

NOAM CHOMSKY HABLA SOBRE PARA QUÉ ES REALMENTE BUENO CHATGPT

Por: Noam Chomsky. 03/07/2023

“Se trata de si el mundo seguirá siendo dominado por un imperio occidental o si permitirá que múltiples centros de poder coexistan y colaboren para abordar los desafíos comunes.” **por Noam Chomsky**

Transcripción de una entrevista realizada a Noam Chomsky el pasado 3 de Mayo del 2023 y publicado en la revista commondreams.org.

La inteligencia artificial (IA) está arrasando en el mundo. Está transformando todos los ámbitos de la vida y planteando importantes problemas éticos para la sociedad y el futuro de la humanidad. ChatGPT, que está dominando las redes sociales, es un chatbot basado en IA desarrollado por OpenAI. Es un subconjunto del aprendizaje automático y se basa en lo que se denomina Large Language Models (grandes modelos lingüísticos), capaces de generar respuestas similares a las humanas. El potencial de aplicación de esta tecnología es enorme, por lo que ya se está pidiendo que se regule la IA como **ChatGPT**.

¿Puede la IA ser más inteligente que los humanos? ¿Constituye una amenaza pública? De hecho, ¿puede convertirse en una amenaza existencial? Noam Chomsky, el lingüista más eminente del mundo y uno de los intelectuales públicos más estimados de todos los tiempos, cuya estatura intelectual se ha comparado con la de Galileo, Newton y Descartes, aborda estas inquietantes cuestiones en la entrevista que sigue.

C. J. Polychroniou: Como disciplina científica, la inteligencia artificial (IA) se remonta a los años 50, pero en las dos últimas décadas se ha ido abriendo paso en todo tipo de campos, como la banca, los seguros, la fabricación de automóviles, la música y la defensa. De hecho, el uso de técnicas de IA ha demostrado en algunos casos superar las capacidades humanas, como en una partida de ajedrez. ¿Es probable que las máquinas lleguen a ser más inteligentes que los humanos?

Noam Chomsky: Para aclarar la terminología, el término “máquina” significa aquí

programa, básicamente una teoría escrita en una notación que puede ser ejecutada por un ordenador, y un tipo de teoría inusual en aspectos interesantes que podemos dejar de lado aquí.

Podemos hacer una distinción aproximada entre ingeniería pura y ciencia. No hay un límite claro, pero es una primera aproximación útil. La ingeniería pura busca producir un producto que pueda ser de alguna utilidad. La ciencia busca la comprensión. Si el tema es la inteligencia humana o las capacidades cognitivas de otros organismos, la ciencia busca comprender estos sistemas biológicos.

Tal y como yo los entiendo, los fundadores de la IA -Alan Turing, Herbert Simon, Marvin Minsky y otros- la consideraban ciencia, parte de las entonces incipientes ciencias cognitivas, que utilizaban las nuevas tecnologías y los descubrimientos en la teoría matemática de la computación para avanzar en la comprensión. Con el paso de los años, estas preocupaciones se han desvanecido y se han visto desplazadas en gran medida por una orientación ingenieril. En la actualidad, las primeras preocupaciones se suelen tachar, a veces con condescendencia, de 'IA a la antigua usanza' [en inglés 'good old-fashioned AI' o su acrónimo 'GOFAI'].

Siguiendo con la pregunta, ¿es probable que se conciban programas que superen las capacidades humanas? Hay que tener cuidado con la palabra "capacidades", por razones a las que volveré. Pero si consideramos que el término se refiere a la ejecución¹ humana, entonces la respuesta es: definitivamente sí. De hecho, existen desde hace mucho tiempo: la calculadora de un ordenador, por ejemplo. Puede superar con creces lo que hacen los humanos, aunque sólo sea por falta de tiempo y memoria. En el caso de sistemas cerrados como el ajedrez, en los años 50 se comprendió perfectamente que tarde o temprano, con el avance de gigantescas capacidades computacionales y un largo periodo de preparación, se podría idear un programa capaz de derrotar a un gran maestro que juega con un límite de memoria y tiempo. El alcance de ese logro, años más tarde, fue más que nada un reclamo publicitario para **IBM**. Muchos organismos biológicos superan las capacidades cognitivas humanas en aspectos mucho más profundos. Las hormigas del desierto de mi patio trasero tienen cerebros minúsculos, pero superan con creces las capacidades humanas de navegación, en los principios que subyacen, no sólo en ejecución. No existe tal cosa como una Gran Cadena del Ser en cuya cúspide esté el ser humano.

Los productos de la **ingeniería de IA** se utilizan en muchos campos, para bien o

para mal. Incluso los más sencillos y conocidos pueden ser bastante útiles: en el área lingüística, programas como el autocompletado, la transcripción en directo, el traductor de google, entre otros. Con una potencia de cálculo mucho mayor y una programación más sofisticada, también deberían surgir otras aplicaciones útiles en las ciencias. Ya ha habido algunas: cómo ha asistido en el estudio del plegamiento de proteínas es un caso reciente en el que la tecnología de búsqueda masiva y rápida ha ayudado a los científicos a abordar un problema crítico y difícil.

Los proyectos de ingeniería pueden ser útiles, o perjudiciales. Ambas cuestiones se plantean en el caso de la ingeniería de la IA. El trabajo actual con grandes modelos lingüísticos (LLM por sus siglas en inglés), incluidos los chatbots, proporciona herramientas para desinformar, difamar y engañar a los no informados. Las amenazas aumentan cuando se combinan con imágenes artificiales y réplicas de voz. Con diferentes preocupaciones en mente, decenas de miles de investigadores de IA han pedido recientemente una moratoria en el desarrollo debido a los peligros potenciales que perciben.

Como siempre, hay que sopesar los posibles beneficios de la tecnología frente a los costes potenciales.

Se plantean cuestiones muy distintas cuando nos referimos a la IA y la ciencia. Aquí se impone la cautela debido a las afirmaciones exorbitantes e imprudentes, a menudo amplificadas en los medios de comunicación. Para aclarar las cuestiones, consideremos casos, algunos hipotéticos, otros reales.

He mencionado la navegación de los insectos, que es un logro asombroso. Los científicos especializados en insectos han avanzado mucho en el estudio acerca de cómo este logro se ha alcanzado, si bien la neurofisiología -una materia muy complicada- sigue sin aclararse del todo, al igual que sucede con la evolución de los sistemas. Lo mismo ocurre con las asombrosas proezas de aves y tortugas marinas que recorren miles de kilómetros y regresan infaliblemente al lugar de origen.

Supongamos que Tom Jones, un partidario de la ingeniería de la IA, viene y dice: "Todo tu trabajo ha sido refutado. El problema está resuelto. Los pilotos de líneas aéreas comerciales consiguen siempre los mismos resultados, o incluso mejores".

Si acaso tomáramos la molestia de responder, nos reiríamos.

Tomemos el caso de las hazañas náuticas de los polinesios, que todavía perviven entre las tribus indígenas, que por medio de las estrellas, el viento y las corrientes pueden desembarcar sus canoas en un punto designado a cientos de millas de distancia. También este caso ha sido objeto de muchas investigaciones para averiguar cómo lo hacen. Tom Jones tiene la respuesta: “Deja de perder el tiempo; los buques de guerra lo hacen todo el tiempo”.

La misma respuesta.

Pasemos ahora a un caso real: la adquisición del lenguaje. En los últimos años se han llevado a cabo investigaciones exhaustivas y muy esclarecedoras que han demostrado que los bebés tienen un conocimiento muy rico de la lengua (o lenguas) de su entorno, muy superior al que exhiben cuando actúan con ella[s]. Se consigue con pocas pruebas y, en algunos casos cruciales, sin ninguna. En el mejor de los casos, como han demostrado cuidadosos estudios estadísticos, los datos de los que disponen son escasos, sobre todo si se tiene en cuenta la rango-frecuencia (“ley de Zipf”).

Entra Tom Jones: “Has sido refutado. Sin prestar atención a tus hallazgos, los LLM que escanean cantidades astronómicas de datos pueden encontrar regularidades estadísticas que permiten simular los datos en los que han sido entrenados, produciendo algo que se parece bastante al comportamiento humano normal. Chatbots”.

Este caso difiere de los demás. En primer lugar, es real. Segundo, la gente no se ríe; de hecho, muchos se asombran. Tercero, a diferencia de los casos hipotéticos, los resultados reales distan mucho de lo que se afirma.

Estas consideraciones sacan a relucir un problema menor con respecto al entusiasmo actual por los LLM: su total absurdidad, como en los casos hipotéticos en los que lo hemos reconocido de inmediato. Pero hay problemas mucho más graves que la absurdidad.

Una es que los sistemas LLM están diseñados de tal manera que no pueden decirnos nada sobre el lenguaje, el aprendizaje u otros aspectos de la cognición, una cuestión de principio, irremediable. Duplique los terabytes de datos escaneados, añada otro billón de parámetros, utilice todavía más energía de California, y la

simulación del comportamiento mejorará, al tiempo que revelará más claramente el fracaso de principio en el planteamiento sobre cómo producir cualquier forma de comprensión. La razón es elemental: los sistemas funcionan igual de bien con lenguas imposibles, tales que los bebés no pueden adquirir, como con aquellas que estos adquieren rápidamente y casi por reflejo.

Es como si un biólogo dijera: “Tengo una nueva gran teoría de los organismos. Esta teoría enumera muchos que existen y muchos que es imposible que existan, y no puedo decirles nada sobre la distinción”.

De nuevo, nos reiríamos. O deberíamos.

No ríe Tom Jones, quien ahora se referirá a casos reales. Persistiendo en su radical alejamiento de la ciencia, Tom Jones responde: “¿Cómo puedes saber nada de esto hasta que no hayas investigado todas las lenguas?”. En este punto, el abandono de la ciencia normal resulta aún más claro. Por un argumento análogo podemos desechar la genética y la biología molecular, la teoría de la evolución y el resto de las ciencias biológicas, que no han estudiado más que las muestras de una ínfima fracción de organismos. Y por si fuera poco, podemos desechar toda la física. ¿Por qué creer en las leyes del movimiento? ¿Cuántos objetos se han observado realmente en movimiento?

Existe, además, la pequeña cuestión de la carga de la prueba. Quienes proponen una teoría tienen la responsabilidad de demostrar que tiene algún sentido, en este caso, demostrando que falla para las lenguas imposibles. No es responsabilidad de otros refutar la propuesta, aunque para el caso parece bastante fácil hacerlo.

Pasemos a la ciencia normal, donde las cosas se ponen interesantes. Incluso un solo ejemplo de adquisición del lenguaje puede aportar una rica visión de la distinción entre lenguas posibles e imposibles.

Las razones son claras y conocidas. Todo crecimiento y desarrollo, incluido lo que se denomina “aprendizaje”, es un proceso que parte de un estado del organismo y lo transforma paso a paso hacia etapas posteriores.

La adquisición del lenguaje es un proceso de este tipo. El estado inicial es la dotación biológica de la facultad del lenguaje, que obviamente existe, incluso aunque pueda ser, como algunos creen, una combinación particular de otras

capacidades. Eso es muy improbable por razones que se comprenden desde hace tiempo, pero no es relevante para lo que nos ocupa aquí, así que podemos dejarlo de lado. Es evidente que existe una dotación biológica para la facultad humana del lenguaje. Una mera obviedad.

La transición avanza hacia un estado relativamente estable, que no cambiará más que superficialmente: el conocimiento lingüístico. Los datos externos desencadenan y configuran parcialmente el proceso. Estudiando el estado alcanzado (conocimiento lingüístico) y los datos externos, podemos sacar conclusiones de gran alcance sobre el estado inicial, la dotación biológica que hace posible la adquisición del lenguaje. Las conclusiones sobre el estado inicial imponen una distinción entre lenguas posibles e imposibles. La distinción es válida para todos los que comparten el estado inicial: todos los humanos, hasta donde sabemos; no parece haber diferencias en la capacidad de adquirir el lenguaje entre los grupos humanos existentes.

Todo esto es ciencia normal, y ha conseguido muchos resultados.

La experimentación ha demostrado que el estado estable se obtiene sustancialmente muy pronto, hacia los tres o cuatro años de edad. También está bien establecido que la facultad del lenguaje tiene propiedades básicas específicas de los humanos, de ahí que sea una verdadera propiedad de especie: común a los grupos humanos y, en aspectos fundamentales, un atributo humano único.

En este relato esquemático se dejan muchas cosas fuera, sobre todo el papel de leyes de la naturaleza en el crecimiento y el desarrollo: en el caso de un sistema computacional como el lenguaje, los principios de eficiencia computacional. Pero ésta es la esencia de la cuestión. De nuevo, ciencia normal.

Es importante tener clara la distinción de Aristóteles entre posesión de conocimiento y uso del conocimiento (en terminología contemporánea, competencia y actuación). En el caso del lenguaje, el estado estable obtenido es la posesión de conocimientos, codificados en el cerebro. El sistema interno determina una serie ilimitada de expresiones estructuradas, cada una de las cuales podemos considerar que formula un pensamiento, cada una externalizable en algún sistema sensomotor, normalmente el sonido, aunque podría ser el signo o incluso (con dificultad) el tacto.

En el uso del conocimiento (“actuación”) se accede al sistema codificado internamente. La actuación incluye el uso interno del lenguaje en el pensamiento:

reflexión, planificación, recuerdo y mucho más. Desde el punto de vista estadístico, éste es, con mucho, el uso abrumador del lenguaje. Es inaccesible a la introspección, aunque podemos aprender mucho sobre él con los métodos normales de la ciencia, desde “fuera”, metafóricamente hablando. Lo que se denomina “habla interior” son, de hecho, fragmentos de lenguaje exteriorizado con el aparato articulatorio silenciado. Es sólo un reflejo remoto del uso interno del lenguaje; cuestiones importantes que no puedo tratar aquí.

Otras formas de uso del lenguaje son la percepción (análisis sintáctico) y la producción, esta última entraña propiedades que siguen siendo tan misteriosas como cuando Galileo y sus contemporáneos las contemplaban con asombro en los albores de la ciencia moderna.

El principal objetivo de la ciencia es descubrir el sistema interno, tanto en su estado inicial de la facultad humana del lenguaje como en las formas particulares que adopta en la adquisición. En la medida en que se comprenda este sistema interno, podremos proceder a investigar cómo entra en funcionamiento, interactuando con muchos otros factores que intervienen en el uso del lenguaje.

La obtención de datos acerca de la actuación ofrece evidencias sobre la naturaleza del sistema interno, en especial cuando se refina experimentalmente, algo habitual en los trabajos de campo. Pero incluso la recopilación más masiva de datos es necesariamente engañosa en aspectos cruciales. Se ciñe a lo que se produce normalmente, no al conocimiento del lenguaje codificado en el cerebro, el objeto primario que investigan quienes quieren comprender la naturaleza del lenguaje y su uso. Ese objeto interno determina una clase infinita de posibilidades que no se utilizarán en el comportamiento normal debido a factores irrelevantes para el lenguaje, como las limitaciones de la memoria a corto plazo; temas estudiados hace 60 años. Los datos observados también incluirán buena parte de lo que queda fuera del sistema codificado en el cerebro, a menudo el uso consciente del lenguaje de formas que violan las reglas con fines retóricos. Son obviedades conocidas por todos los investigadores de campo, que recurren a técnicas de recolección directa de datos con sujetos informantes, básicamente experimentos, para obtener un corpus refinado que excluya las restricciones irrelevantes y las expresiones desviadas. Lo mismo ocurre cuando los lingüistas se utilizan a sí mismos como fuentes de información, un procedimiento perfectamente razonable y normal, habitual en la historia de la psicología hasta la actualidad.

Si seguimos avanzando con la ciencia normal, descubrimos que los procesos y elementos internos de la lengua no pueden detectarse mediante la inspección de los fenómenos observados. A menudo, estos elementos ni siquiera aparecen en el habla (o la escritura), aunque sus efectos, a menudo sutiles, pueden detectarse. Esa es otra razón por la que restringirse a los fenómenos observados, como en los enfoques LLM, limita marcadamente la comprensión de los procesos internos que son objeto central de la investigación sobre la naturaleza del lenguaje, su adquisición y su uso. Pero eso no es relevante si se ha abandonado la preocupación por la ciencia y la comprensión en favor de otros objetivos.

Más en general, en las ciencias, durante milenios se ha llegado a conclusiones mediante experimentos -a menudo experimentos mentales-, siendo cada uno de ellos una abstracción radical de los fenómenos. Los experimentos se basan en la teoría y tratan de descartar los innumerables factores irrelevantes que intervienen en los fenómenos observados, como en la actuación lingüística. Todo esto es tan elemental que rara vez se discute. Y familiar. Como se ha señalado, la distinción básica se remonta a la distinción de Aristóteles entre posesión del conocimiento y uso del conocimiento. El primero es el objeto central de estudio. Los estudios secundarios (y bastante serios) investigan cómo sirve a la actuación el sistema de conocimiento almacenado internamente, junto con los muchos factores no lingüísticos que entran en lo que se observa directamente.

También podemos recordar una observación del biólogo evolutivo Theodosius Dobzhansky, famoso sobre todo por su trabajo con la *Drosophila*: cada especie es única, y los humanos son los más únicos de todos. Si nos interesa comprender qué clase de criaturas somos -siguiendo el mandato del Oráculo de Delfos de hace 2.500 años-, nos ocuparemos sobre todo de lo que hace que los humanos seamos los más únicos de todos, principalmente el lenguaje y el pensamiento, estrechamente entrelazados, como reconoce una rica tradición que se remonta a la Grecia clásica y la India. La mayor parte del comportamiento es bastante rutinario y, por tanto, hasta cierto punto predecible. Lo que realmente nos hace únicos es lo que no es rutinario, y lo encontramos, a veces mediante experimentos, a veces mediante la observación, desde los niños en condiciones normales hasta los grandes artistas y científicos.

Un último comentario al respecto. La sociedad lleva un siglo plagada de campañas masivas por parte de empresas para fomentar el desprecio por la ciencia, temas bien estudiados por Naomi Oreskes entre otros. Comenzó con compañías cuyos

productos eran letales: el plomo, el tabaco, el amianto y, más tarde, los combustibles fósiles. Sus motivos son entendibles. El objetivo de una empresa en una sociedad capitalista es el beneficio, no el bienestar humano. Es un hecho institucional: si no sigues el juego, te echan y te sustituye otro que sí lo haga.

Los departamentos de marketing de las empresas reconocieron muy pronto que sería un error negar las crecientes pruebas científicas de los efectos letales de sus productos. Eso sería fácilmente refutable. Mejor sembrar la duda, fomentar la incertidumbre, el desprecio por esos superexpertos de traje y corbata que nunca han pintado una casa, pero vienen de Washington a decirme que no use pintura con plomo, destruyendo mi negocio (un caso real, fácilmente multiplicable). Eso ha funcionado demasiado bien. Ahora mismo nos está llevando por el camino de la destrucción de la vida humana organizada en la Tierra.

En ambientes intelectuales, la crítica posmoderna de la ciencia, desmontada por Jean Bricmont y Alan Sokal, pero aún muy viva en algunos círculos, ha producido efectos similares.

Puede que sea una sugerencia poco amable, pero creo que es justo preguntarse si los Tom Jones y aquellos que repiten acríticamente e incluso amplifican sus descuidadas proclamas están contribuyendo a las mismas tendencias nefastas.

CJP: ChatGPT es un chatbot basado en el lenguaje natural que utiliza la inteligencia artificial para permitir conversaciones similares a las humanas. En un reciente artículo publicado en The New York Times, junto con otros dos autores, usted tachaba de bombo publicitario a los nuevos chatbots porque, sencillamente, no pueden igualar la competencia lingüística de los humanos. ¿No es posible, sin embargo, que las futuras innovaciones en IA produzcan proyectos de ingeniería que igualen y quizá incluso superen las capacidades humanas?

NC: El crédito por el artículo debería atribuirse al autor real, Jeffrey Watumull, un excelente matemático-lingüista-filósofo. Los dos coautores que figuran en la lista son asesores que están de acuerdo con el artículo, pero no lo han escrito.

Es cierto que los chatbots no pueden en principio igualar la competencia lingüística de los humanos, por las razones que hemos repetido antes. Su diseño básico les impide alcanzar la condición mínima de adecuación para una teoría del lenguaje humano: distinguir los lenguajes posibles de los imposibles. Dado que se trata de

una propiedad del diseño, no puede ser superada por futuras innovaciones en este tipo de IA. Sin embargo, es muy posible que futuros proyectos de ingeniería igualen e incluso superen las capacidades humanas, si nos referimos a la capacidad humana de actuación [performance] en el uso del lenguaje. Como se ha señalado más arriba, algunos ya lo han hecho desde hace tiempo: las calculadoras automáticas, por ejemplo. Y lo que es más interesante, como ya se ha mencionado, insectos con cerebros minúsculos superan las capacidades humanas entendidas como competencia.

CJP: En el citado artículo también se observaba que los proyectos de IA actuales no poseen una facultad moral humana. ¿Este hecho tan obvio hace que los robots de IA sean una amenaza menor para la raza humana? Creo que se puede argumentar que los hace incluso más peligrosos.

NC: En efecto, es un hecho evidente, entendiendo la “facultad moral” en sentido amplio. A menos que se controle cuidadosamente, la ingeniería de la IA puede plantear graves amenazas. Supongamos, por ejemplo, que se automatizara el cuidado de los pacientes. Los inevitables errores que el juicio humano es capaz de subsanar podrían dar lugar a una historia de terror. O supongamos que se eliminara a los humanos de la evaluación de las amenazas determinadas por los sistemas automatizados de defensa antimisiles. Como nos informa un registro histórico estremecedor, eso sería el fin de la civilización humana.

CJP: Organismos reguladores y fuerzas de seguridad en Europa están mostrando su preocupación por la propagación de ChatGPT, mientras que un texto legislativo de la Unión Europea presentado recientemente intenta hacer frente a la IA clasificando dichas herramientas según su nivel de riesgo percibido. ¿Está de acuerdo con los que temen que ChatGPT suponga una grave amenaza pública? Además, ¿cree realmente que se puede detener el desarrollo de herramientas de IA hasta que se introduzcan salvaguardias?

NC: Puedo simpatizar muy fácilmente con los esfuerzos por intentar controlar las amenazas que plantea la tecnología avanzada, incluido este caso. Sin embargo, soy escéptico sobre la posibilidad de hacerlo. Sospecho que el genio está fuera de la botella. Es probable que los agentes maliciosos -institucionales o individuales- puedan encontrar formas de eludir las salvaguardias. Estas sospechas no son, por supuesto, razón para no intentarlo y para mantener la guardia.

[LEER EL ARTÍCULO ORIGINAL PULSANDO AQUÍ](#)

Fotografía: bloghemia

Fecha de creación

2023/07/03