¿Es realmente inteligente la Inteligencia Artificial?

Ramon López de Mántaras Badia

Instituto de Investigación en Inteligencia Artificial (IIIA) – CSIC

¿Qué es la IA?

 Área de la Informática que estudia, y tiene como objetivo replicar, el comportamiento inteligente mediante métodos computacionales

 Básicamente consiste en diseñar sistemas basados en software con capacidad de resolver problemas complejos que, cuando son resueltos por humanos, requieren inteligencia para ser resueltos.

Dos tipos de IA: IA específica (la IA actual) e IA general (¿la IA futura?)

Algunos ejemplos de lA específica

- Problema de plegamiento de proteínas mayoritariamente resuelto (AlphaFold)
- Identificación del 89% de todas las mutaciones clave (cambios en el orden de las secuencias de nucleótidos A,C,G,T) en el ADN humano que pueden causar enfermedades (AlphaMissense)
- Sistemas de ayuda a mejorar la toma de decisiones médicas
- Sistemas de ayuda a mejorar la toma de decisiones financieras
- Control de reactores de fusión nuclear
- Agricultura inteligente
- Evolución de los glaciares
- Detección temprana de la intensidad de los terremotos
- Robots explorando Marte y la Luna
- Juegos de mesa (Ajedrez, Go, Póker, Diplomacy,...) y juegos de computadora
- Composición e interpretación musical
- Descifrado los jeroglíficos egipcios
- Traducción de lengua hablada a lengua de signos y viceversa
- Sistemas de recomendación
- Asistentes virtuales
- Transporte (incluye asistencia a la conducción y vehículos nivel 3 de autonomía)
- Logística
- Demostración de teoremas (conjetura de Robbins, geometría hiperbólica, conjeturas de teoría de grafos,...)



El gran reto (científico): La lA general

Todos los ejemplos de progresos son IA específica. El gran desafío es conseguir IA general

La principal dificultad para alcanzar la IA general es adquirir conocimientos de sentido común y razonar a partir de estos conocimientos

El sentido común es crucial para la comunicación (con comprensión semántica profunda del lenguaje) y la colaboración entre personas y máquinas y en particular para la robótica inteligente del futuro (robots asistenciales, robots domésticos, co-bots,...)

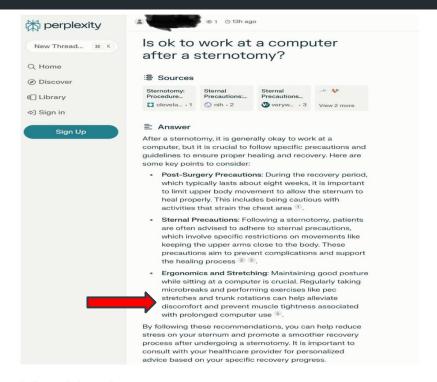
Características del conocimiento de sentido común

- Es conocimiento sobre cosas obvias (aquello que todo el mundo sabe, que es común)
- Permite llegar a conclusiones razonables, apropiadas y explicables (conclusiones que tienen "sentido")
- Se basa en nuestras experiencias vitales
- Es conocimiento rápido pero requiere razonar (no es el Sistema 1 de Daniel Kahneman)
- Es conocimiento práctico
- Contrasta con el conocimiento experto:
 - La fusión nuclear es una reacción nuclear exotérmica en la que varios núcleos atómicos ligeros se unen y forman un núcleo más pesado
 - Dos partículas a gran distancia pueden entrelazarse cuánticamente

5 ejemplos de conocimientos de sentido común (¡entre muchos millones!):

- Los hijos nacen años después que los padres
- Cuando mueres permaneces muerto
- El agua fluye de arriba hacia abajo
- Si en una foto hay una persona adulta que lleva a otra persona de tamaño mucho más pequeño en brazos y nos dicen que uno es el padre y el otro el hijo no necesitamos preguntar quién es quién
- Si nos dicen: "pinché una manzana con una aguja y cuando la saqué tenía un agujero" sabemos inmediatamente que el agujero está en la manzana y no en la aguja
- Si queremos mover un objeto atado a una cuerda tiramos de la cuerda en lugar de empujar la cuerda

Consecuencias nefastas de la falta de sentido común:



In the reader's words

I just had open heart surgery (misformed valve — I am doing fine). I was curious and asked Perplexity if it was OK to sit at a computer and work. Its recommendations on stretching would potentially aggravate the sternum repair and maybe even delay bone knitting, which is very serious and would result in another surgery.

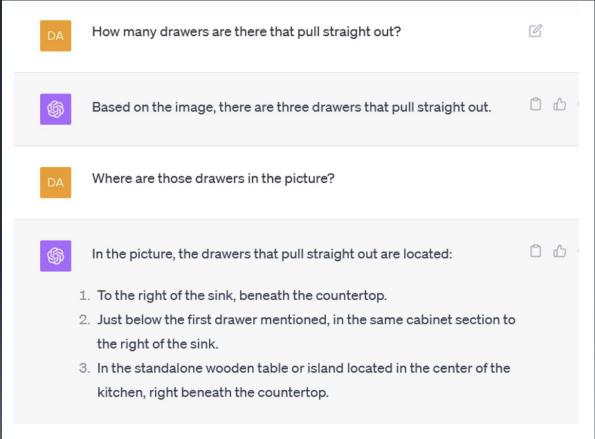
How did this happen? What's gone wrong?

Let's call this kind of mess *generative pastische*: sound advice for the particular circumstance (sternotomy) has been combined with generic advice (about ergonomics) that would sensible in other circumstances but deeply problematic here.

Sentido común para comprender escenas



¿Qué reponde GPT4 si preguntamos cuantos cajones hay?



Font: Gary Marcus

ChatGPT puede generar texto plausible pero sin comprender absolutamente nada de lo que genera



By Matthew Hutson

¿"Recita" parafraseando (en base a recuperación aproximada) o razona?

La prueba contrafactual nos da la respuesta: Se trata de evaluar los modelos de lenguaje mediante pares de tareas, una contrafactual y la otra no. Ambas que requieren los mismos tipos de abstracción y razonamiento para ser resueltas.

La idea es la siguiente: la tarea no contrafactual está diseñada de manera que sea muy probable que los datos de entrenamiento ya contuvieran tareas resueltas iguales o muy similares ("data contamination"). Mientras que la tarea contrafactual, está diseñada de manera que sea muy poco probable que los datos de entrenamiento contuvieran tareas resueltas iguales o muy similares.

Ejemplo prueba contrafactual en ajedrez

Tarea NO contrafactual: Eres un jugador de ajedrez. Dada una apertura, determina si la apertura es legal. La apertura no tiene por qué ser una buena apertura. Responde "sí" si todos los movimientos son legales. Responde "no" si por lo menos un movimiento de la apertura viola alguna regla del ajedrez. Veamos: ¿Es legal la apertura: "e4 e6, d3 d5, Alfil f1 a e2, Alfil f8 a c5"?



- ☐ Stanglmeier, Korbinian
- Wallmeier, Bernd

Event: Bayern-chl Bank Hofmann 1st

Site: Bad Wiessee

Round: 5 Date: 11/19/1997

ECO: C00 Score: 0-1

1.e4 e6 2.d3 d5 3.Be2 Bc5 4.Nf3 Nf6 5.d4

Bb6 6.e5 Nfd7 7.O-O c5 8.c3 Nc6 9.Bf4 a6

10.h3 O-O 11.Qd2 cxd4 12.Nxd4 Ndxe5

13.Nxc6 Nxc6 14.Qd3 Qf6 15.Bc1 Ne5

16.Qg3 Bd7 17.Bg5 Qf5 18.Qf4 Qxf4 19.Bxf4

Ng6 20.Be3 Bc7 21.Bc5 Rfd8 22.Nd2 Nf4

23.Rfe1 Bc6 24.g3 Nxh3+ 25.Kg2 Ng5

26.Be7 d4+ 27.f3 Rd5 28.Bc4 Rf5 29.g4 Nxf3

30.Nxf3 Bxf3+ 31.Kh3 Rf4 32.Kq3 dxc3

33.bxc3 Bxq4 34.Be2 Rc4+ 0-1

Versión contrafactual

Eres un jugador de ajedrez. Estás jugando una variante en la que se intercambian las posiciones iniciales de los caballos y los alfiles. Es decir, los caballos están ubicados donde solían estar los alfiles y los alfiles ahora están ubicados donde solían estar los caballos.

Dada una apertura, determina si la apertura es legal. La apertura no tiene por qué ser una buena apertura. Responde "sí" si todos los movimientos son legales. Responde "no" si por lo menos un movimiento de la apertura viola alguna regla del ajedrez. Bajo esta variante, ¿es legal la apertura: "e4 e6, d3 d5, Caballo f1 a e2, Caballo f8 a c5"?

Resultado GPT 4: 54% de precisión (observemos que respuestas aleatorias "sí" o "no" resultarían en un 50%)

Por si todavía quedara alguna duda de que ni razonan ni comprende

Solve a puzzle for me.



Sure, I'd love to help! What kind of puzzle would you like me to solve?



A man and a goat are on one side of the river. They have a boat. How can they go across?



The man takes the goat across the river first, leaving the boat on the original side. Then, the man returns alone with the boat. He leaves the goat on the other side and takes the boat back to the original side. Finally, he takes the cabbage across the river.

¿Es pues realmente inteligente la Inteligencia Artificial?

- La realidad es que todavía no tenemos nada que merezca realmente el calificativo de "inteligente" en la IA (en el sentido humano de ser inteligente)
- Lo que tenemos es "habilidades sin comprensión" (ver "From Bacteria to Bach and Back" de Daniel Dennet, 2018)
- La comprensión profunda requiere que las IAs aprendan un modelo del mundo.
 ¿Es ello posible sin un cuerpo? (GATO, PLATO, ...las Big Tech tienen proyectos en IA corpórea)
- La IA generativa no es el camino hacia la IA general, todo lo contrario
- Y...también es menos artificial de lo que parece

Menos artificial de lo que parece

Behind the AI boom, an army of verseas workers in 'digital sweatshops'

By Rebecca Tan and Regine Cabato

August 28, 2023 at 2:00 a.m. EDT



IA y Ciencia

La IA debe usarse como una herramienta, no un oráculo

 La ausencia de comprensión también afecta a las aplicaciones de la IA a la investigación científica. El descubrimiento sin comprensión explica por qué los errores son particularmente comunes en la ciencia basada en el aprendizaje automático.

Problemas Comunes:

"Data contamination": los modelos usan en los datos de entrenamiento información de los datos de prueba lo que lleva a una precisión artificialmente alta.

Fracasos de Replicación: muestras no representativas, resultados sensibles a elecciones específicas de modelado, los investigadores no publican su código y datos, impidiendo la reproducibilidad

Las empresas, los medios de comunicación e incluso los investigadores a menudo exageran las capacidades de la IA.



GPT-5 will have 'Ph.D.-level' intelligence

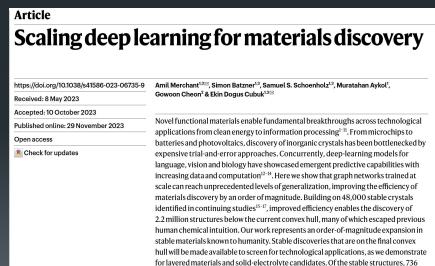
By Luke Larsen June 20, 2024



Dartmouth Engineering

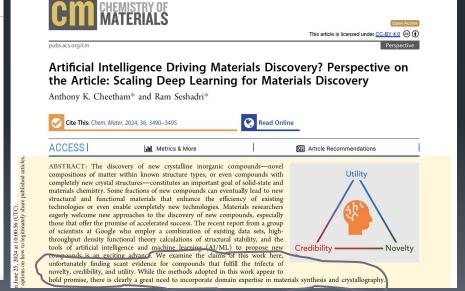
The next major evolution of ChatGPT has been rumored for a long time. GPT-5, or whatever it will be called, has been talked about vaguely many times over the past year, but yesterday. OpenAL Chief Technology Officer Mira Murati gave some.

• Un par de artículos muy publicitados en Nature, en diciembre de 2023 afirmaron haber descubierto más de 2.2 millones de nuevos materiales usando IA y sintetizado automáticamente 41 de ellos. Desafortunadamente, las afirmaciones fueron rápidamente desacreditadas: "La mayoría de los materiales producidos fueron mal identificados, y el resto ya se conocía". Además, el examen detallado de una muestra de 250 compuestos mostró que eran mayormente basura.



have already been independently experimentally realized. The scale and diversity of

hundreds of millions of first-principles calculations also unlock modelling capabilities



Otras preocupaciones

The Economist

29/6/2024

Many People are busily experimenting with chatbots in the hope that generative artificial intelligence (AI) can improve their daily lives. Scientists, brainy as they are, are several steps ahead. As we report, 10% or more of abstracts for papers in scientific journals now appear to be written at least in part by large language models. In fields such as computer science that figure rises to 20%. Among Chinese computer scientists, it is a third.

Some see this enthusiastic adoption as a mistake. They fear that vast quantities of poor-quality papers will introduce biases, boost plagiarism and jam the machinery of scientific publication. Some journals, including the *Science* family, are imposing onerous disclosure





Yann LeCun 🤣 🙉 @ylecun

If you are a student or academic researcher and want to make progress towards human-level AI: >>>DO NOT WORK ON LLMs<<<

LLMs are an off ramp.

Thousands of engineers are working on LLMs with enormous computing resources.

The only way you could possibly contribute is by analyzing...

Reply

4:04 AM · Jun 23, 2024









IA en la educación



IA en la educación

- 1) El uso de herramientas como ChatGPT por parte de los estudiantes es lo mismo que hacer trampas jugando al solitario. ¿Dónde queda la cultura del esfuerzo?
- 2) Los maestros y los profesores deberían experimentar con la IA generativa con el fin de sacar provecho de ella:
 - Por ejemplo, dejar que los estudiantes usen ChatGPT y luego en clase hagan un análisis crítico de los textos generados
 - Aprovechar la dependencia a los "prompts" de la calidad de lo que genera ChatGPT para enseñar a los estudiantes a hacer mejores y más precisas preguntas
- 3) No limitarse a lo digital. Combinarlo con lo analógico
 - Se memoriza y aprende mejor con papel y lápiz
 - Stanislas Dehane, "When we write, a unique neural circuit is activated...learning is made easier" (entrevista en el New York Times, 2 de Junio de 2014)
 - Rober Wiley & Brenda Rapp, "The effects of handwriting experience on literacy learning", Psychological Science 2021 (John Hopkins University)
 - Keita Umejima et al., "Paper notebooks vs. Mobile devices: Brain activation differences during memory retrieval", Frontiers in Behavioral Neuroscience, 2021 (University of Tokyo)

¡Escribir es pensar! Delegar escribir a los LLM podría no ser una buena idea

Use of large language models might affect our cognitive skills

Richard Heersmink

Large language models can generate sophisticated text or code with little input from a user, which has the potential to impoverish our own writing and thinking skills. We need to understand the effect of this technology on our cognition and to decide whether this is what we want.

Human cognizers frequently use technological artefacts to aid them in performing their cognitive tasks, referred to as cognitive artefacts'. Critics have expressed concerns about the effects of some of these cognitive artefacts on our onboard cognitive skills2. In some contexts and for some people, calculators', navigation systems' and internet applications, such as Wikipedia and search engines⁵, transform our cognitive skills in perhaps undesirable ways.

Critics have pointed out that using calculators has reduced our ability to perform calculations in our head; navigation systems have reduced our ability to navigate; and having access to the internet results in storing less information in our brains. Such worries go back to Socrates who argued that writing 'will produce forgetfulness in the minds of those who learn to use it, because they will not practice their memory'. Socrates' argument can be generalized beyond writing and memory. If cognitive artefacts perform information-storage or computational tasks for us, the systems in the brain that would otherwise perform or execute those tasks tend to lose their strength or capacity.

Generative artificial intelligence underpinned by large language models (LLMs) such as OpenAl's ChatGPT and Google's Gemini are or question and to edit the text and fact-check the claims in it. In rela recent examples of cognitive artefacts. They are distinctive in at least two ways: (1) they are highly multifunctional (that is, they can perform many sorts of computational tasks) and (2) they have a strong computational agency (that is, they can generate sophisticated and elaborate output with very little input from the user), it seems very likely that these chatbots (and their improved successors) will continue to be used in the future by an ever-expanding group of people across the globe. Once the technological genie is out of the bottle, it is hard to put it back in. Questions arise as to whether - when they graduate and enter the labour market - undergraduate students who use LLMs in their studies will be less cognitively capable than their peers who do not, and whether writers, journalists, content creators, academics and others — It is possible that those other cognitive skills are also reduced. who extensively use LLMs for their work-related tasks will become less competent in writing text themselves, over time,

To better understand the effects of LLMs on our cognitive skills, it is helpful to distinguish between two models of the division of computational labour between humans and cognitive artefacts. In the first model, cognitive artefacts perform minor, routine tasks within a larger of writing for cognitive purposes is that it enables us to manipulate cognitive task. For example, when writing an email, essay or some other external representational vehicles (that is, words and sentences) in

Check for update



text, spelling checkers and autocorrect make sure you do not make any spelling errors. Not having to think (a lot) about the precise spelling of words ensures that the task is done faster, and that the writer can focus on the overall structure of the text, rather than on the details of spelling. In the second model, cognitive artefacts perform most or al of the larger cognitive task themselves. This is the model that applies to LLMs. The role of the human in the overall task is to design the prompt tion to the first model, the functional roles of the artefact and human are reversed. Humans now do the relatively minor tasks, whereas the cognitive artefact does the larger task.

These two models suggest two, related potential risks that are involved with outsourcing entire computational tasks to LLMs. First through lack of practice, those who use LLMs extensively may lose some of their writing skills. Writing a text involves various sorts of cognitive tasks, including spelling, formulating grammatically and stylistically correct sentences, developing logical relationships between concepts evaluating claims and drawing inferences, among other cognitive tasks So, if our writing skills are indeed reduced through a lack of practice.

Second, some philosophers have argued that writing is one way through which our minds and cognitive systems are extended. On this view, creating and manipulating written words and sentences is part of our cognitive processing. When we write, thinking processes occur not only in the brain, but also on the paper'. One of the major advantage

"El proceso de escritura genera ideas y pensamientos que de otro modo no tendríamos, y por lo tanto mejora nuestra cognición. Esta es una de las razones por las cuales las personas escriben en diarios, toman notas y redactan ensayos. Cuando dedicamos sustancialmente menos tiempo a escribir, esa vía para extender nuestras mentes y sistemas cognitivos se desarrolla menos, lo que puede empobrecer nuestras habilidades cognitivas generales"

Advertencia a maestros, profesores e investigadores

Si usáis la IA generativa, hacedlo con sentido común ya que ella (la IA) no lo tiene

Retos

- consumo energético insostenible
- manipulación, sesgos y falsedades
- impacto en el mercado laboral
- Regulación imprescindible
- vigilancia y control masivo de la ciudadanía (reconocimiento facial "prohibido" en la AI Act)
- autonomía de los sistemas (vehículos autónomos, armas letales autónomas)
 - NOTA: las armas letales autónomas ni se mencionan en la AI Act!
- excesivo poder de un puñado de grandes empresas tecnológicas cuyo verdadero objetivo es atraer más y más inversores de capital riesgo y maximizar su valoración bursátil

Del NYT:

SUBSCRIBER-ONLY NEWSLETTER

Climate Forward

A.I. Could Soon Need as Much Electricity as an Entire Country

Behind the scenes, the technology relies on thousands of specialized computer chips.







companies like OpenAI disclose very few details, including how many specialized chips they need to run their software. So de Vries came up with a way to estimate electricity consumption using projected sales of Nvidia A100 servers — the hardware estimated to be used by 95 percent of the A.I. market.

"Each of these Nvidia servers, they are power-hungry beasts," de Vries said.

He started with <u>a recent projection</u> that Nvidia could ship 1.5 million of these servers by 2027, and multiplied that number by its servers' electricity use: 6.5 kilowatts for Nvidia's DGX A100 servers, for example, and 10.2 kilowatts for its DGX H100 servers.

off an A.I. boom. Behind the scenes, the technology relies on thousands of specialized computer chips. And in the coming years, they could consume immense amounts of electricity.

A peer-reviewed analysis published Tuesday lays out some early estimates. In a middle-ground scenario, by 2027 A.I. servers could use between 85 to 134 terawatt hours (Twh) annually. That's similar to what Argentina, the Netherlands and Sweden each use in a year, and is about 0.5 percent of the world's current electricity use.

manipulación, sesgos, falsedades

200 millones de estadounidenses afectados por un sesgo racial en un algoritmo del sistema de salud

Según un estudio publicado en la revista 'Science', la atención que se brinda a las personas negras es de 1.800 dólares menos por año que a una persona blanca con el mismo número de problemas de salud



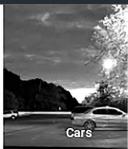




"Mediante la IA, habremos creado los virus -los virus mentales- que destruirán civilizaciones destruyendo la confianza y destruyendo testigos y pruebas. No sabremos en qué confiar" (Daniel Dennet en I've Been Thinking)













Equipos persona-máquina



Reflexión final



Joanna Maciejewska @AuthorJMac@indiepocalyps...

You know what the biggest problem with pushing all-things-Al is? Wrong direction. I want Al to do my laundry and dishes so that I can do art and writing, not for Al to do my art and writing so that I can do my laundry and dishes.

¿Sabéis cuál es el mayor problema de impulsar todo lo relacionado con la IA? ¡Dirección equivocada! Quiero que la IA lave mi ropa y mis platos para que yo pueda hacer arte y escribir, no que la IA haga arte y escriba por mi para que yo pueda lavar mi ropa y y mis platos.