Transversal

José García Montalvo

Catedrático de Economía (UPF)

IA: "Sempre negatifo"

"Sempre negatifo, nunca positivo". Esta frase, que ha quedado en la historia como reflejo de la personalidad del que fuera entrenador del FC Barcelona Louis van Gaal, se puede aplicar a una gran cantidad de analistas que hablan sobre inteligencia artificial y, en particular, sobre los LLM (*large language models*) tipo ChatGPT. Que si alucina, que si favorecen la desinformación, que si consume mucha energía eléctrica y favorece el calentamiento global, que si tiene sesgos, que si podría facilitar la formación de ejércitos de robots, que si podría acabar controlando el mundo y esclavizando a la raza humana... Esto ha llevado a una inflación de centros y conferencias sobre ética e inteligencia artificial, y llamadas a su estricta regulación, especialmente en la Unión Europea. Esto recuerda un poco a los inicios de la revolución de la mecánica cuántica y la polémica sobre su interpretación. A pesar de ser una teoría con predicciones increíblemente precisas que han servido para el desarrollo de tecnologías de computación y comunicación cada vez más potentes, una parte de las discusiones se han centrado en su interpretación: que si la versión de Copenhague, que si los mundos paralelos, que si interpretaciones informacionales, que si el colapso espontáneo... Hasta que alguien exclamó el famoso "calla ya la boca y ponte a calcular". De forma muy simplista, podemos decir que mientras la UE regula y aprueba normas, en Estados Unidos se investigan nuevas aproximaciones para mejorar los LLM y se crean start-ups. Unos hablan y otros calculan.

No soy tan ingenuo como para pensar que la IA no puede presentar problemas en su aplicación y deberían adoptarse algunas salvaguardias, pero parece claro que los efectos positivos de esta tecnología son abrumadores; ver el vaso siempre medio vacío o totalmente vacío no es una visión ajustada a la realidad. Es cierto, como ya les he contado desde esta misma columna en alguna ocasión, que los efectos positivos de la IA sobre la productividad no se observan todavía. Pero también es cierto que la adopción de las tecnologías por el sistema productivo precisa de un cierto tiempo. Este mismo efecto de impacto con retardo sobre la productividad se ha producido en el pasado con muchas otras tecnologías.

Pero vayamos por partes. En primer lugar, los LLM alucinan... aunque quizás por poco tiempo. El problema fundamental es que cuando reciben una pregunta o intentan resolver un problema siempre buscan una respuesta o solución. Esto es consecuencia directa del funcionamiento interno del modelo, que acaba determinando unas probabilidades y seleccionando la palabra que tiene mayor probabilidad. En muchos casos, la respuesta debería ser que no se puede dar una respuesta. Pero las técnicas de aprendizaje reforzado pueden resolver este problema. Si cuando se hace una pregunta se puede valorar si la respuesta es adecuada o no, el sistema finalmente podría acabar con respuestas del tipo "no lo sé". Además, el aprendizaje del modelo en entornos limitados a un conjunto documental relevante para el tipo de preguntas que se deben contestar minimiza la probabilidad de alucinar. Pensemos en una aplicación que responde a preguntas de empleados de una compañía que tienen que aplicar la normativa sectorial y de la empresa ante solicitudes de sus clientes y que ha tenido un aprendizaje a partir de dicha normativa.

En segundo lugar, es cierto que se puede crear bots basados en IA para expandir bulos y desinformación. Pero, como muestra un artículo publicado hace una semana en la revista *Science*, los chatbots basados en IA son mejores que los humanos en convencer de la verdad a personas que creen en todo tipo de conspiraciones absurdas. En una conversación de tres interacciones con GPT-4 turbo, los creyentes en una conspiración concreta reducían un 21,4% su creencia. El grupo de control, aquellos que conversaron con GPT-4 de otras cosas y no de su particular creencia, no mostraron ningún cambio significativo en su creencia conspiranoide. Además, un 27,4% de los participantes que conversaron sobre su visión conspirativa se mostraron inseguros con posterioridad sobre la verdad de dicha conspiración frente a solo el 2,4% del grupo de control. Y este cambio de creencia se mantenía dos meses después de la interacción con la IA. Los resultados eran similares para conspiraciones muy diferentes, como la relativa a la muerte de la princesa Diana de Gales, el falso alunizaje del Apollo 11 o la participación de la CIA en los atentados del 11 de septiembre en EE.UU.

Eficiencia

La IA demanda mucha energía, pero aportará infinidad de posibles soluciones para favorecer la transición energética

En tercer lugar, es cierto que las demandas energéticas del funcionamiento de estos modelos son elevadas. Pero, como les conté hace unos meses, las posibilidades de la inteligencia artificial para favorecer la transición energética son inmensas. Los algoritmos de IA permitirán una gestión más eficiente de la red eléctrica que genera predicciones en tiempo real de la demanda de energía y ajusta dinámicamente la generación y la distribución. Se podrá optimizar la generación de energía en proyectos híbridos calculando millones de configuraciones entre fuentes de energía renovable y almacenamiento. La IA también colabora en la búsqueda de nuevos materiales para hacer baterías más eficientes que permitan superar la intermitencia característica de las energías renovables y en la optimización de la red eléctrica en función de la distancia a los productores y del origen de la energía. Y otras cosas como la búsqueda de mejores conductores para las redes o el control de la seguridad de las centrales nucleares.

En cuarto lugar, la IA tiene sesgos. Esto es innegable, pero la crítica sería más sustancial si los humanos no tuvieran sesgos. De hecho, ya hay mucha investigación científica que muestra casos donde un sistema de IA produce resultados menos sesgados que un humano, por ejemplo, jueces concediendo fianzas. Y los sesgos humanos son más difíciles de explicar que los generados por los sistemas de IA. En principio, se podría hacer auditorías y utilizar otros LLM para analizar los motivos de los sesgos de un sistema de IA. Estos procedimientos son más precisos que un psicoanalista intentando averiguar qué trauma infantil genera un particular sesgo humano. Bueno, les dejo pensando, que yo me voy a calcular.

Realidades

La UE regula y aprueba normas, en Estados Unidos se investigan nuevas aproximaciones para mejorar los LLM y se crean start-ups. Unos hablan y otros calculan