

Inteligencia Artificial: ¿Qué tan Inteligente?

Una perspectiva crítica sobre
la inteligencia “humana”



Autor: Diego Llontop Céspedes

Facultad de Psicología

Universidad de Lima

doi:<https://doi.org/10.26439/piedepagina2023.n009.6461>

Hace poco, durante una capacitación sobre los potenciales usos educativos del ChatGPT, se mencionó la posibilidad de incluir dicha herramienta como una fuente más en la lista de referencias, en caso esta haya sido utilizada por los estudiantes para escribir un texto. Mi primera reacción frente a dicha posibilidad fue la negación. El ChatGPT es un motor estadístico de búsqueda y combinación de diferentes textos existentes en la web, una búsqueda mecánica y acrítica. Una simple combinación de textos en función de potentes algoritmos diseñados, estos sí, por inteligencias humanas. La conclusión inicial era la siguiente: no se puede equiparar los esfuerzos investigativos de autores humanos con un motor de texto.

Sin embargo, ¿qué nos hace pensar que el cerebro humano no es un potente generador estadístico de texto, acrítico, sin control centralizado? Solo una simple combinación de subprocesos neuronales (Dennett, 1991) que llegan a una conclusión de manera eficiente, pero sin la menor participación por parte del supuesto usuario, es decir, uno mismo. Mientras escribo este texto, me pregunto qué tanta participación tengo en las ideas que van surgiendo, en los autores que van apareciendo vinculados a estas ideas que escribo. ¿Vienen a mi mente de manera controlada o espontánea? Si bien parece haber un tipo de control y regulación durante el proceso, la génesis de las ideas parece estar fuera de mi control e, incluso, de mi conciencia.

El ChatGPT –Generative Pre-trained Transformer o Transformador Generativo Preentrenado– es una red neural artificial. Un tipo de inteligencia que replica, en algún grado, el procesamiento de las redes neurales reales de los cerebros biológicos y que es posible ser entrenada –a través del llamado *deep learning* o *aprendizaje profundo*– para llegar a los resultados esperados. En el caso del ChatGPT, estos resultados son básicamente la producción de texto en función de la demanda de los usuarios.

Por ejemplo, se le puede pedir al sistema que haga una reseña de la producción literaria peruana de los últimos diez años. El *input* será recibido por la primera capa de “neuronas”, unidades inferenciales mecánicas que transmiten impulsos eléctricos formando un circuito. Esta primera capa está conectada con otras capas que filtran la información hasta llegar al resultado esperado, es decir, una predicción acorde al *input* recibido reflejado en un texto. Previamente, el programa ha sido entrenado masivamente durante meses de trabajo. Según Cooper (2021), para preparar el circuito se utilizaron 175 000 millones de parámetros y 45 *terabytes* de información textual extraída de diferentes fuentes de Internet y libros. Un *terabyte* es equivalente a un millón de *megabytes*. Un libro promedio (200 páginas), en formato PDF, pesa alrededor de un *megabyte*. Estamos hablando de los sueños de Jorge Luis Borges llevados al máximo extremo con esta suerte de biblioteca de Babel electrónica.

Hablando de Borges, Noam Chomsky, Roberts y Watumull (2023) utilizan una idea de este autor para descalificar al programa en sus supuestas pretensiones de “igualar la inteligencia humana”. Según el argentino citado por los autores, vivir en momentos de alto riesgo supone experimentar tanto la tragedia como la comedia. Chomsky et al. (2023) consideran que el ChatGPT representa este escenario, puesto que los supuestos logros intelectuales del programa no alcanzan ni siquiera a lo que podríamos llamar “conducta inteligente”: el programa es incompetente en términos productivos, pues es incapaz de regular o inhibir sus respuestas, al ser estas tanto verdaderas como falsas indistintamente. En esta línea, los autores indican que esto aplica también para sus “decisiones morales” emitiendo moralidad e inmoralidad a discreción, sin ningún atisbo de responsabilidad o cálculo de consecuencias. Para

los autores, la marca de la verdadera inteligencia supone lograr explicaciones en base al juicio crítico de implicar causalidades y descartar lo imposible en base a dicho juicio. El programa no es más que una máquina estadística que emite respuestas en función de su base de datos sin discriminar o elegir “racionalmente”. Esto, concluyen los autores, nos lleva a poder “llorar” o “reír” frente a los supuestos “grandes logros” del programa (Chomsky et al., 2023).

Descartando la evidente falacia de espantapájaros incluida en el artículo en mención –la función del programa es emitir texto, no verdad o moralidad–, vemos la expresión hasta cierto punto conmovedora del escepticismo humano frente a cualquier posibilidad de que un sistema producido artificialmente pueda igualar sus propias capacidades. Este tipo de argumentos no son nuevos, pues ocurren desde mediados del siglo XX con el surgimiento de los sistemas computacionales y la desafiante posición funcionalista en el campo de la filosofía de la mente.

El funcionalismo ataca el chauvinismo humano de atribuirse a sí mismo una capacidad intelectual exclusiva. Para este chauvinismo, el cobre y la silicón de las computadoras nunca podrán igualar procesos biológicos cerebrales, los cuales aún están bastante lejos de nuestra comprensión o, en su defecto, se asume que un mecanismo artificial no podrá igualar jamás un proceso mental “abstracto” o “subjetivo”, es decir, la vivencia interna de nuestra propia conciencia de la cual carecen los mecanismos que producimos los seres humanos prácticamente por definición. El funcionalista, en contraste, no atiende a la base material que produce la inteligencia o a una supuesta subjetividad humana, sino a lo que, efectivamente, puede hacer el sistema. Un reloj de cuarzo y una persona que dan la misma hora son funcionalmente idénticos, independientemente de sus soportes materiales o el hecho de que el reloj muestre el *output* en la pantalla y la persona exprese su respuesta hablando. Independientemente de que podamos saber si el reloj tiene una experiencia subjetiva.

El argumento funcionalista ataca al argumento chomskiano en lo referente a apelar a la cualidad interior de los procesos mentales que emiten la respuesta en humanos, en contraste con sistemas artificiales. Para el funcionalismo, no es importante lo que ocurre dentro, con tal que

la respuesta sea la que se espera del sistema. En contraste, para Chomsky et al. (2023), el ChatGPT es un mero cálculo estadístico ciego, masivo, pero torpe, comparado con la inteligencia humana, un “sorpresivamente eficiente e incluso elegante sistema que opera con pequeñas cantidades de información; no busca inferir correlaciones brutas entre puntos de información sino crear explicaciones”. Sin embargo, cabe preguntarse lo siguiente: ¿no se podría llegar a explicaciones exitosas en función de “correlaciones brutas” ajustando adecuadamente los parámetros de aprendizaje del sistema si ese fuera el objetivo? ¿Es importante la elegancia o la economía informacional al llegar a una respuesta o lo es la respuesta en sí?

En efecto, si comparamos los sistemas que podemos crear artificialmente en la actualidad con la biología cerebral humana, la distancia es muy notoria. Una red neural artificial como ChatGPT tiene, a lo sumo, una interconexión de mil neuronas o unidades de procesamiento (Kerthana, 2021). Un cerebro promedio humano cuenta con 86 000 millones de neuronas (Alonso, 2018). El solo pensar en la cantidad de interconexiones de un sistema de este tipo nos hace entender cómo se puede pensar en el cerebro humano como la máquina más compleja del universo conocido y dejar muy a la zaga los esfuerzos de ingenieros por producir un sistema similar. Considerando esta complejidad, el Proyecto Conectoma Humano (NIH Blueprint for Neuroscience Research, 2009) busca mapear áreas en el cerebro en contraste con neuronas individuales y sus, prácticamente, infinitas interconexiones.

Algunos otros piensan que se podría compensar la desventaja de procesamiento al utilizar la mayor velocidad de computadoras cuánticas para simular un sistema, por lo menos, parecido al cerebro humano. Está por verse hasta dónde puede llegar este tipo de tecnologías. Lo real es que incluso la metáfora científica de “procesador de información” —el modelaje del cerebro en función de una computadora eléctrica— queda corto dado el carácter biológico del cerebro, una máquina electroquímica. El cerebro humano no funciona en base a unidades que expresan un código simplemente binario, de transmisión discreta de impulsos eléctricos como en los transistores (Wissell, 1999). Al parecer, las neuronas funcionan en base a grados, pudiendo demorar o acelerar la



La IA ha reabierto el debate conceptual sobre la inteligencia.

respuesta del impulso y transmitir la información en base a la latencia del impulso o en base a su duración absoluta. Esta variabilidad impide modelar satisfactoriamente el cerebro, es decir, entender cómo funciona la codificación y ser capaz de replicarla (Churchland, 2005).

No obstante, incluso en estos radicales contrastes, no podemos dejar de pensar en el cerebro humano como una máquina infinitamente más compleja que cualquiera que podamos diseñar y construir en la actualidad, pero, inevitablemente, una máquina. Si bien la noción de inteligencia que maneja Chomsky —la capacidad de llegar a explicaciones y por lo mismo, descartar las que no pueden serlo— resulta justificada y razonable. ¿Qué nos hace pensar que una máquina artificial no pueda llegar a realizar estos procesos? Los mismos autores deslizan la posibilidad de que esto pueda ser posible en un futuro, con lo cual no niegan de plano que la naturaleza intelectual humana no pueda ser replicada.

Además de las evidentes limitaciones fácticas mencionadas previamente, podemos ubicar el escepticismo sobre las posibilidades de la inteligencia artificial en más de una circunstancia. Una de ellas podría ser el miedo a lo desconocido, del cual ya han hablado notorios científicos como

Stephen Hawking o celebridades como Elon Musk: la visión apocalíptica en la cual las máquinas cobran conciencia de sí mismas y deciden dejar de atender a las ordenes de sus diseñadores e ir en contra de ellos. Dejando a un lado esta circunstancia, pasaremos a desarrollar dos que considero esenciales.

La primera tiene que ver con el aspecto que acabamos de mencionar, la *conciencia de uno mismo*. La tradición en la cual dicha posición se sustenta es de larga data, llega por lo menos hasta Descartes (2002). Fue el filósofo francés el primero en darle una relevancia absoluta en términos epistemológicos a la información asequible “interiormente” sin la contribución de sus sentidos. La evidencia del pensar, prueba perfecta dada su supuesta indubitabilidad interior, le permite llegar a la primera convicción absoluta de su filosofía: su propia existencia. O, en palabras del filósofo: “Pienso, luego existo”.

El problema aquí –notado de forma resaltante por Ryle (2005), Dennett (1991) y Damasio (1994) desde un ángulo diferente– es darle tanta preponderancia a la racionalidad interior humana. Esto nos lleva a sobredimensionar los datos que podemos recolectar

desde dentro y su potencial informativo o explicativo. Estipular la total superioridad de los datos subjetivos conlleva el riesgo de suponer que estas intuiciones interiores deben determinar nuestra teorización acerca de lo mental/cerebral. Incluso, en el caso de Descartes, llegar a pensar que estos datos subjetivos tienen una naturaleza diferente a la física, dado que llegamos a ellos a través de una vía no empírica, inmaterial. Este es el camino de entender la inteligencia humana como algo de distinta naturaleza y, por lo tanto, uno de los soportes del chauvinismo objeto de la denuncia funcionalista.

Esta circunstancia está íntimamente vinculada con el segundo condicionante que podríamos mencionar: las dificultades para llevar esta experiencia subjetiva –supuestamente de naturaleza tan excéntrica– al lenguaje objetivo de la ciencia y, por ende, de desentrañar la “naturaleza” de lo mental en dichos términos. Pero este parece ser un modo inadecuado de definir el problema.

El problema “duro” (Chalmers, 1996) en la comprensión de la conciencia humana puede ser perfectamente un error conceptual, que arrastra esta tendencia a sobredimensionar nuestra vivencia mental interior queriendo poner los



La IA empieza a exhibir resultados indistinguibles de la inteligencia humana.

datos producidos de esta forma al mismo nivel objetivo del discurso científico. En contraste, para neurocientíficos muy notorios actualmente, como Stanislas Dehaene (2014), la vivencia subjetiva no puede formar parte de la perspectiva neurocientífica a menos que sea vista como dato contrastable, por ejemplo, con imágenes computarizadas del cerebro. Para dicho autor, sería un error asumir que la vivencia interior supone potenciales conceptos asimilables a nuestras teorías sobre lo mental.

En buena cuenta, el calificar de “simulación” a la inteligencia computarizada parece arrastrar no solo la evidente diferencia entre un sistema artificial y un sistema natural, sino ciertos prejuicios filosóficos profundamente enraizados en la idiosincrasia humana. A pesar de las limitaciones tecnológicas actuales referidas al diseño de inteligencia artificial, podemos tener en mente la lección funcionalista: nadie puede estar en la mente de nadie y esto parece aplicar también para las computadoras. En términos prácticos, la inteligencia no se mide en función de lo que ocurre adentro, sino en función de los resultados obtenidos. En otras palabras, y tal como lo sugiere la interesante película *Ex Machina* (2014), si parece que tiene conciencia, quiere decir que la tiene. O, en términos más simples: “Si lo ves, es cierto”.

REFERENCIAS

- Alonso, J. (2018). *Historia del cerebro. Una historia de la humanidad*. Guadalmazán.
- Chalmers, D. (1996). *The conscious mind. In search of a fundamental theory*. Oxford University Press. https://personal.lse.ac.uk/ROBERT49/teaching/ph103/pdf/Chalmers_The_Conscious_Mind.pdf
- Chomsky, N., Roberts, I., & Watumull, J. (2023, 8 de marzo). Noam Chomsky: the false promise of ChatGPT. *Portside*. <https://portside.org/2023-03-08/noam-chomsky-false-promise-chatgpt>
- Cooper, K. (2021). *OpenAI GPT-3: everything you need to know*. Springboard. <https://www.springboard.com/blog/data-science/machine-learning-gpt-3-open-ai/>
- Damasio, A. (1994). El error de Descartes, la emoción, la razón y el cerebro humano. *Panamerican Journal of Neuropsychology*, 5(2), 173-178. <https://www.redalyc.org/pdf/4396/439642488006.pdf>
- Dehaene, S. (2014). *Consciousness and the brain. Deciphering how the brain codes our thoughts*. Penguin Books.
- Dennett, D. (1991). *Consciousness explained*. Back Bay Books.
- Descartes, R. (2002). Meditations on first philosophy. En D. Chalmers (Ed.), *Philosophy of mind, classical and contemporary readings* (pp. 10-21). Oxford University Press. <http://course.sdu.edu.cn/G2S/eWebEditor/uploadfile/20140227112306002.pdf>
- Kerthana, V. (2021). Artificial neural network, its inspiration and the working mechanism. *Analytics Vidhya*. <https://www.analyticsvidhya.com/blog/2021/04/artificial-neural-network-its-inspiration-and-the-working-mechanism/>
- NIH Blueprint for Neuroscience Research. (2009). *Connectome program*. <https://neuroscienceblueprint.nih.gov/human-connectome/connectome-programs>
- Ryle, G. (2005). *El Concepto de lo mental*. Paidós. <https://cideps.com/wp-content/uploads/2018/12/Ryle-El-concepto-de-lo-mental.pdf>
- Wissell, L. (1999, 21 de octubre). How does a logic gate in a microchip work? A gate seems like a device that must swing open and closed, yet microchips are etched onto silicon wafers that have no moving parts. So how can the gate open and close?. *Scientific American*. <https://www.scientificamerican.com/article/how-does-a-logic-gate-in/>