

Artificial. La nueva Inteligencia y el contorno de la humano, de Mariano Sigman y Santiago Bilinkins

Artificial. La nueva Inteligencia y el contorno de la humano, by Mariano Sigman and Santiago Bilinkins

Danilo Miranda Baires

Universidad Centroamericana José Simeón Cañas

El Salvador

dmiranda@uca.edu.sv

ORCID: 0000-0003-4881-3740

La Inteligencia Artificial (IA) se ha posicionado en el ojo del huracán tecnocapitalista cuando está completándose el primer cuarto del siglo XXI. Las narrativas sobre esta revolución tecnológica oscilan entre las que abrazan sin cuestionar la radiante utopía por venir hasta las que avizoran catástrofes distópicas. La evidencia empírica apoya ambas proyecciones y tal como en tecnologías precedentes, engendra posibilidades tanto para contribuir a resolver grandes problemas que aquejan a la humanidad, pero también como amenaza igualmente a la misma especie que la creó, al igual que el doctor Frankenstein y su criatura, o el aprendiz de brujo, incapaces de controlar a las fuerzas que desataron.

Una excelente introducción a un tema tan complejo es *Artificial. La nueva inteligencia y el contorno de lo humano*, del neurocientífico Mariano Sigman y el tecnólogo Santiago Bilinkins, publicado en Barcelona por Penguin Random House a través de su sello Debate en 2023. Como un valioso ejercicio de divulgación científica, los autores se valen de una exposición clara procurando que los múltiples tecnicismos utilizados no oscurezcan el texto al gran público. Con todo, el libro no carece de rigor y ofrece una perspectiva interdisciplinaria que busca abordar la cuestión desde diversos campos del saber. Sus líneas dialogan con una diversidad de exponentes de la ciencia, la tecnología, la política, la economía, la filosofía y la literatura para construir un relato multidimensional de la IA. Los autores estructuran su obra en un breve prólogo, once capítulos y un epílogo.

El libro tiene como punto de partida el origen de la inteligencia artificial y reconstruyen su proceso sociohistórico. Esto incluye a pioneros como el británico Alan Turing y su máquina Enigma, que formó parte de un proyecto para descifrar códigos nazis durante la Segunda Guerra Mundial. También hace referencia a Eliza, un bot conversacional programado por el profesor de informática Joseph Weizenbaum en 1966 para ser usado en psicoterapia. En plena Guerra Fría, cuando la carrera atómica y espacial eran el centro de la atención, la IA ocupaba una agenda marginal en reductos académicos, donde físicos, matemáticos y neurofisiológicos trabajaban en la noción de "redes neuronales", que a la postre se convertirían en un motor de la IA. De ahí se desprendió el *deep learning* o aprendizaje profundo, que cambió la manera de aprender de las máquinas, que ya no siguen a pie juntillas las instrucciones humanas sino que pasan por un proceso de entrenamiento que se apoya en la retroalimentación humana para explorar los patrones de conexiones neuronales que las hagan más efectivas y de esta manera aprenden



por su cuenta: “La clave es que el *input* humano a la red neuronal es definir la función de valor, indicándole qué es lo que debe maximizar. [...] Una vez que le decimos el qué, la IA encuentra el cómo” (Sigman y Bilinkis, 2023, p. 27).

Otros hitos que potencian esta tecnología son el desarrollo de hardware e internet. En el primero, cobró importancia la unidad de procesamiento de gráficos (GPU, por sus siglas en inglés) que aligeró el trabajo del procesador central y posibilitó trabajar paralelamente en varias instancias de una tarea varias cosas. El segundo alimentó a las redes neuronales de una enorme cantidad de datos e hizo que la IA explotara. Un momento decisivo tuvo lugar en 2015, cuando la máquina AlphaGo derrotó al coreano Lee Se-dol, campeón de Go, en una pérdida más impactante que la sufrida años antes por Gary Kasparov ante DeepBlue. AlphaZero superó a su predecesora, dominando tanto el ajedrez como el Go. Aprendió jugando contra sí misma.

Los autores describen los acontecimientos tecnológicos han dado paso a una nueva era y la manera en la cual se desarrollaron “máquinas capaces de abstraer, calcular, general ideas propias y originales, concebir objetos y, en última instancia, de conversar” (Sigman y Bilinkis, 2023, p. 47). Los autores se preguntan cómo se construye una inteligencia y cómo trasciende de la simple imitación y ponen de relieve la importancia que cobra educar a una máquina: ¿cómo descubre una red neuronal las palabras, la ideas y el estilo de un escritor clásico, o la jerga usada por un grupo social específico? ¿Cómo se entrena una mente artificial? En estas preguntas reside un elemento clave para la comprensión de estas tecnologías, su autonomía. Hay una intuición común según la cual las máquinas no pueden superar la inteligencia humana

[...] porque justamente fueron pensadas, diseñadas y entrenadas por personas. Este razonamiento esconde una gran falacia que ya visitamos: sabemos que el discípulo puede superar al maestro. Las redes neuronales pueden identificar atributos que son indistinguibles para nosotros y así realizar tareas con habilidad sobrehumana. Pueden descubrir y aprender cosas que nadie les ha enseñado (Sigman y Bilinkis, 2023, p. 55)

Otra observación importante que el libro plantea es que la IA se usaba para cuestiones eminentemente operativas. Pero con los llamados *Large Language Models* (LLM) estas tecnologías no sólo irrumpen en tareas mecánicas sino en actividades que requieren creatividad e ingenio. A los humanos les inquieta que una máquina piense mejor porque la inteligencia es una facultad definitoria del Homo Sapiens. El texto sugiere que al no ser una especie rápida, fuerte y resiliente, la inteligencia es la herramienta con la que los humanos han gobernado el mundo y se han impuesto sobre otras formas de vida, hasta ahora. Para Sigman y Bilinkis “hacemos bien en temer lo que pueda pasar cuando alguien o algo nos supere, y pueda ser quien decida si andamos sueltos, con correas o enjaulados” (Sigman y Bilinkis, 2023, p. 60). Estos escenarios catastróficos ya han sido advertidos por expertos que consideran que la IA pueda llegar a convertirse en una amenaza para la humanidad (The Future of Life Institute, 2023; Yudkowsky, 2023). Con todo, todavía en el presente los humanos tienen cierto margen de maniobra al definir la función de valor que orientará el comportamiento de las máquinas. Pero esta impronta también incluye sesgos e intereses, la tecnología no es neutral.

Los autores dedican un espacio importante a la reflexión sobre la IA conversacional, como el ChatGPT. La entrada de texto que inicia la interacción es conocida como *prompt*, y su calidad “es la clave para determinar el valor del resultado que obtengamos. Por lo tanto, aprender a dar buenas instrucciones se convertirá pronto en una habilidad fundamental” (Sigman y Bilinkis, 2023, p. 68). La pregunta o entrada que se haga puede generar respuestas superficiales o interesantes. Esta IA puede identificar relaciones y conexiones entre enfoques que a primera vista no parecen vínculo y tiene también la facilidad para ofrecer matices, comparaciones y miradas que no es sencillo imaginar. Más allá de lo anterior, para que la respuesta del *prompt* sea efectiva y personalizado “no basta con saber qué y cómo preguntarle [...] cuanto más información le proporcionamos más precisa y adecuada a nuestra necesidad será la respuesta” (Sigman y Bilinkis, 2023, p. 73). Ahora bien, conversar con IA conversacionales tiene vicios y virtudes. Con respecto a los primeros,

en primer lugar, los LLM tienen una alta efectividad en tareas relacionadas con el lenguaje y la atención; en segundo, el chat traduce muy bien y en tercero, las IA basadas en *transformers* son muy buenas para la programación. Entre sus vicios, el libro hace referencia a las “alucinaciones” de la IA, cuando se inventan lo que no saben; también a que los chats son pre-entrenados y esto supone que no se actualiza más allá de la fecha de entrenamiento.

Otro tema que tuvo una atención especial es uno de los campos donde se avizora una transformación radical: la educación. El capítulo que la aborda en exclusiva lleva el nombre de “terremoto educativo”, tal es la magnitud de los cambios esperados en la práctica educativa. Una de las principales preocupaciones radica en lo que el texto denomina “sedentarismo cognitivo”. Debido a que las personas pierden capacidades esenciales, es necesario no perder de vista que “hay muchos pilares básicos de la cognición que son vitales para el pensamiento [...] la capacidad de concentrarnos, la competencia lectora, el bien uso del lenguaje y el pensamiento lógico y matemático” (Sigman y Bilinkis, 2023, p. 105). Para los autores es clave no dejar enteramente en la tecnología esas habilidades que las personas no deberían perder, aquellas que proporcionan autonomía en diversas áreas del pensamiento y que resulta pertinente conservar ante un futuro tan incierto. En tanto todo aquello que no se utiliza se pierde es ineludible la cuestión sobre qué habilidades necesitamos conservar y cuáles podemos dejar a las máquinas. Según el texto se pueden aprovechar las potencialidades de la IA en tres aspectos: para que produzca material desde la perspectiva de algún personaje en particular, para vincular la información relevante con los intereses de formas creativas e innovadoras, y finalmente, se puede recuperar el valor de hacer buenas preguntas.

Otro desafío de gran envergadura es el impacto de la IA en el mundo del trabajo. El desempleo tecnológico es una realidad preocupante, si bien quienes son más optimistas se centran en aquellos empleos que se crearán en el marco de la IA. La IA tiene, además, el potencial de usar su capacidad en el procesamiento de datos para ayudar a las organizaciones a encontrar soluciones a problemas de alta complejidad. En el escrito se plantean las siguientes interrogantes:

¿Qué actividades actuales dejarán de estar en manos de seres humanos? ¿Qué nuevos empleos sustituirán las actividades que ya no se realicen? ¿Quién puede asegurar que la cantidad de puestos que se creen serán suficientes para compensar los que se destruyan? ¿Cómo afectará este proceso a los salarios? ¿Vamos hacia la utopía tan anhelada de los antiguos griegos de liberarnos finalmente de todos los menesteres elementales de la vida para dedicarnos plenamente al ejercicio de la virtud? ¿O, por el contrario, nos dirigimos irremediamente hacia una distopía poscapitalista con desempleo estructural masivo e incremento de los niveles de pobreza y desigualdad? (Sigman y Bilinkis, 2023, p. 120)

Pese a que la irrupción de la tecnología en el mundo del trabajo no es algo nuevo, los autores consideran que la IA puede traer transformaciones cualitativas diferentes a las previas. Será un impacto “que pegará en lo más hondo de la humanidad, en un sentido ontológico” (Sigman y Bilinkis, 2023, p. 123). Tiene que ver con interrogantes sobre para qué los humanos están aquí, qué vienen a hacer, qué presente quieren vivir y qué futuro están construyendo.

Otro campo de aplicación de la tecnología que me ocupa en este texto, es la psicología, que como ciencia o como terapia, “está en la esencia misma de lo humano” (Sigman y Bilinkis, 2023, p. 141). Desde Eliza, la IA conversacional podría tener aplicaciones útiles en este campo, por ejemplo, como en el diagnóstico temprano de problemas de salud mental. Sin embargo, como en los campos anteriores, también tiene un lado B con su permanente exigencia para que se le comparta información sensible y las estrategias algorítmicas de las plataformas para captar la atención, promover los consumos triviales o tóxicos, la obsesión por la belleza física o la popularidad, o la falsa ilusión de estar conectados, pueden ser perjudiciales y traer efectos colaterales: “el rápido aumento de los trastornos de ansiedad como los ataques de pánico los desórdenes alimenticios y la depresión especialmente entre los jóvenes nos llevan a pensar que en interacción con la idea tenemos mucho que ganar pero también que perder” (Sigman y Bilinkis, 2023, p. 152)

Para el libro, la creciente interacción cotidiana con este tipo de máquinas conlleva errores, pero también las crecientes tensiones geopolíticas y económicas pueden privilegiar un uso potencialmente destructivo de la IA, como puede avizorarse en el desarrollo de tecnología militar, que enfrentaría a humanos contra humanos, apoyados con sus respectivas máquinas. Pero también hay un riesgo inherente en que el eje del conflicto futuro se traslade a humanos versus máquinas, sobre todo con el advenimiento de una inteligencia artificial general, que daría un poder sobrehumano del cual nadie garantiza que será respetuoso de las literarias "leyes de la robótica", postuladas por Isaac Asimov y citadas por los autores.

Se me escapan muchos más elementos de esta interesante obra. Sirva esta breve reseña como pretexto para debatir sobre una cuestión cuyo abordaje es ya ineludible. Convendría aprovechar los recursos tecnológicos disponibles, hasta cierto punto, y sin dejar de lado el examen crítico de su uso. Este es apenas un libro en el mar de publicaciones sobre el tema. Uno escrito en un lenguaje asequible y ameno que puede ser muy valioso para una primera aproximación.

Referencias bibliográficas

- Sigman, M. y Bilinkis, S. (2023). *Artificial. La nueva Inteligencia y el contorno de lo humano*. Debate.
- The Future of Life Institute (2023). Pause Giant AI Experiments: An Open Letter. *Future of Life Institute*. <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>
- Yudkowsky, E.(2023, 29 de marzo) Pausing AI Developments Isn't Enough. We Need to Shut it All Down. *Time*. <https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/>