mdt.

Ex Datis, Claritas

¿Seremos asesinados por una IA, próximamente?

Un estudio confirma un secreto a voces: a los LLM aún les queda mucho para ser fiables en el uso diario, y más todavía para ser una IA realmente inteligente.

Manuel Delgado Tenorio · 24 de septiembre de 2023

Publicado originalmente en www.manueldelgado.com/blog/seremos-asesinados-por-una-ia-proximamente



Un estudio pone negro sobre blanco un secreto a voces: que a los LLM aún les queda mucho camino por recorrer para ser fiables en el uso diario. Más lejos queda, por tanto, la IA realmente inteligente.

Si quieres comprender de verdad el estado actual de los grandes modelos de Inteligencia Artificial Generativa (GenAI), no te puedes perder este artículo de Gary Marcus y el *paper* al que enlaza.

De forma muy resumida, el artículo académico pone de manifiesto cómo los grandes modelos de lenguaje (LLM) actuales **fallan estrepitosamente** en una tarea que, para los seres humanos, es trivial: las relaciones del tipo "*si Carlos es el hijo de María, entonces María es la madre de Carlos*".

Que los modelos actuales más famosos, como GPT-4, fallen de forma tan burda en una tarea cognitiva tan básica debe servirnos para dos cosas, principalmente:

- Ser conscientes de las capacidades reales de estos sistemas y, en consecuencia, modular nuestras expectativas sobre ellos, lo que incluye establecer las salvaguardas necesarias para, al usarlos, protegernos de sus habituales errores y debilidades.
- Reafirmar la idea de que estamos MUY, MUY LEJOS de nada parecido a la Inteligencia Artificial General (AGI), que a menudo se nos vende como máquinas inteligentes, conscientes de sí mismas y con voluntad propia.

Me centraré ahora solo en el segundo punto. Si para cualquiera mínimamente metido en este mundo de la IA es tan evidente que el estado del arte actual está TAN alejado de esa AGI, **¿por qué sigue siendo tema habitual de conversación en entornos sobre los que se toman decisiones relevantes sobre estas cosas?**

Mi opinión: porque a muchos les interesa exagerar las capacidades potenciales de sus sistemas y, también, que los políticos y los reguladores se distraigan con la fantasía de las máquinas pensantes en lugar de fijarse bien en los riesgos, más numerosos y reales, de la tecnología actual.

Por eso, con cierta frecuencia vemos a algún personaje relevante lanzar un huesecillo para desviar la atención y tener a la gente preocupada por si su Thermomix se transforma, de repente, en un Terminator.

A estas alturas de la vida, ya ni me sorprende ese uso interesado de la idea de una inteligencia artificial realmente inteligente.

Todo viene de la ciencia-ficción

Me resulta más interesante **entender cómo el concepto de máquina inteligente (y, comúnmente, malévola) ha calado tanto en el imaginario cultural compartido como para que resulte tan eficaz a la hora de distraer nuestra atención** de los problemas cercanos y reales.

Su potencia y eficacia radican en que la discusión sobre la AGI se nutre de cientos de relatos fantásticos a los que hemos estado expuestos durante toda nuestra vida. Sin embargo, esos relatos son solo eso: fantasía, sin relación alguna con las realidades técnicas en las que supuestamente se basan.

Echemos la vista atrás: la posibilidad de que las máquinas cobren vida y, además, se rebelen contra nosotros alimenta la imaginación de escritores y cineastas **desde hace ya más de un siglo.**

Por ejemplo, la palabra *robot* la acuñó un escritor checo de ciencia-ficción, Karel Čapek, **en 1920**: hablamos de una época en la que el medio de transporte más común todavía era el caballo y, sin embargo, ya se **fantaseaba** con la fabricación masiva de máquinas androides.

LA CLAVE ESTÁ EN QUE NUNCA ESAS NOVELAS O PELÍCULAS HAN SIDO UN REFLEJO DE LAS CAPACIDADES TECNOLÓGICAS DEL MOMENTO PERO, POR REPETICIÓN Y HABITUALIDAD, SE HAN INCORPORADO A NUESTRA VISIÓN DEL MUNDO COMO SI LO FUERAN

Ahora, interactuamos con ChatGPT y creemos estar hablando con HAL9000, el cerebro inteligente de la nave de "2001 — Odisea en el espacio", pero la realidad es que 2001 es una película del año 68, basada en una historia escrita veinte años antes, cuando los aviones aún eran de hélice y se empezaba a generalizar el uso de antibióticos.

Las capacidades de ese supercomputador de novela que ahora queremos ver reflejadas en la tecnología actual fueron el producto de la fantasía de su autor, del mismo modo que visualizamos sin esfuerzo dragones voladores o hadas del bosque, sin relación alguna con las posibilidades reales de que algo así exista.

Así que debemos ser conscientes de que, cada vez que hablamos de la posibilidad de construir máquinas inteligentes, todas esas referencias literarias y cinematográficas con las que hemos crecido distorsionan nuestra capacidad de análisis. **Bajamos las defensas de nuestro juicio crítico y damos por cercano y posible algo que, por el momento, sigue siendo lejano y teórico.**

Con ese sustrato, incluso personas con conocimientos robustos sobre tecnología se maravillan ante lo que perciben como "atisbos de consciencia" o, peor aún, "señales de humanidad" al utilizar modelos de lenguaje cuyo funcionamiento, si se me permite la simplificación, no consiste más que en elegir la palabra más adecuada para continuar una secuencia de otras palabras. Y que, como demuestra el artículo que mencionaba al principio, se aturullan si, simplemente, les cambiamos el orden de las palabras.

Con todo lo bonito, útil y práctico que podemos sacar de estos recientes avances tecnológicos, debemos desterrar de una vez el estéril e infundado debate sobre la amenaza de las máquinas inteligentes.



Sobre el autor

Manuel Delgado Tenorio

Manuel Delgado Tenorio es asesor independiente de estrategia de datos e inteligencia artificial, profesor e investigador. Cofundó y vendió una *startup* de inteligencia artificial, fue socio de la práctica de Data & Analytics para servicios financieros en EY y responsable de Data Analytics para el sur de Europa en AWS. Es profesor asociado en IE University, donde imparte la primera asignatura de grado sobre Customer Lifetime Value en España, investiga sobre el CLV en su doctorado en la Universidad Complutense de Madrid y es autor del libro "Customer Lifetime Value (CLV): No todos tus clientes valen lo mismo".

manueldelgado.com · LinkedIn · X

Recibe cada artículo en tu correo con la newsletter Ex Datis, Claritas:
manueldelgado.com/newsletter/suscribete