

Groquesco

FABRIZIO MEJÍA MADRID :: 11/08/2025

Las 'inteligencias' artificiales como Grok tienen incrustado el racismo, el neocolonialismo y el sexismo en su estructura

Hace dos semanas sucedió algo en principio divertido en la red X. Un usuario le pidió a su modelo de lenguaje de inteligencia artificial (IA), llamado Grok, que le explicara algo al caricaturista José Hernández sobre el *Che* Guevara. El dispositivo le recomendó al monero *Hernández* que se pusiera a leer la biografía que sobre el revolucionario había escrito John Lee Anderson. Grok nunca supo que a quien le estaba aconsejando que leyera ese libro lo había convertido en novela gráfica en tres tomos, en coautoría precisamente con John Lee Anderson. Por más que *Hernández* le daba pistas, Grok re combinaba sus frases algorítmicas sin sentido alguno como en una especie de oráculo Monty Python.

El incidente me dejó pensando en qué tipo de IA nos venden como si fuera la síntesis de todo. Grok no mencionó más que a Anderson, periodista estadounidense. Obvió al coautor, el propio monero mexicano, y a otros grandes biógrafos del *Che*, como Paco Ignacio Taibo II. A los algoritmos se les ha llamado optimizadores del sesgo porque sólo aprenden de las fuentes tendenciosas con que alguien de carne y hueso los alimentó y con el tiempo las van acentuando. Tienen incrustado el racismo, el neocolonialismo y el sexismo en su estructura.

Grok, por supuesto, no sabe nada ni miente ni tiene ningún tipo de sabiduría, pero nos lo venden como un software autónomo que decide cuál es el contenido valioso que debes conocer. Los grandes modelos de lenguaje (LLM, por sus siglas en inglés), por ejemplo, han recomendado comer piedras como si fueran alimentos nutritivos, han asegurado que Gibraltar es una creación humana o han tenido la osadía de emitir un manual para saber qué hongos son venenosos. También aconsejaron a un usuario utilizar pegamento para adherir a la masa los ingredientes de una pizza.

No son lo que prometen los magnates de la IA. Maciej Ceglowski, ex empleado de Yahoo y crítico de Silicon Valley, lo tiene muy claro: Cuando observamos dónde la IA realmente está teniendo éxito, no es en algoritmos complejos que se automejoran recursivamente. Es en el resultado de volcar cantidades ingentes de datos en redes neuronales relativamente simples. Las construcciones que usamos en IA son bastante opacas después del entrenamiento. No funcionan como el escenario de superinteligencia requiere que funcionen. No hay forma de ajustarlas recursivamente para mejorarlas, salvo reentrenándolas con aún más datos.

Y el texto con que fue entrenada la IA es la Internet, así que muy pronto estará convertida en basura informativa, llena de xenofobia, teorías conspirativas, neocolonialismo intelectual y animaciones de gente alimentándose con piedras. Será como leer la copia de una copia de una copia: cada vez más tonterías borrosas generadas por computadora.

El asunto podría tornarse desastroso. Google, por ejemplo, ha decidido no distinguir entre el contenido real y el generado por IA en sus noticias, lo que implica una irresponsabilidad

mayúscula que puede generar reacciones reales a partir de una mentira. Como ejemplo está lo que ocurrió en Gran Bretaña el 29 de julio del año pasado, cuando se divulgó por las redes que el asesino de tres niñas en Southport era un inmigrante musulmán sin papeles que había llegado en un bote por el Canal de la Mancha. Manchester, Liverpool, y Londres vieron arder mezquitas y hoteles de solicitantes de asilo cuando el criminal real era un joven nacido en Gales.

Si los usuarios sólo consultan la IA que, a su vez, se alimenta de Internet, no pueden ya diferenciar una falsedad de algo comprobable. Si existe ya una generación que obtiene toda su información de las redes y no de libros o de autoridades epistémicas, y que está desprotegida ante la eventualidad de que ya no sepamos que la Tierra es realmente esférica, achatada en los polos y con un eje de rotación, estamos en la puerta de una crisis de ignorancia planetaria. Los antivacunas y la ultraderecha han crecido de la mano de las verdades que son improbables, salvo por el yo le creo a Reddit o lo dice la Wikipedia.

Hasta ahora, la IA sirve de juego para animar a Trump como si fuera un bebé o hacerte una autocaricatura al estilo del estudio de animación japonés Ghibli, pero al entusiasmo le ha seguido una especie de pareidolia, es decir, de esa capacidad que tenemos los seres humanos de ver a fenómenos aleatorios como si fueran familiares, como ver personajes en las nubes. Con los modelos del lenguaje tenemos la desacertada impresión de que hay alguien con un juicio válido detrás del masacote de conexiones de la conversación simulada. Estamos siempre dispuestos a atribuirle autonomía a las cosas.

Como sólo conocemos a humanos hablándonos de forma inteligible, le atribuimos intención, pensamiento y consciencia a entidades que usan el lenguaje y con las que creemos que estamos conversando. Discutimos con la tele, le gritamos a una aplicación, le mentamos la madre a Alexa. El 80 por ciento de los dueños de una aspiradora marca Roomba le han puesto nombre. Con modelos como Grok vemos la ilusión de significado: el texto debe de haber sido escrito por alguien que entiende, si no del tema, por lo menos de lo que está escribiendo, pero no es así. Los LLM sólo van haciendo plausible una relación de frases que nos da la impresión de que entiende como nosotros. El problema es cuando se utiliza para decidir a quién darle un crédito o un trabajo o, peor, para predecir dónde y quién cometerá un delito y expedir una sentencia de cárcel o un asesinato de civiles en Gaza.

Pero nuestra reacción ante las respuestas algorítmicas es que no importa lo que avance la tecnología, seguimos siendo los mismos: tendemos a deshumanizar a las personas, mientras humanizamos a los no-humanos. El problema ahora es pensar quién decide cuál es la humanidad y cuál no. La tradición del robot, desde el inicio, cuando Homero describe a las ayudantes voladoras de Hefesto, es que hacen los trabajos peligrosos o aburridos. Pero los LLM nos acercan a algo más perverso: que hagan música, poemas, ensayos, definiciones de sentido, que nos digan qué hacer, como si las personas fuéramos todas uniformes, con las mismas necesidades, valores, y metas. Muy pronto la uniformidad a la mano serán la ignorancia y la aburrición.

El episodio que más me indigna es cuando, en 2017, los saudíes le entregaron la ciudadanía a una robot llamada Sophia. La rubia androide se había reunido con Angela Merkel en la Conferencia de Seguridad en Munich y en una asamblea general de Naciones Unidas

hablando con modelos LLM. Pensando en los millones de migrantes a los que se les niegan las ciudadanías y hasta los permisos de estancia todos los días, lo de Sophia fue un acto de crueldad. Pero señala ese otro espejismo de la IA: la idea de que es más sencillo para los blancos y billonarios deshacerse de los problemáticos humanos y sustituirlos con algo que simule que les desea salud cada vez que estornuden.

La Jornada

https://www.lahaine.org/mm_ss_mundo.php/grokesco