligro existencial?

- Marzo 2023. Carta abierta de Future of Life Institute, firmada por 27500 personas (S. Wozniak, E. Musk, Y. Harari, etc.), pide pausar todo entrenamiento de IA que supere a GPT-4
- Expresaron preocupación por posible peligro existencial.

¿Peligro existencial?

- Marzo 2023. Carta abierta de Future of Life Institute, firmada por 27500 personas (S. Wozniak, E. Musk, Y. Harari, etc.), pide pausar todo entrenamiento de IA que supere a GPT-4
- Expresaron preocupación por posible peligro existencial.
- Abril 2023. E. Musk planea lanzar una nueva compañía para entrenar un LLM (Large Language Model).





Dr. Victor Frankenstein



Peter Cushing en *La maldición de Frankenstein* (1957)

Conclusiones

- Conciencia: el misterio más grande y más frecuente.
- ¿Ilusión?
- Si se resuelve el misterio, se podría reproducir.
- Cambios y avances en la ciencia se reflejan en el cine y la literatura de ciencia ficción (algo que sucedió antes del boom de la IA).
- Mecánica estadística: importancia de fluctuaciones críticas.

Conciencia artificial en la ficción

- Stanislaw Lem (1961), *Memoirs of a Space Traveler: Further Reminiscences of Ijon Tichy* (se simulan agentes conscientes, o personoides, para testear la validez de la hipótesis de simulación de la realidad, o sea, de la idea de solipsismo)
- Frederik Pohl (1954), *The Tunnel under the World* (se simula una ciudad para testear campañas de publicidad, un habitante, un *simularum*, se da cuenta)
- Greg Egan, *Permutation City* (1994), *Diaspora* (1997), *Schild's Ladder* (2002)
- M. H. (2004) *Siccus*.