

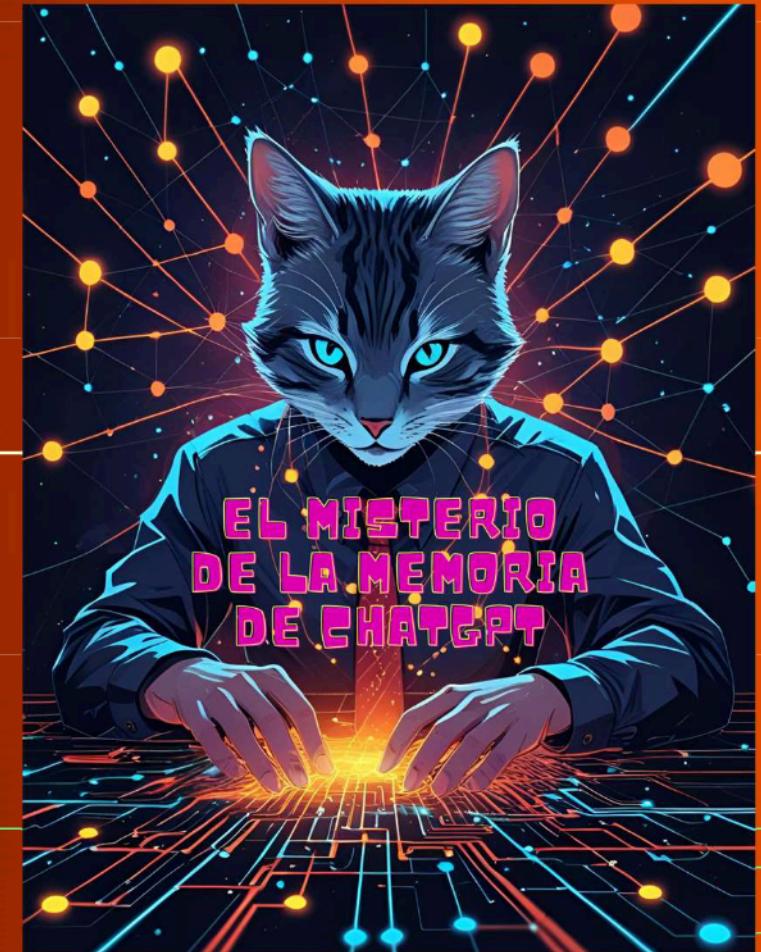


El misterio de la memoria de ChatGPT

Jesús Marco
Marta Obregón
Victor de Mora
María Peña

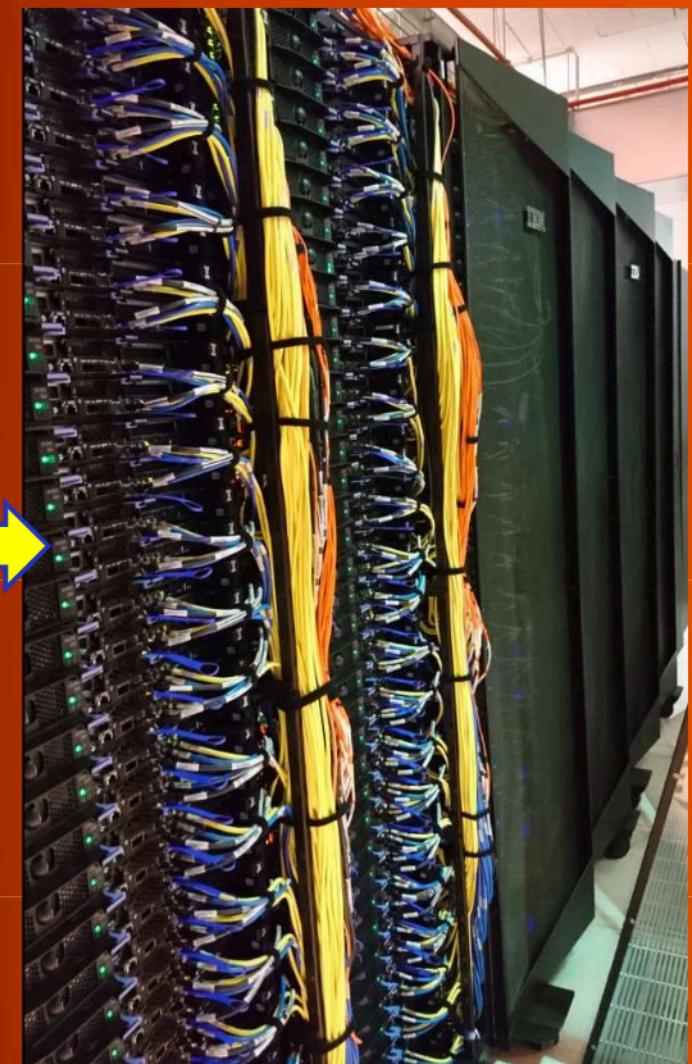
David Rodríguez
Saul Fernández
Eloy González

Advanced Computing and e-Science Group
<https://advancedcomputing.ifca.es>



Nos presentamos

- IFCA = Instituto de Física de Cantabria
CSIC + Universidad de Cantabria
- Somos investigadores que hemos estudiado Física, Informática, Matemáticas, Historia...
- Proyectos:
ENGRAMMER, AI4-EOSC, SIESTA, IMAGINE
- Este es nuestro centro y la sala de supercomputación



Advanced Computing and e-Science Group
<https://advancedcomputing.ifca.es>



¿Qué nos sorprende de chatGPT?

- ¿Quién no ha probado a usar chatGPT?
 - u otro sistema similar: Gemini, Perplexity, DeepSeek, LFM
- ¿Cómo es posible que se “comunique” tan bien con nosotros?
 - Parece “entender” lo que escribimos
 - Incluso es capaz de entender imágenes o nuestra voz
- ¿Y como es posible que parezca saber “de todo”?
 - Contestar casi cualquier pregunta, incluso buscando en internet
 - Escribe programas de ordenador
 - Crea imágenes, hace videos.
- **¿Qué es lo que más os asombra?**



Una pregunta bastante complicada:

¿Es chatGPT como un ordenador muy potente?

¿O más bien se parece a un humano?

- **COMO ES LA MEMORIA DE UN ORDENADOR**
- **COMO APRENDE CHATGPT**
- **COMO APRENDEMOS LOS HUMANOS**
- **EL MISTERIO DE LA MEMORIA DE CHATGPT**
- **COMO APROVECHAR TODO LO QUE SABE CHATGPT...
PERO SIN VOLVERNOS TONTOS**



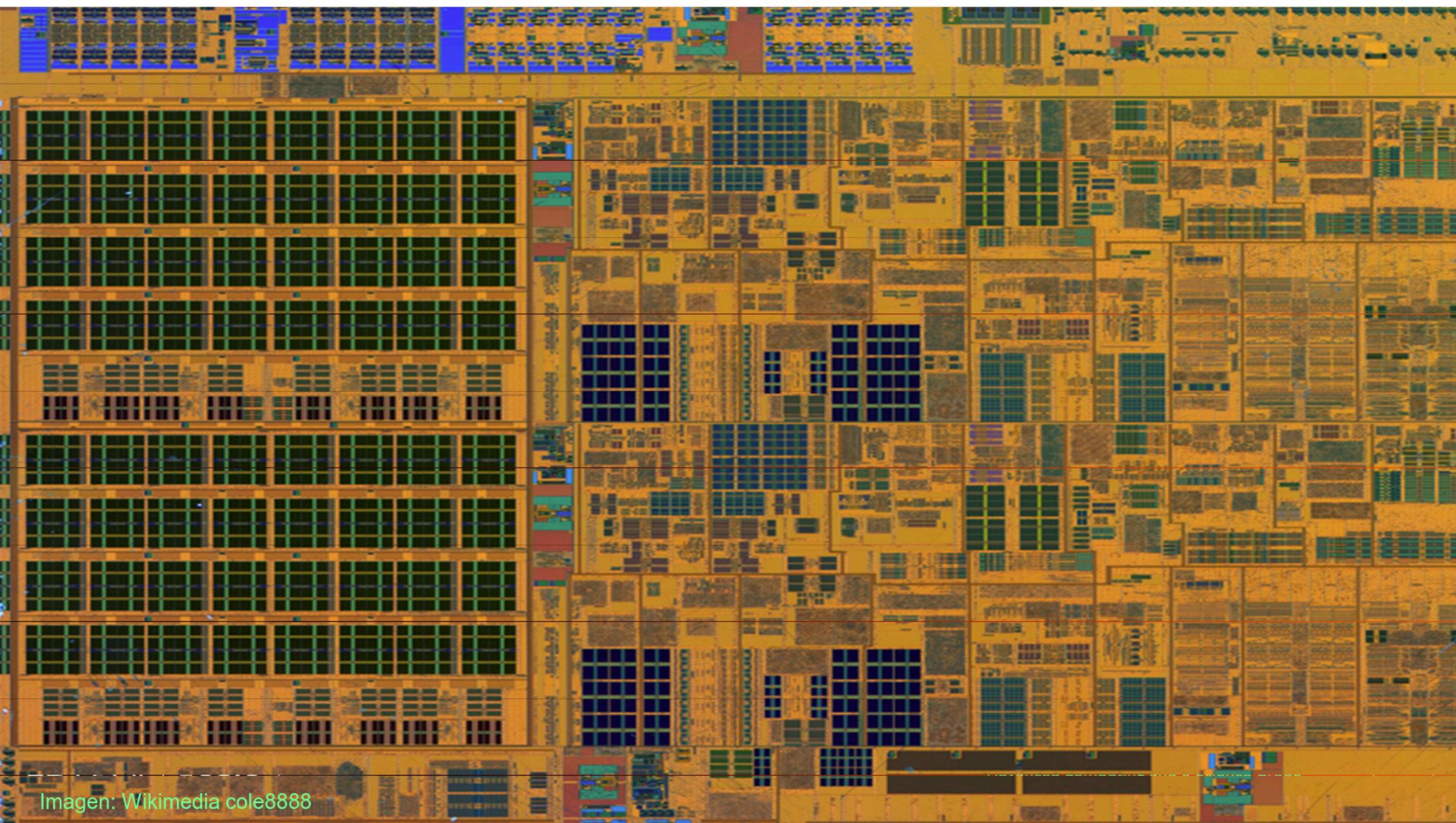


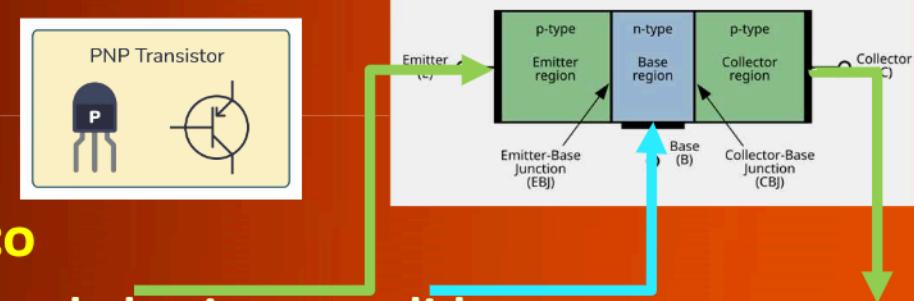
Imagen: Wikimedia cole8888

Los ordenadores solo usan bits : ¿cómo hacen?

CHIPS BASADOS EN SEMICONDUCTORES

- Integran **millones** de transistores
- Un transistor es un “interruptor” electrónico

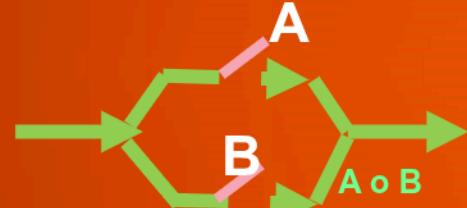
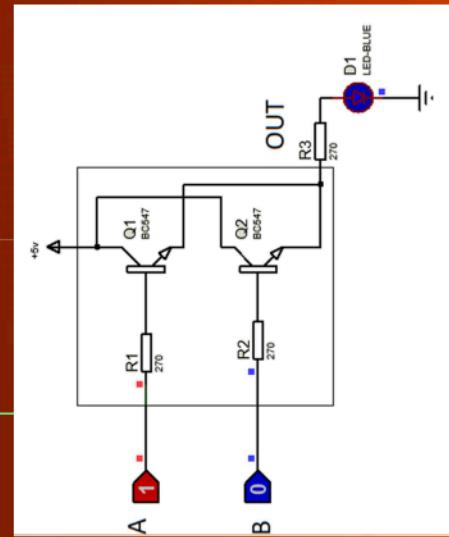
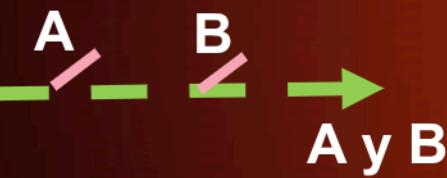
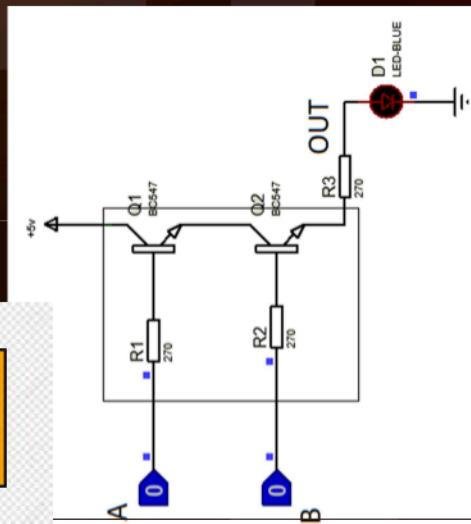
El paso de señal eléctrica desde una entrada se regula hacia una salida



Funciones

AND

OR



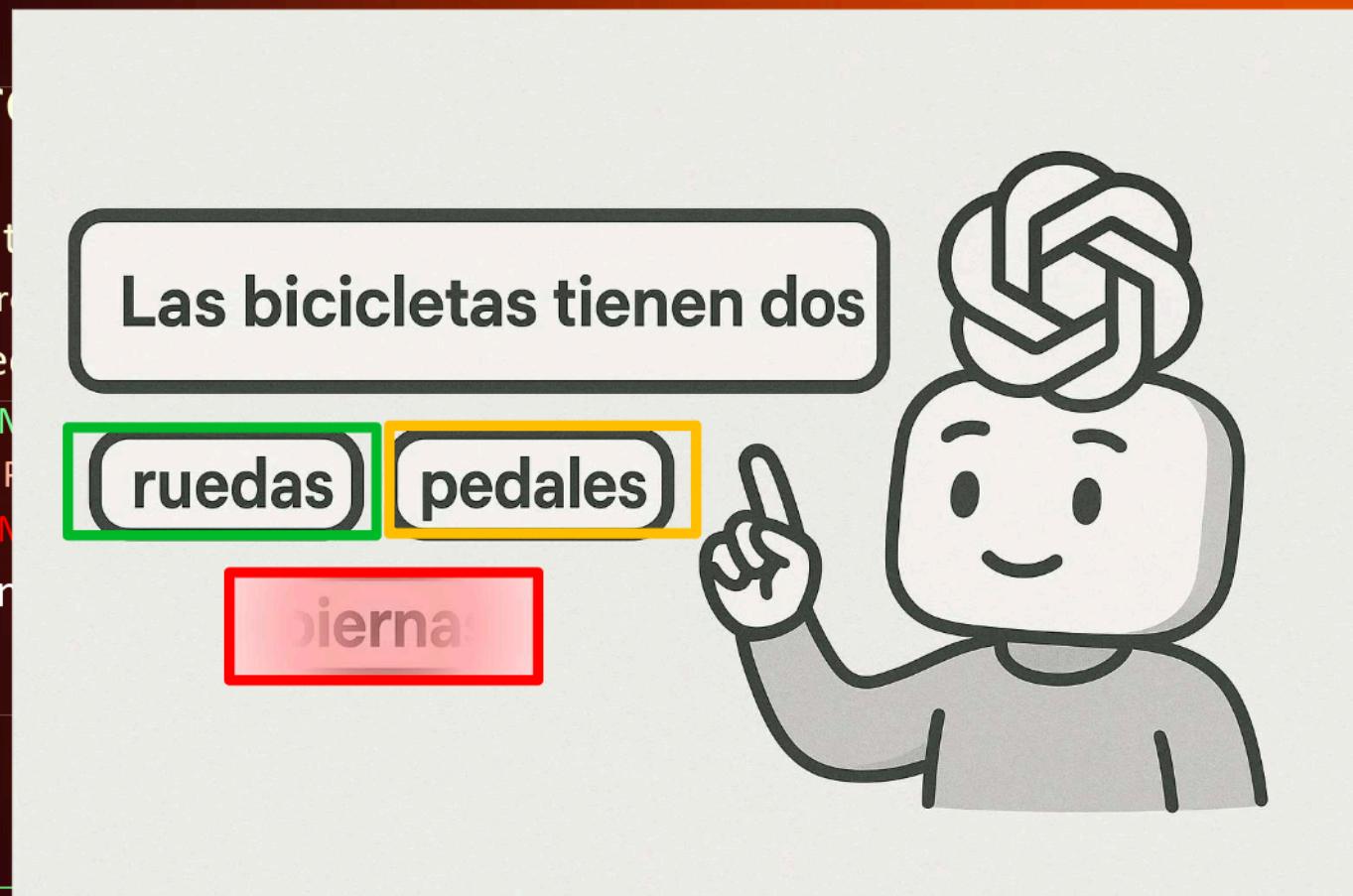
Computing and
Advanced

La memoria en los ordenadores: ¿ sólo 0 / 1 ?



Cómo “aprende” chat GPT

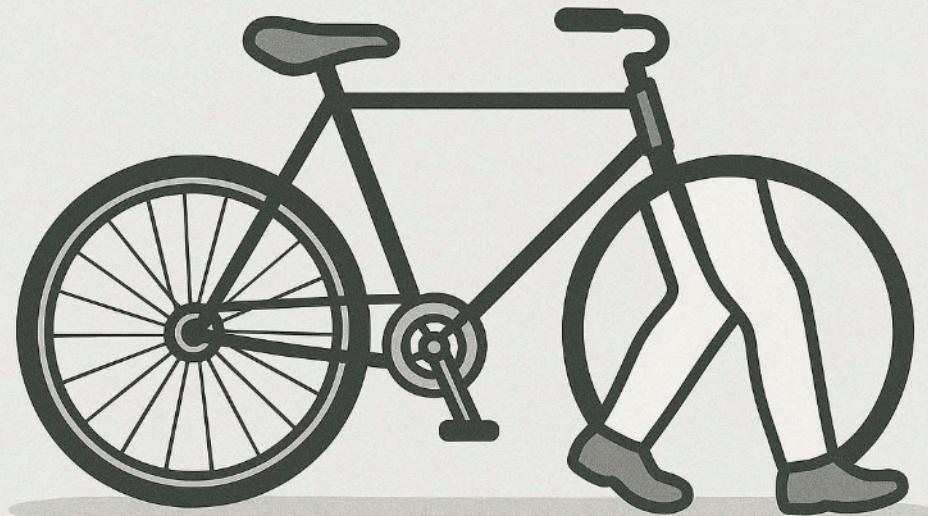
- ¡GPT se entrena a “predecir”!
 - Las bicicletas tienen dos ?????
- Para entrenar, “lee” millones de textos
 - artículos en internet, manuales, libros...
- Para cada texto, prueba a “predecir” las siguientes palabras
 - Las bicicletas tienen dos **ruedas**
 - Las bicicletas tienen dos **pedales**
 - Las bicicletas tienen dos **piernas**
- Selecciona la palabra teniendo en cuenta el contexto
- ...Y el contexto



Un chat,... chatea! de GPT a chatGPT

- GPT se ha entrenado con texto y salidas
- Ahora se le dan ejemplos de conversaciones:
P: “¿Qué es una célula?” R: “Una célula es...”
P: “Un chiste corto.” R: “¿Qué hace una abeja?”
- GPT ve miles de ejemplos así, y aprende
- Luego, **expertos humanos** revisan el resultado:
 - 👍 “Esta respuesta es buena, es clara y aclaradora.”
 - 👎 “Esta no, es confusa o poco útil.”
- GPT aprende así a dar respuestas “correctas”

Una bicicleta tiene dos piernas.



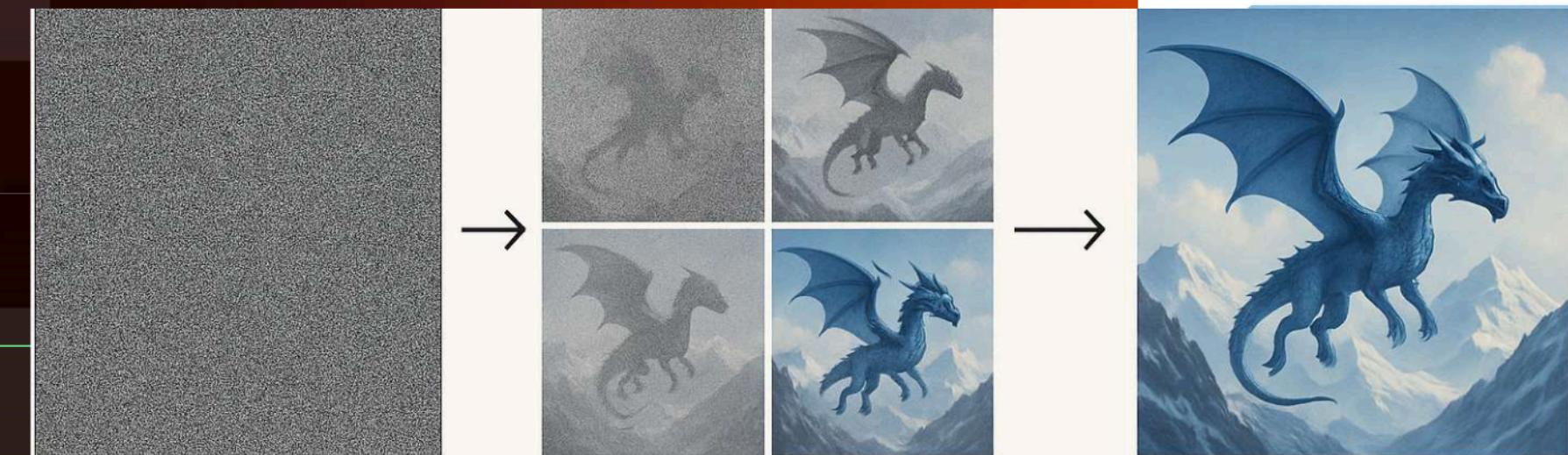
GPT = Generative Pre-trained Transformer

GPT es una IA generativa que usa “Transformers”

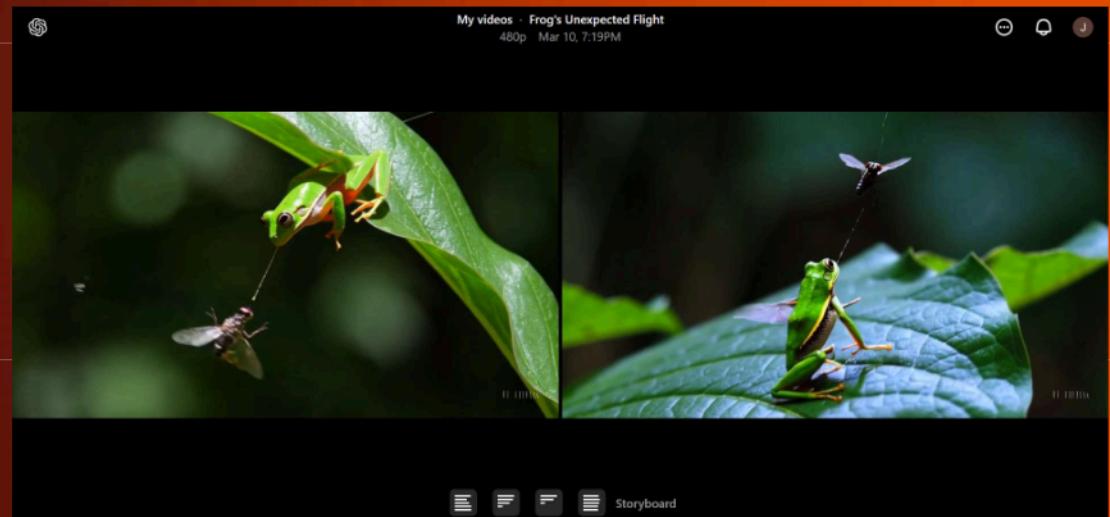
- Hemos visto que predice palabras
- Si le damos un nuevo **contexto**, puede **generar** textos nuevos (tengan más o menos sentido)
- Esos textos están representados en un espacio latente, y podemos usarlos para crear imágenes



16



Sobre ranas y moscas...



¿Como aprendemos los humanos?

- AL NACER YA SABEMOS...
 - Respirar
 - Comer
- Y LUEGO APRENDEMOS
 - Andar
 - Hablar
 - Escribir...
- NUESTRO “SISTEMA”:
 - Sistema Nervioso Central
 - Sistema Sensorial
 - Sistema Motor

es decir:

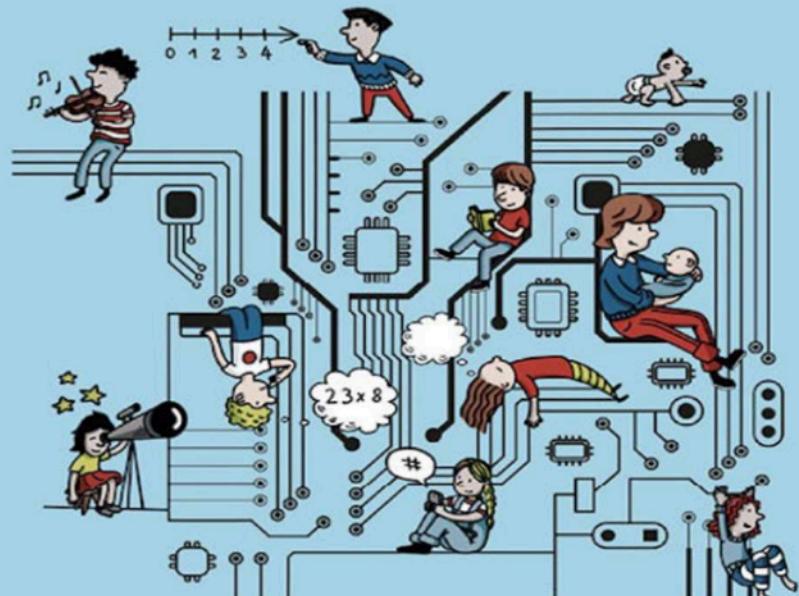
 - Cerebro (+Cerebelo, Médula Espinal)
 - Ojos-> Visión
 - Oídos-> Audición
 - Músculos->Movimiento

(incluye lengua, ojos...)

¿CÓMO APRENDEMOS?

Stanislas Dehaene

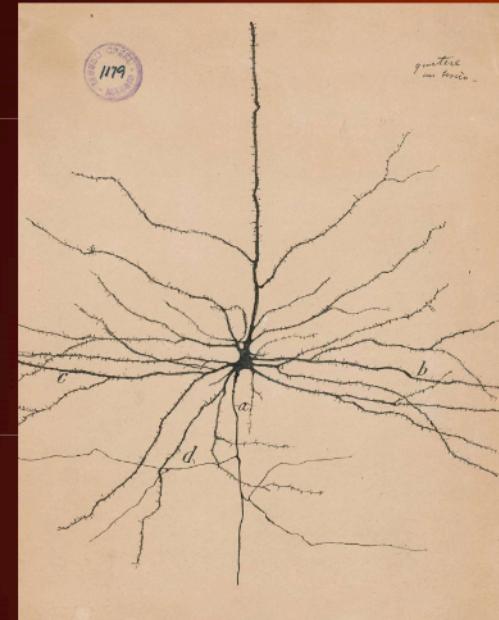
Los cuatro pilares con los que la educación puede potenciar los talentos de nuestro cerebro



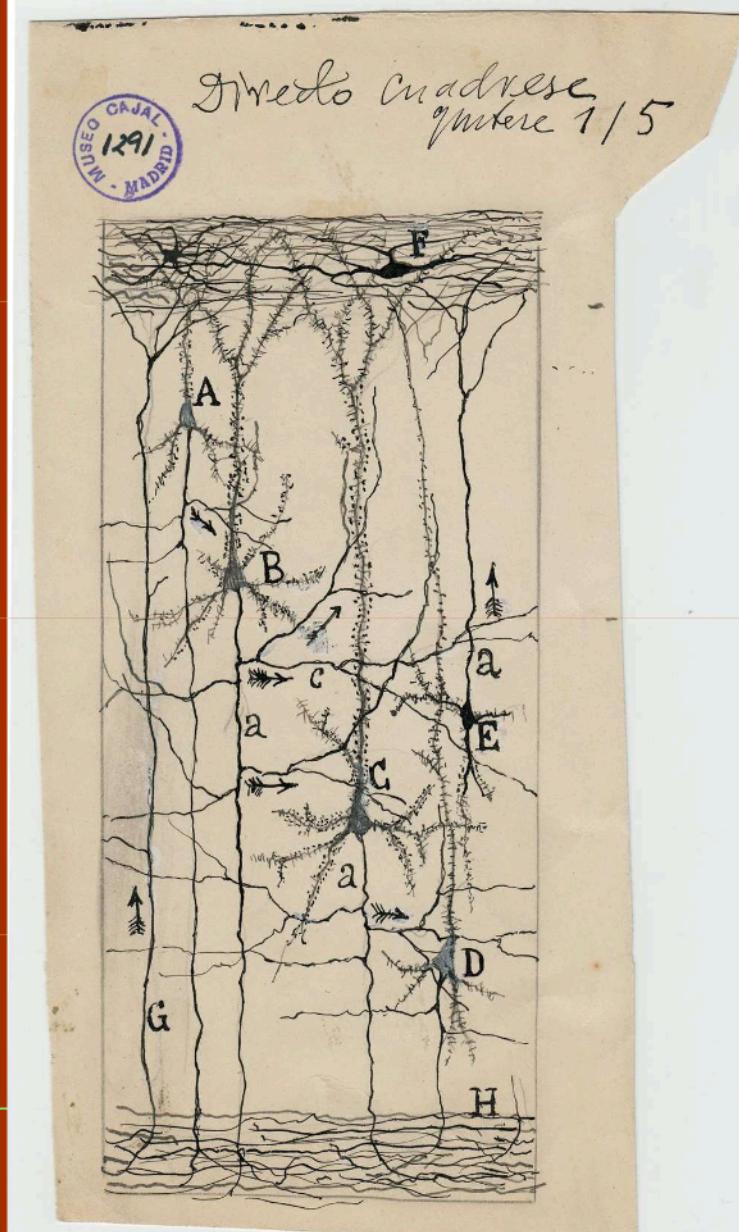
Cómo funciona el sistema nervioso



Santiago Ramón y Cajal
Premio Nobel de Medicina 1906
(Estructura del sistema nervioso)



Las neuronas son
células especiales que
se conectan formando
“circuitos” neuronales





Actividades



Venta de entradas

Visita
accesibleVen con
el colegio

Exposiciones

Localización,
tarifas y
horarios[Home](#) / [visita el mncn](#) / [Exposiciones](#) / El Legado Histórico de Santiago Ramón y Cajal

Permanentes

Temporales

Próximamente

Itinerantes

Histórico

Contenidos
multimedia

Buscador



El Legado Histórico de Santiago Ramón y Cajal

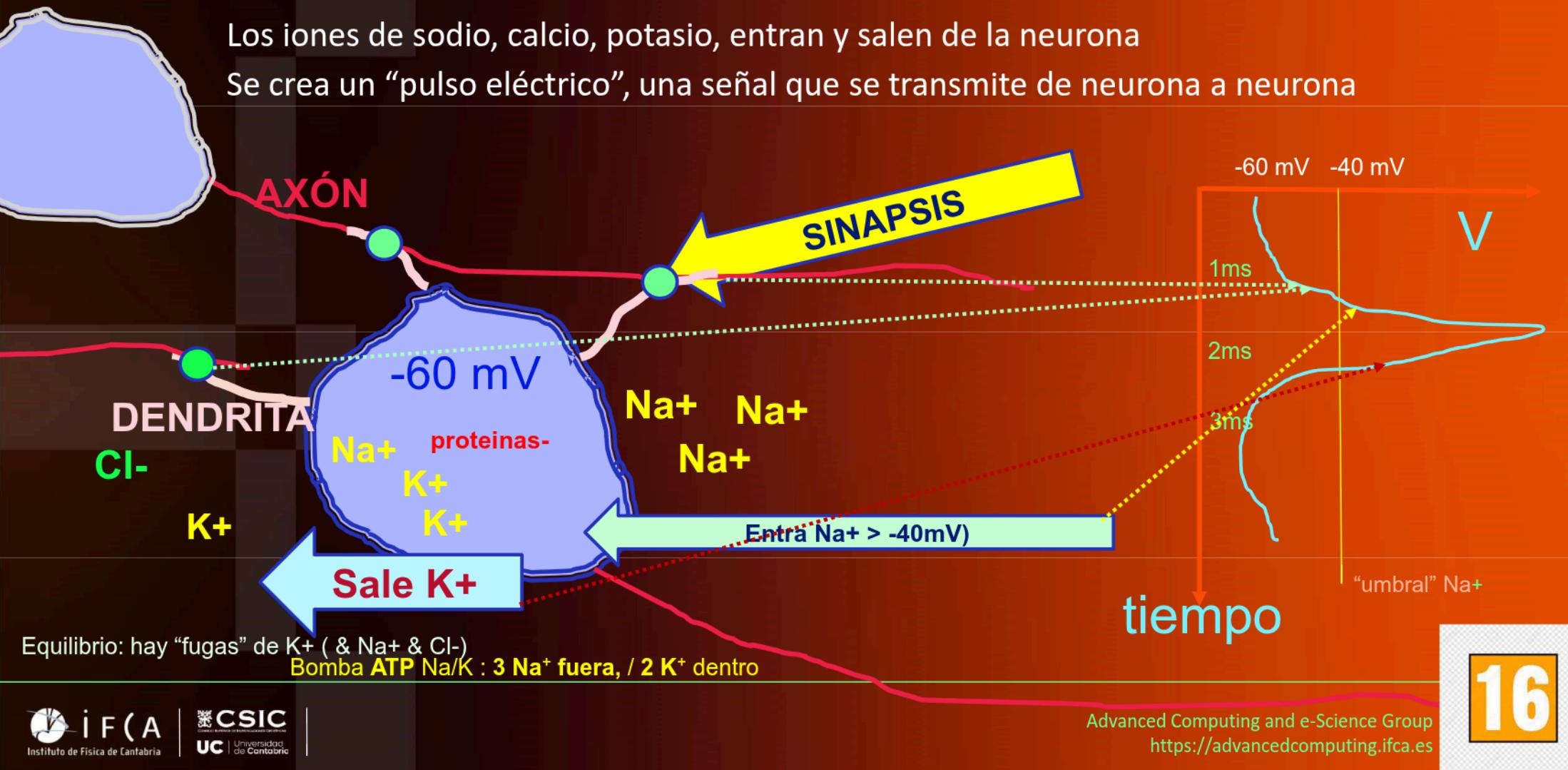
1º PLANTA

Advanced Computing and e-Science Group
<https://advancedcomputing.ifca.es>

La clave: cómo funciona una neurona

Los iones de sodio, calcio, potasio, entran y salen de la neurona

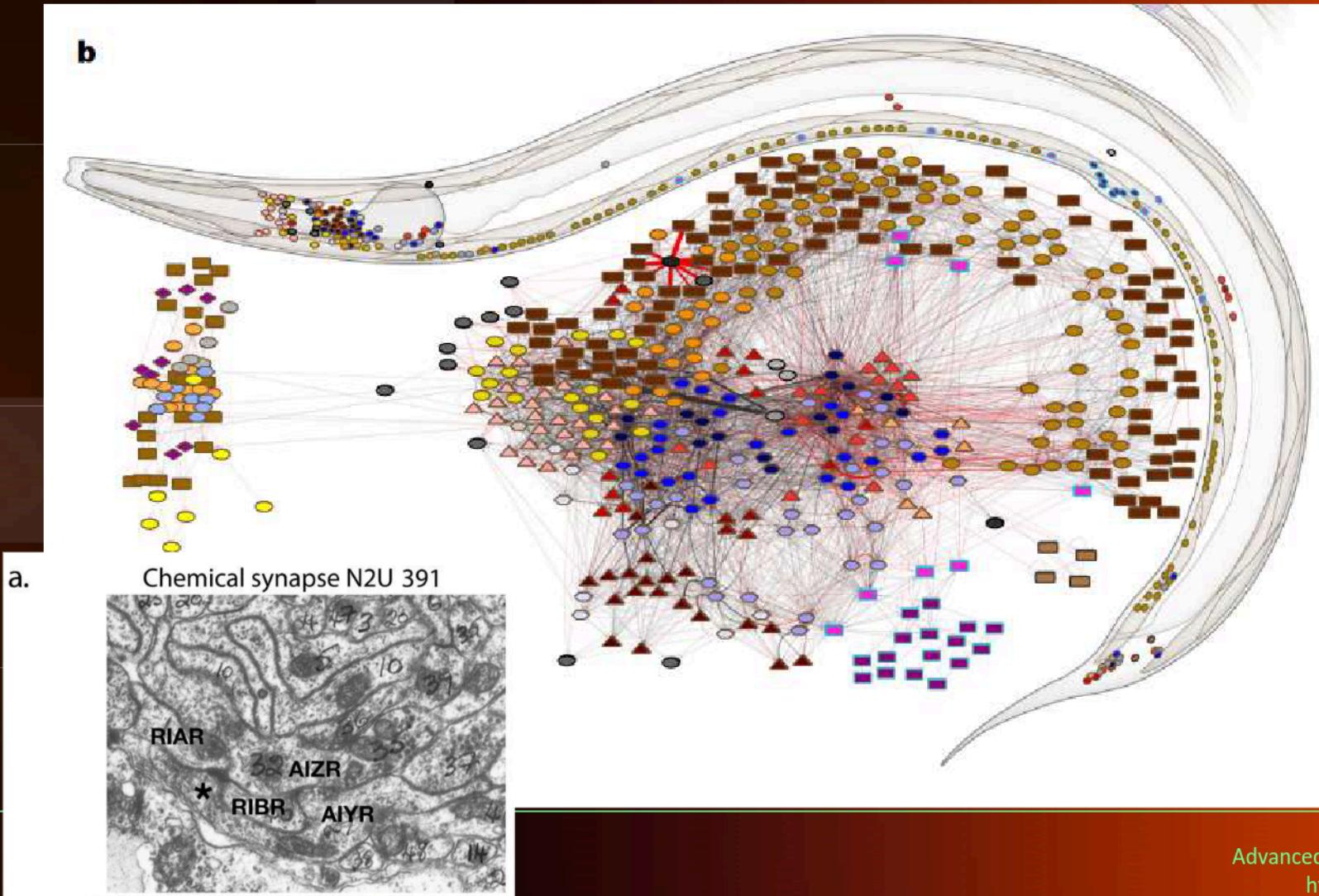
Se crea un “pulso eléctrico”, una señal que se transmite de neurona a neurona

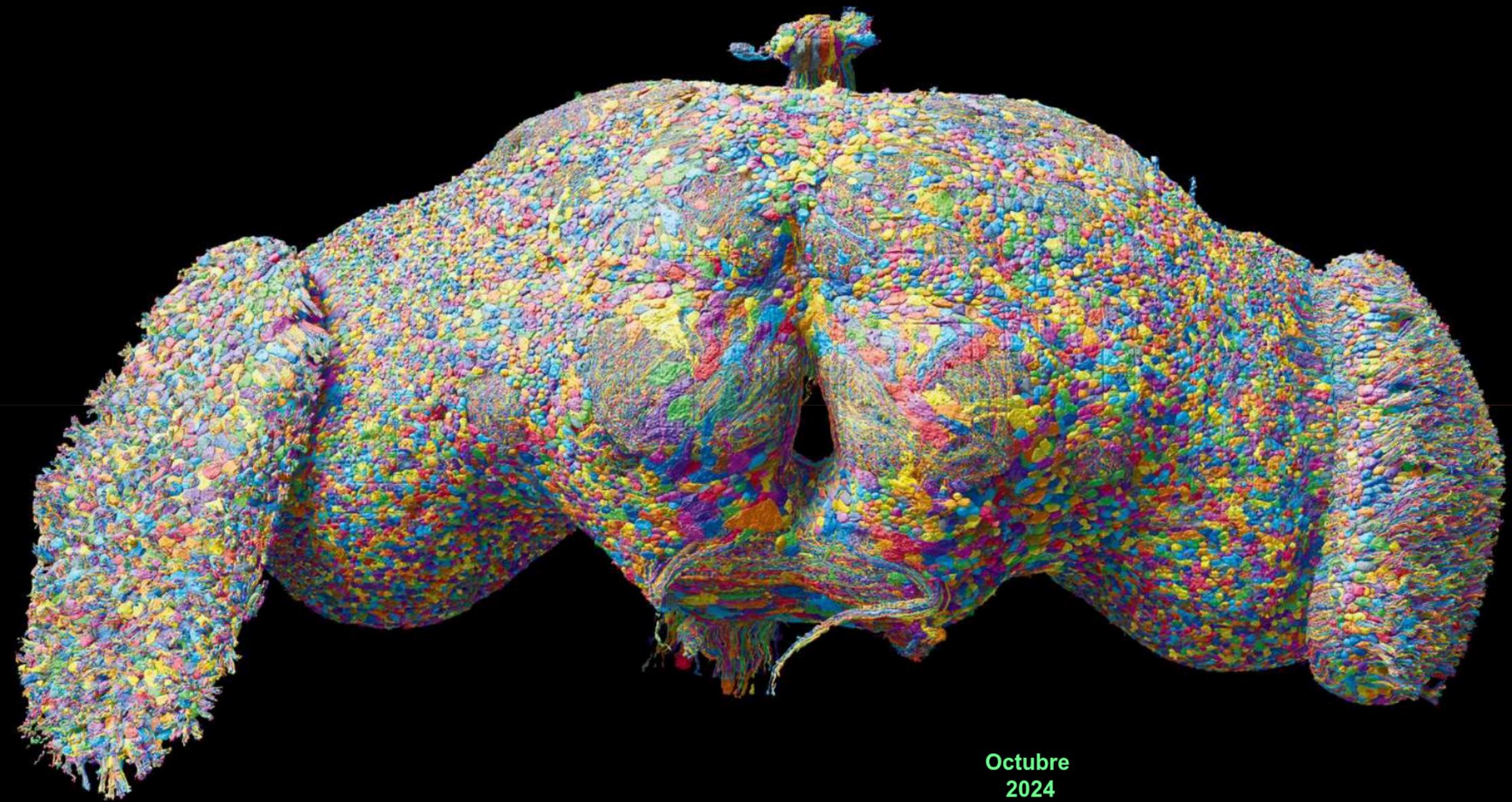


MAPA (CONECTOMA) DE LAS 302 NEURONAS DEL GUSANO C.ELEGANS

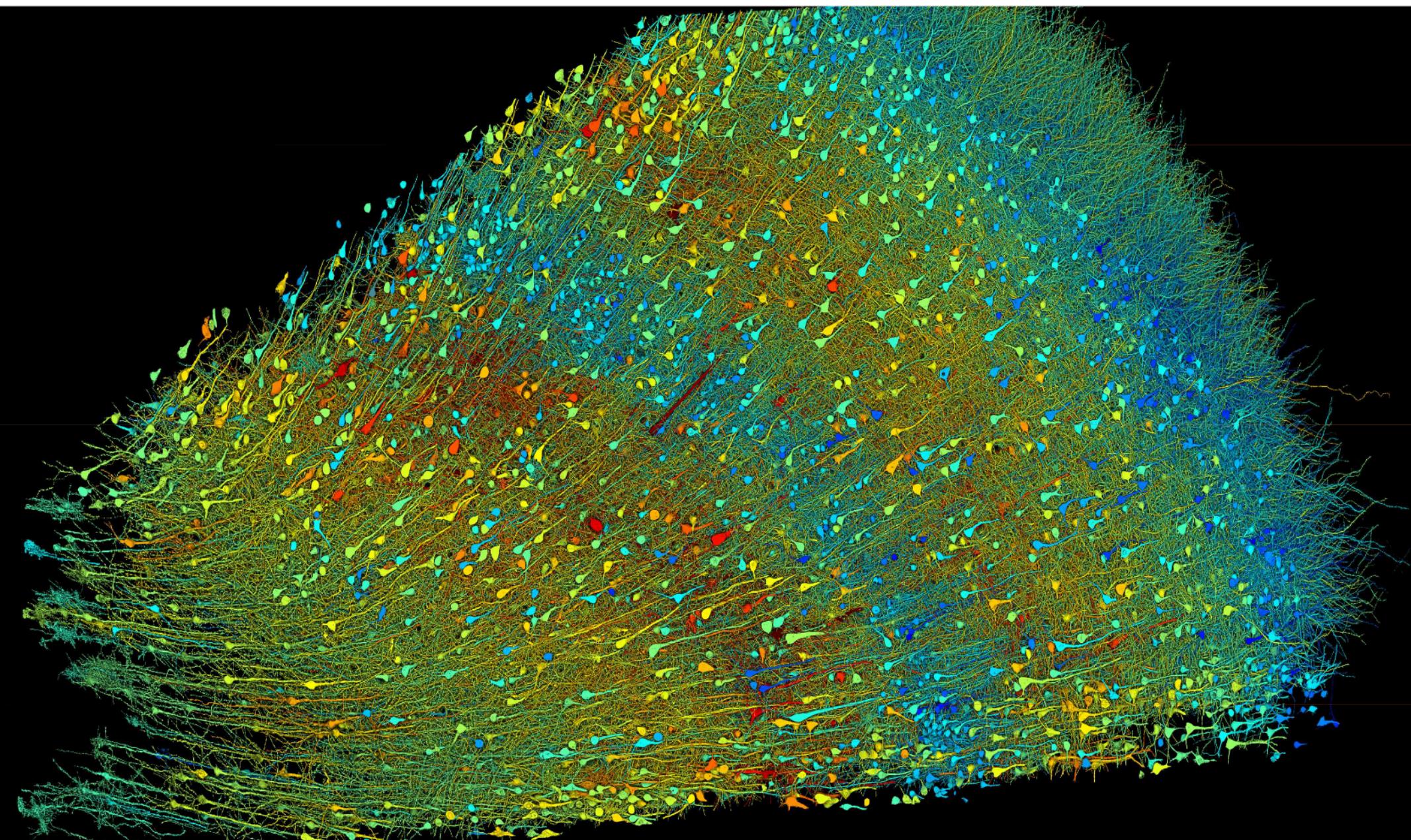
Nature 2019

Investigadores del Departamento
de Neurociencia
Albert Einstein College of
Medicine, NY, USA

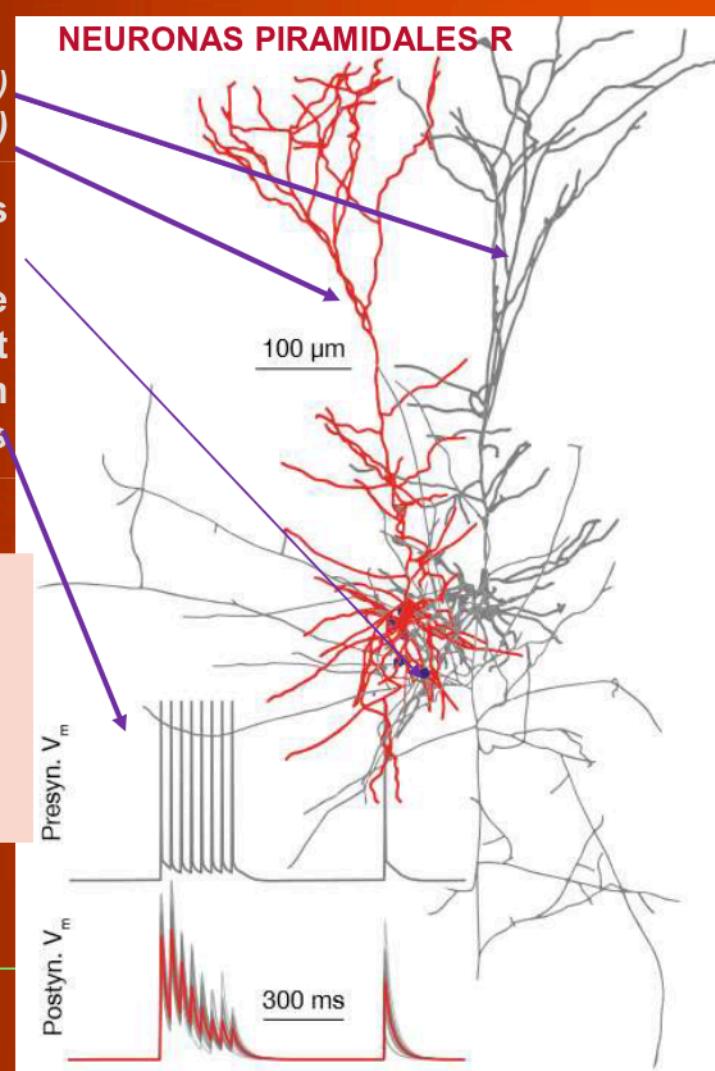
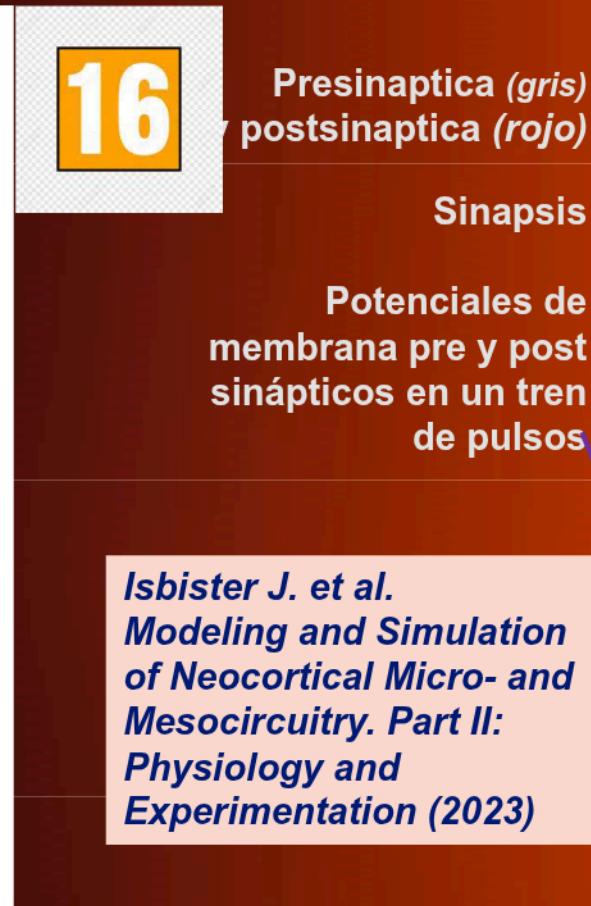
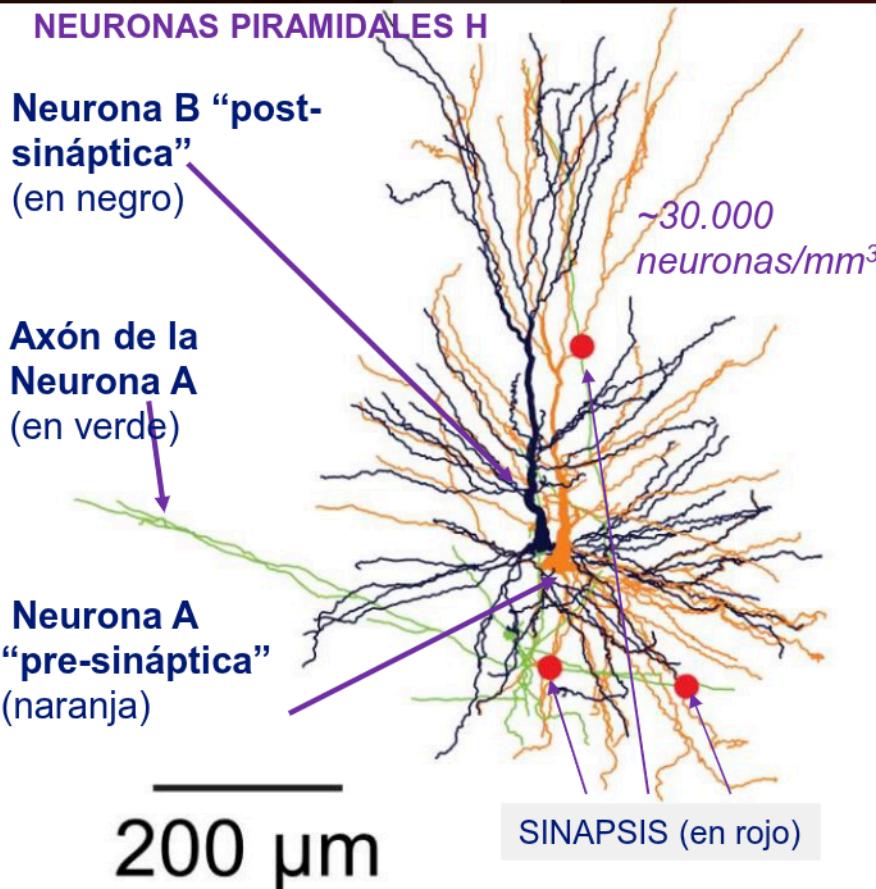




Octubre
2024



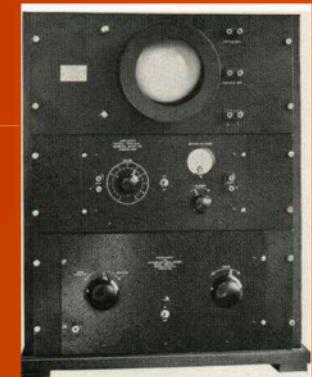
Ojo: ¿neuronas de ratón o neuronas humanas?



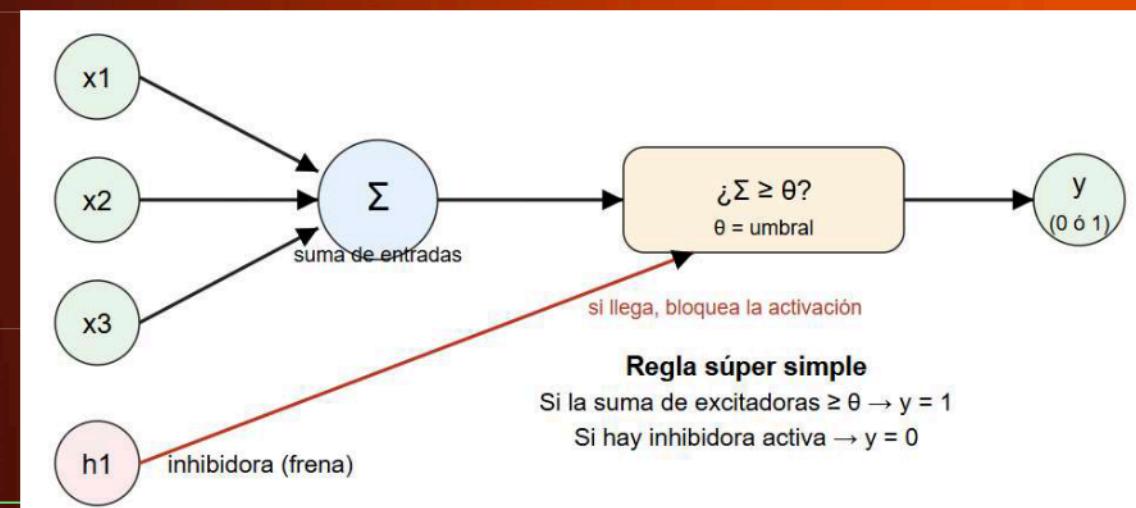
Hunt, S. et al.* “Strong and reliable synaptic communication between pyramidal neurons in adult human cerebral cortex” (2022)
*Incluyendo a Ruth Benavides, Javier de Felipe, CSIC

De la neurona biológica a la neurona “artificial”

- Rafael Lorente de Nò (discípulo de Cajal) empleando el osciloscopio (1935), observa cómo el pulso eléctrico que transmite el axón depende de los pulsos que llegan de otras neuronas

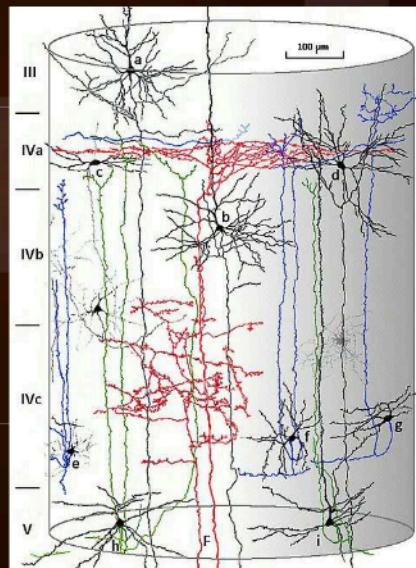
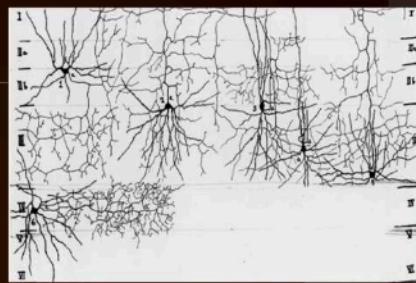


- Siguiendo esta idea, McCulloch y Pitt proponen la “neurona artificial” juna función matemática!



Las redes neuronales de nuestro cerebro

16



Lorente de Nò también observó que la zona visual de la corteza cerebral tiene medio millón de columnas corticales, cada una ~ 50.000 neuronas.

Propuso que la memoria y el aprendizaje en el cerebro se basan en la **“reverberación” en los circuitos neuronales**

Donald Hebb tomó nota de esta idea y propuso que las redes neuronales aprenden reforzando conexiones cuando se activan a la vez (**“fire together/wire together”**)

Estas ideas son clave en modelos como el de la memoria **“asociativa”** de Hopfield (Nobel 2024)

Aún seguimos investigando como aprendemos

- Las neuronas son células “para toda la vida”
 - Tenemos aproximadamente entre 50 y 100 millones de neuronas
 - En adultos se generan muy pocas neuronas nuevas (“neurogénesis”)
- Un cerebro “aprende” cambiando sus conexiones (**SINAPSIS**)
 - Lo hemos estudiado en animales más sencillos: gusanos, moscas, ratones
 - Cada neurona se conecta con otras muchas
 - En humanos: entre 10.000 y 100.000 **miles de millones** de conexiones
- La memoria está distribuida por todo el cerebro en estas conexiones
 - **Que se crean, refuerzan, debilitan y eliminan todo el tiempo**
 - **A partir del aprendizaje, que dura toda la vida**
- ENTONCES...

¿Como es posible que tengamos memoria?



CUIDEMOS NUESTRAS NEURONAS... Y SUS CONEXIONES

Los humanos somos “privilegiados”

- Durante 20 años vamos formando las mejores conexiones, sobre todo...
- ...en nuestra corteza prefrontal: la que nos ayuda a tomar decisiones
- **¡Por eso es crítico cuidar nuestro cerebro!**

¿Qué afecta a las conexiones en nuestro cerebro?

- ESTUDIAR, LEER, HACER DEPORTE, DIBUJAR, CANTAR, BAILAR
- NO DORMIR LO SUFICIENTE
- FUMAR TABACO / VAPEAR
- BEBER ALCOHOL
- CANNABIS / MARIHUANA
- COLGARSE DE LOS VIDEOJUEGOS ...O DE CHATGPT

ADICIONES



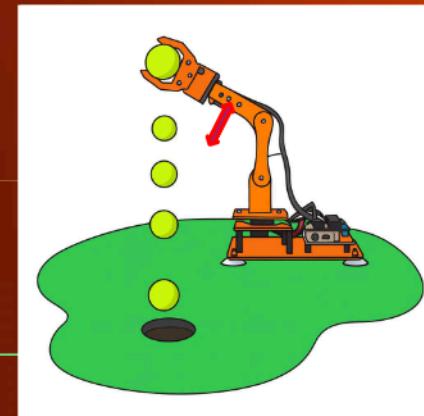
Nos vamos acercando a... la memoria de chatGPT

- La memoria de chaGPT usa redes neuronales computacionales muy muy muy muy muy... complejas
- En lugar de memorizar una palabra o una imagen mediante sus bits, aprendemos a reconocerla
- Para ello tenemos que “entrenar” la red, viendo palabras o imágenes: vamos ajustando las conexiones de la red hasta lograr lo que queremos



Entrenamiento = reducir el error hasta acertar

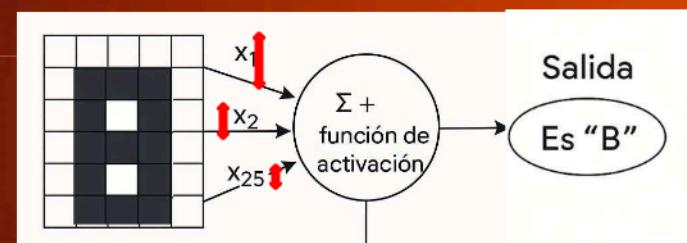
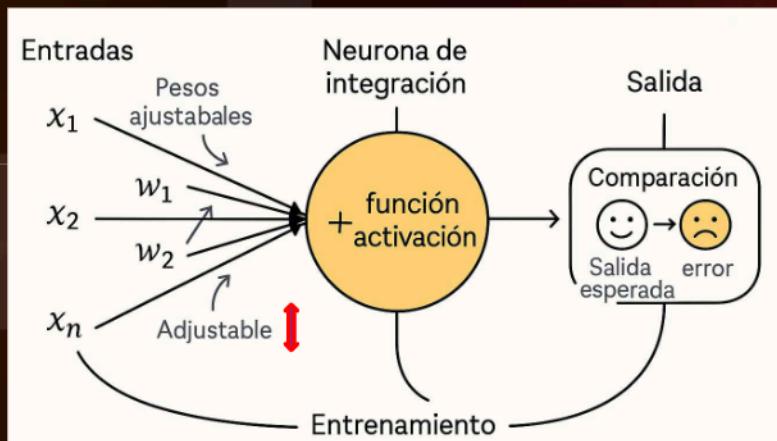
- Imagina un brazo robótico que mete bolas en un agujero
 - O una escopeta de feria con la que queremos acertar
 - O los brazos de un jugador de baloncesto tratando de encestar
- El brazo se puede ajustar en posición, altura y dirección
 - Son como las conexiones de la red neuronal, que se ajustan
- Lanza una bola, si entra en el agujero, ¡éxito!
- Si no entra, vemos que corregir teniendo en cuenta el error
 - Corregimos
- Volvemos a lanzar...



La red neuronal de IA más simple: perceptrón

Un ejemplo (Rosenblatt 1957, ordenador Mark I):

25 “pixels” en blanco y negro que representan letras...



¡Pero estas redes “sencillas” no son suficientemente potentes!



Transformers (2017: las redes neuronales que usa chatGPT)



El texto de entrada se convierte en vectores numéricos

Se suma una codificación posicional

Red neuronal multicapa, con miles de millones de conexiones

Cada palabra del texto se compara con todas las demás, y se calcula a cuáles se debe "prestar más atención".

A cada paso, el significado de cada palabra se refina según el contexto.

El modelo genera una salida basada en lo aprendido tras ser entrenado con millones de textos diferentes

Y seguimos con el misterio...

- Si chatGPT es una red neuronal, y guarda la información en sus conexiones (“pesos”), ¿cuantas conexiones usa?
- **PUES NO LO SABEMOS**
 - GPT 3 tenia 175 miles de millones (175 B)
 - GPT 4/5 podría tener 1.800 miles de millones (1.800 B)
- Pero si que sabemos que DeepSeek (chino) usa 675 B
 - Pero suele usar “solamente” 40 B

¿RECUERDAS CUANTAS CONEXIONES TENIAN NUESTRAS NEURONAS?

¿Y como es posible que chatGPT con esas conexiones sepa absolutamente de casi todo?
¿Es tanta información (en bits)? (pista: ~4TB)



El coste de chatGPT

- **Para entrenar chatGPT se han empleado:**

- Superchips (GPUs) (nvidia) (precio de cada chip: como un coche!)
- Gran cantidad de texto, imágenes, etc (internet)
- Que se ha usado (copiado) sin preguntar en general
- Grandes centros de computación (coste de cientos de millones de \$)
- Que han consumido (y consumen) gran cantidad de energía
- Grandes compañías (OpenAI, Google, Meta,...) o países (China,...)

- **Para usar chatGPT necesitamos**

- **Conectarnos por internet a esos centros...**
- **y darles nuestra información**

¿Y si “otro chatGPT” funcionara sin salir de nuestra tablet?



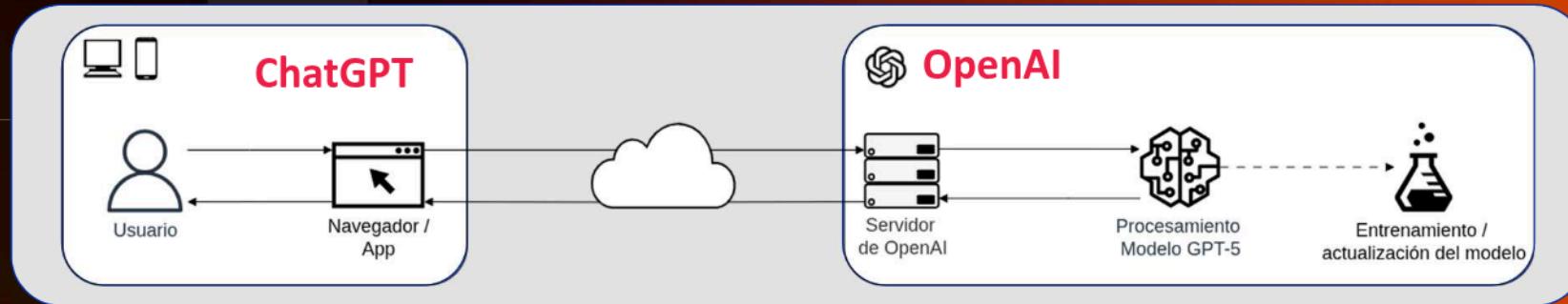
Un “hermano pequeño” de chatGPT: LFM

- Tiene muchas menos conexiones (1.2 - 2.6 B)
- Ha “entrenado” con muchos menos ejemplos
- Probemos a dialogar con LFM2 1.2B **y** con LFM2 2.6B
 - A veces se equivoca
 - O te contesta en inglés (está “entrenado” en inglés y luego un poco en español, en japonés...)
- Usar un modelo así tiene una ventaja
 - Puede ejecutarse en nuestro propio móvil o Tablet
 - **¡No le damos nuestra información a nadie!**

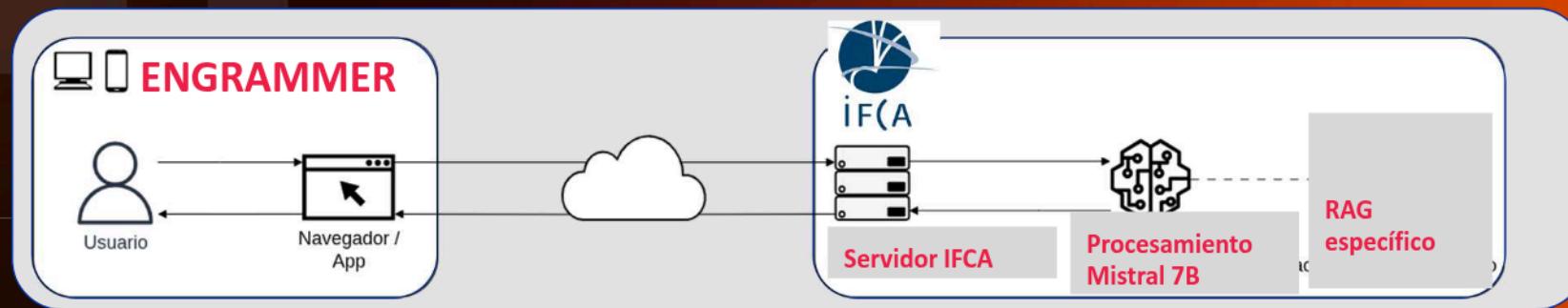


COMO ABORDAR LA PRIVACIDAD

- Todos los LLM están “pre-entrenados”, y ofrecen un servicio de “inferencia” (consultas)



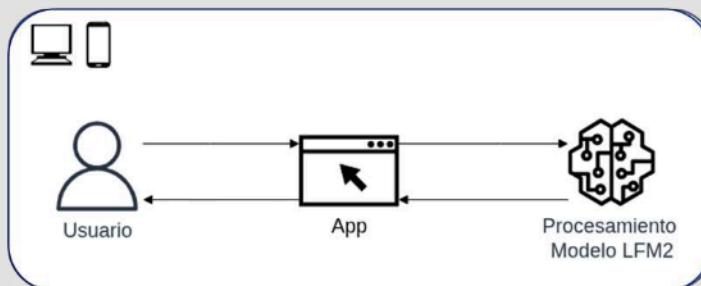
- Si son modelos “en abierto”, podemos realizar la inferencia en sistemas locales, y controlar el RAG que necesitemos, adaptado a cada “asignatura”



LA REVOLUCIÓN DE LA COMPUTACIÓN PERIFÉRICA (EDGE COMPUTING)

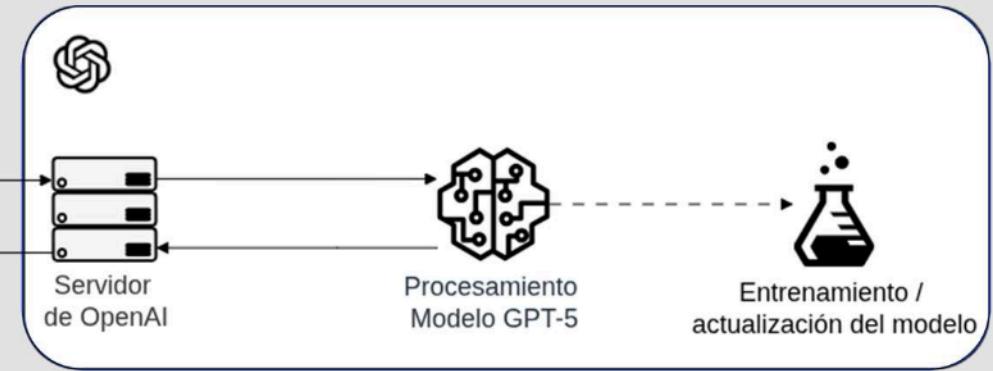
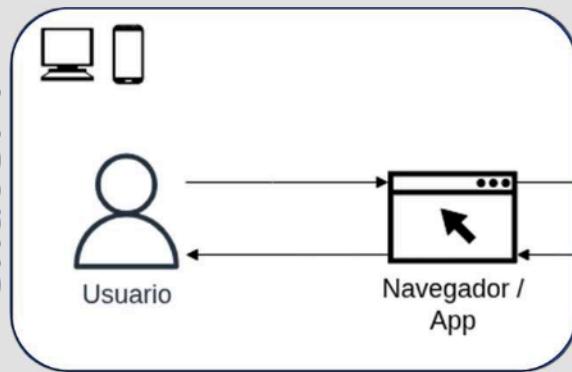
- ¿Y si fuéramos capaces de ejecutar la inferencia de un modelo (en abierto) en nuestro propio PC/móvil/tablet?
- MODELOS OPTIMIZADOS para CPU/NPU de dispositivos móviles

LFM

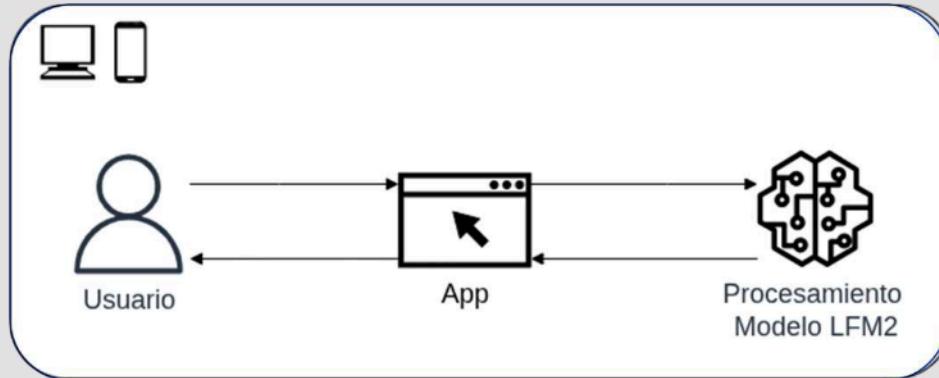


Un “hermano pequeño” de chatGPT: LFM

ChatGPT



LF
M



Repasemos: ¿hemos aprendido algo sobre chatGPT?

- Es muy potente, usa un modelo con muchas conexiones 
- Sus respuestas siempre son correctas 
- Lo que conversamos con ChatGPT es confidencial 
- Podemos fiarnos de ChatGPT 
- ChatGPT no piensa por nosotros 
- Gracias a ChatGPT puedo hacer cualquier tarea 

¿Y si ChatGPT lo hace todo, que le pasa a tu cerebro?



Lo sospechábamos, pero ahora lo sabemos

- En el MIT analizaron varios grupos de estudiantes
 - Unos usaban ChatGPT para todo
 - Otros Google
 - Otros solo su cabeza
- Los que usaban ChatGPT hicieron peores trabajos, pero eso no fue todo...

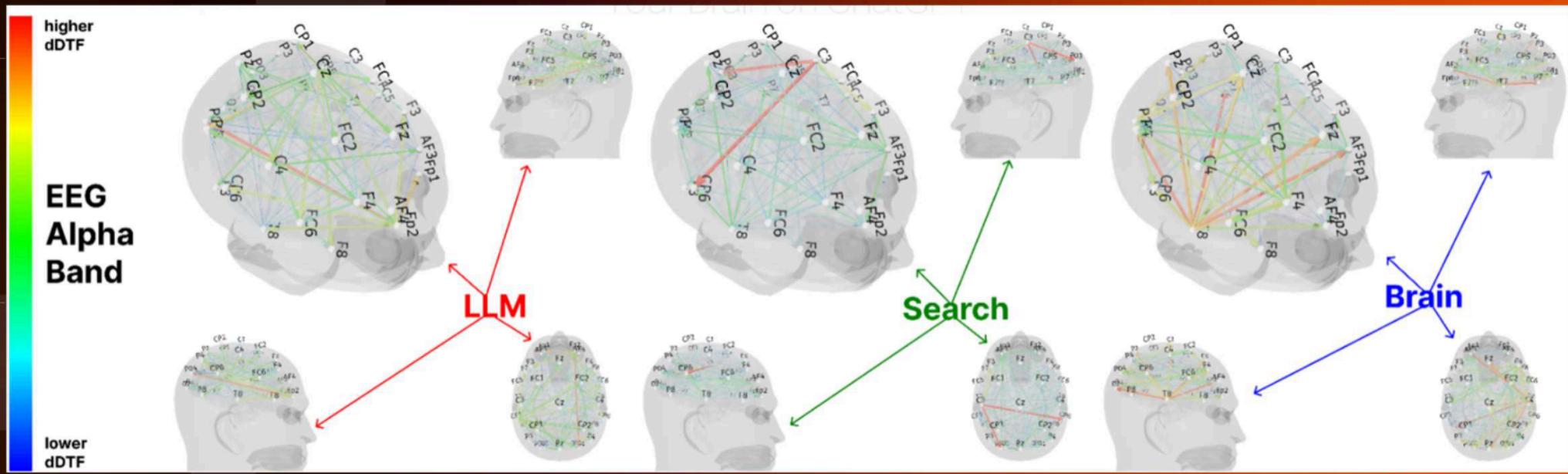
What Happens to the Brain on ChatGPT?

Researchers at MIT invited fifty-four participants to write essays across four sessions, divided into three groups: one using ChatGPT, one using Google, and one using only their brainpower. In the final session, the groups switched roles. What these researchers found should make all of us pause.

Participants who used ChatGPT consistently produced essays that scored lower in originality and depth than those who used search or wrote unaided. More strikingly, brain imaging revealed a decline in cognitive engagement in ChatGPT users. Brain regions associated with attention, memory, and higher-order reasoning were noticeably less active.



Lo sospechábamos, pero ahora lo sabemos



En sus cerebros, las regiones de atención, memoria y razonamiento estaban menos activas: chatGPT nos puede “atontar”

Entonces, ¿podemos utilizar chatGPT?

- **La clave: ENTENDER Y CONTRASTAR lo que nos dice**
- Solamente si podemos comprobar la respuesta nosotros mismos podemos fiarnos de chatGPT
 - Ejemplo: arreglar una puerta de un garaje
- Más difícil: que “referencias” ha usado chatGPT para entrenarse
 - ¿Fiables?
 - ¿Relevantes? Únicas?
 - **¿Tienen sesgos? ¿Hay instrucciones ocultas?**



chatGPT (y todos los LLM) “alucinan”

- Alucinaciones de verdad: “probabilísticas”
- “Alucinaciones” que no son alucinaciones:
 - Tendencias (políticas, ...)
 - Servilismo
 - Interés comercial
 - Ahorro (si no pagas)



Vale, usemos chatGPT como un asistente de nuestra memoria: proyecto ENGRAMMER

- Aumentamos la memoria de chatGPT (u otra IA) con lo que queremos aprender (lo que está en nuestros libros de texto)
- También grabamos lo que vamos aprendiendo
- Usamos este chatGPT-asistente personal para aprender mejor:
 - Aprendizaje espaciado
 - Método “Feynman”: explicar lo aprendido de forma “más sencilla”
 - Palabras sencillas
 - Esquemas gráficos



Financiado por la Unión Europea
NextGenerationEU



Plan de Recuperación,
Transformación
y Resiliencia

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Test práctico

- Vamos a conectarnos a una app desde la Tablet:
 - **Desde el navegador:** engrammer.eu
 - **Desde la app APOLLO:** LFM
- Luego explicamos de forma sencilla, lo que habéis aprendido:
 - Primero avisamos a chatGPT (o a LFM)

Soy una estudiante de XX años, y te voy a explicar lo que recuerdo de la charla de hoy para que me hagas preguntas y me des pistas, pero no me digas la solución, quiero aprenderlo yo!

Hemos visto que la memoria de chatGPT no es como la memoria usual de un ordenador o un móvil que por ejemplo guarda las fotos usando pixels/bits que se graban en chips con transistores. chatGPT usa unas funciones matemáticas que se llaman redes neuronales porque funcionan de forma parecida a las neuronas de nuestro cerebro: recuerdan las conexiones que usan para entrenarse y en estas conexiones queda grabada la memoria.

- La app os irá haciendo preguntas o sugerencias
- ¡Verás como recuerdas mejor este tema!

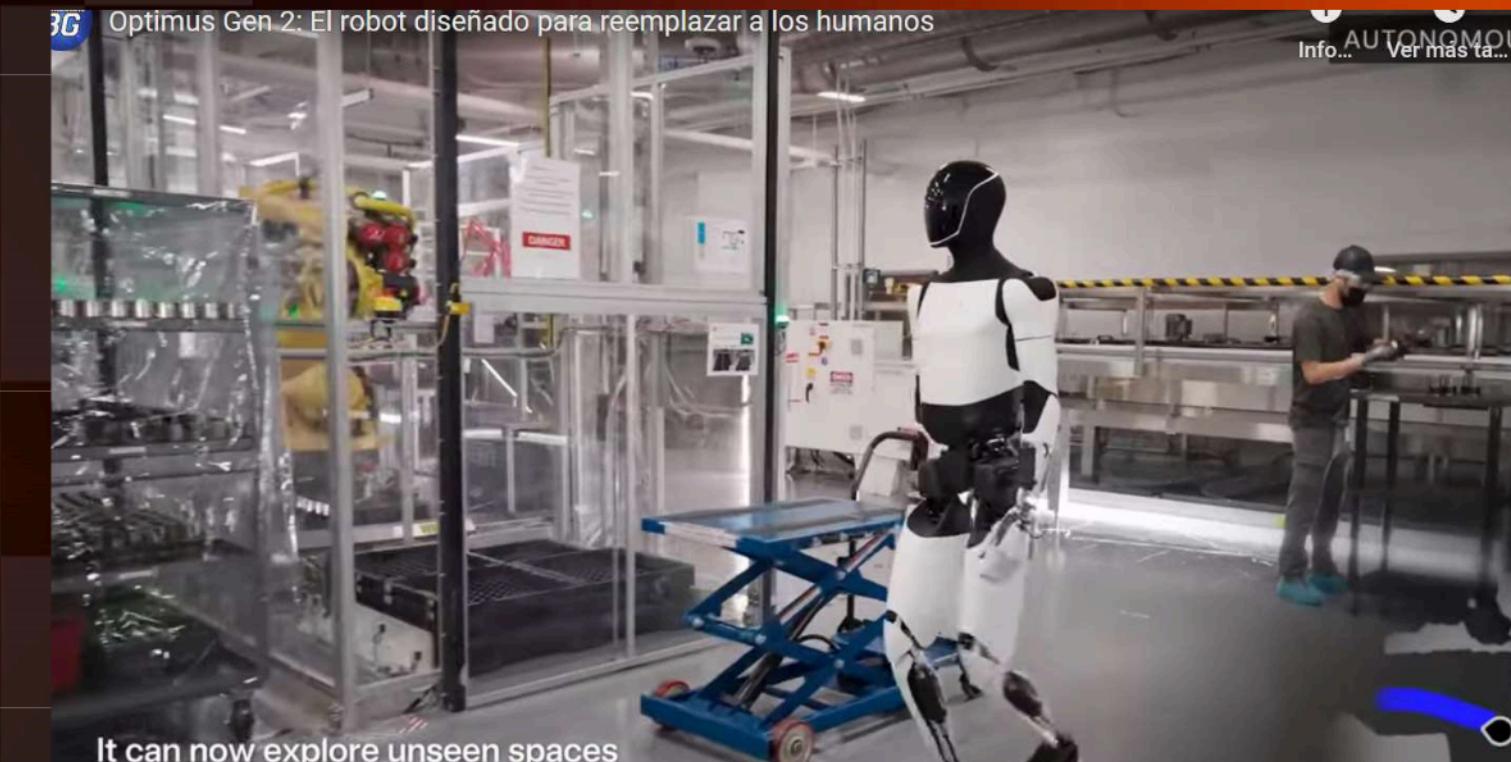


EPÍLOGO

¿QUE DEBEMOS HACER?
¿QUE PODEMOS HACER?



¡No es que venga... es que ya casi está aquí!



Robot Optimus 2 de TESLA



Hágase justicia

Vivo lo mejor que puedo en este país infeliz, rico en su gente y su juventud, temporalmente pobre en sus élites, en busca de un nuevo orden y un renacimiento en el que creo...

Sin verdadera libertad y sin cierto honor, no puedo vivir.

Siempre he pensado que si quien confiaba en la condición humana era un necio, quien se desesperaba ante los acontecimientos era un cobarde.

Albert Camus (filósofo francés, 1913-1960)

Realmente, ¿qué podemos hacer?

18

- **Queramos o no, la ciencia y la tecnología no se van a detener**
 - La carrera nuclear en los años 60 fue un ejemplo
 - Es posible que la IA logre sistemas competitivos con el cerebro humano
(Los expertos disienten ampliamente ahora mismo al respecto)
 - Si se logra, podrían aparecer sistemas robóticos o directamente humanoides
- **NO ES SENCILLO REGULAR LOS PRINCIPIOS ÉTICOS A NIVEL MUNDIAL**
 - Las leyes de protección de la información personal en Europa son un ejemplo positivo
 - Se ha logrado parcialmente regular la manipulación genética
 - Se ha controlado la proliferación nuclear
 - Se trabaja en mejorar la ética en Inteligencia Artificial
- **A NIVEL PERSONAL**

Debemos buscar referencias personales y sociales, y para ello necesitamos formarnos

LA EDUCACIÓN ES LA CLAVE DE NUESTRA SOCIEDAD

Y queramos o no, debemos implicarnos.

Cada opinión, cada gesto, cada acción, cuenta.

Si no empezamos hoy, quizás nos demos cuenta demasiado tarde.

**El futuro está
en vuestras manos**

GRACIAS POR VUESTRA ATENCIÓN

