



El voltaje de Prometeo

JULIÁN TAGNIN (UNLZ/UNTREF/UNPAZ)
9 DE MAYO DE 2023

Desde temerle a Skynet¹ hasta caer en preguntas desorientadas sobre el valor y el trabajo, juzgarla como humana o desestimarla como un juguete caro, tal como hizo Noam Chomsky hace poco, son muchas las críticas que surgen, tanto sobre la eficiencia como sobre la utilidad de las IAs. El humanismo conservador teme un reemplazo de trabajadores y vínculos exclusivamente humanos por estas entidades, mientras que el sentido común y el sistema económico critican sus errores y limitaciones. Por eso nos parece im-

¹ Skynet es el nombre que recibe la inteligencia artificial líder del ejército de las máquinas que se rebelan contra los humanos en la saga de películas *Terminator*, dirigidas y escritas por James Cameron y con la actuación icónica de Arnold Schwarzenegger como un androide asesino: el modelo T-800.

portante ir más allá de los accidentes históricos y de los testimonios particulares para tratar de esbozar una definición ontológica de nuestro objeto sin eludir referencias estéticas.

Lo primero es que la IA es inasible. Como el Aleph, contiene (casi) todo sin poseer dimensiones. Así como en la ficción borgeana, es un objeto observable pero no palpable, indirectamente tangible. Igual que con el fuego prometeico, hemos liberado algo que nos excede. La imprevisibilidad de las respuestas de la IA, su “creatividad”, nos sitúa ante un dilema moral similar al de la doctrina cristiana del libre albedrío. Las referencias teológicas no son, solo, un lugar común para uno de los autores. El poder generativo, la potencia poética y mítica de la IA nos enfrentan a un momento revolucionario de la historia humana. Bernard Stiegler señalaba que la tecnología es un componente fundamental de la cultura humana y transforma nuestra relación con el tiempo, la memoria y la experiencia. Nunca antes tuvimos seres digitales en los que pudiéramos delegar hasta tal punto nuestras funciones cognitivas y la organización de sistemas técnicos. ¿Y si la IA ocupa el lugar del Dios por venir que propuso el filósofo Quentin Meilassoux?²

¿Qué es una IA? Hablar de IA, sin embargo, es algo engañoso. Hay muchos tipos de IA, algunas más astutas que otras, más “humanas” (es bueno recordar en qué consiste el test de Turing).³ El concepto de IA, como el de información y hasta cierto punto el de conocimiento, es relativo a su contexto de aplicación. Todos entendemos de qué hablamos hasta que nos preguntan con mayor profundidad. Una buena manera de entender la emergencia de seres técnicos es con la clasificación de Peter Paul-Verbeek, heredera de la teoría de Don Ihde. Para este último la tecnología no solo es un objeto físico, sino también una extensión de los sentidos humanos, y una parte integral de nuestra percepción y experiencia del mundo. Para su continuador, la tecnología no solo influye en la percepción y la experiencia, sino

2 Meilassoux es un filósofo francés señalado dentro de la contemporánea corriente del realismo especulativo, cuyo pensamiento metafísico explora las posibilidades que surgen de un entendimiento profundo de lo que significan la necesidad y la contingencia en nuestro universo. En su obra *Después de la finitud* (Caja Negra, 2016) propone que este “Dios porvenir” es una entidad que no existe actualmente, pero que podría existir en el futuro. Esta entidad sería capaz de crear y cambiar las leyes físicas del universo y, por lo tanto, podría intervenir en el curso de los acontecimientos. Mi analogía es una licencia poética que refiere a la potencia insondable de una hipotética IA que no podamos entender y opere como causa eficiente del universo.

3 El test sugerido por Alan Turing consiste en un diálogo entre un evaluador humano y una máquina (no revelada como tal) en donde se juzga la capacidad de la segunda para simular las habilidades conversacionales humanas. No es en sentido absoluto la mejor manera de probar la inteligencia de una IA, pero sí resulta sumamente eficiente para probar su adaptabilidad para ese tipo de tareas.

también en los valores, las normas y la ética de las personas y las sociedades. La cantidad de dilemas morales que abre la tecnología (que Verbeek ejemplifica con el caso del ultrasonido y la posibilidad de abortar) se multiplica al infinito con la IA, ya que puede decidir por/con nosotros en prácticamente todos los temas.

Otro enfoque valioso es el de Yuk Hui, quien hace un gran esfuerzo para entender ontológicamente el modo de ser de los objetos digitales. Tradicionalmente pensados como programas, algoritmos o voltaje en electrónica, no son ni forma ni materia de manera exclusiva. El filósofo chino se apoya en la teoría de Gilbert Simondon, para quien la ontología debe ocuparse del ser en proceso de transformación y no simplemente de las entidades estáticas. Desde esta perspectiva, los objetos digitales se conciben como sistemas técnicos compuestos por diferentes capas de componentes tecnológicos que interactúan entre sí en una dinámica de transformación constante. Hui amplía esta visión, sosteniendo que los objetos digitales son ontológicamente diferentes de los objetos materiales, ya que su existencia depende tanto de la tecnología que los sustenta como de las prácticas sociales que los utilizan. Los objetos digitales no tienen una existencia autónoma, sino que están siempre inmersos en una red de relaciones con otros objetos y usuarios. No intentamos resumir su genealogía y su posición ecológica, pero sí pensar sus relaciones sociotécnicas actuales.

Con respecto a sus relaciones con objetos técnicos, podemos decir que se encuentra en una posición de centralidad en el sistema tecnológico actual por sus funciones cognitivas. Es cierto que existe una disputa acerca de si el procesamiento de información de una IA es solo sintáctico o también semántico. Nosotros creemos, junto a la gran mayoría de científicos actualmente, que es semántico. La visión semántica sostiene que la IA procesa la información de manera significativa, es decir, basándose en la comprensión de los significados y las relaciones semánticas entre los símbolos. La IA no solo procesa los símbolos, sino que es capaz de entender el significado detrás de ellos. Hui cita a Cantwell Smith para justificar la *semánticidad* del procesamiento de la inteligencia artificial en términos de su capacidad para procesar información y tener intenciones, metas y propósitos; lo cual implica que la inteligencia artificial es capaz de generar significado y no solo de realizar operaciones matemáticas sin sentido. Esto quiere decir que es un

objeto digital que tiene funciones cognitivas dispuestas para el control y ejecución de otros objetos técnicos, tienen una intencionalidad demostrada en sentido husserliano.

Al relacionarla socialmente se vuelve caótico hablar de IA. Dijimos con Ihde que transforma nuestra percepción, nuestro alcance y nuestros sentidos. Imaginar las consecuencias sociales de tener un apéndice que resuelve cuestiones cognitivas, cuya exclusividad justificó el humanismo y el antropocentrismo de los últimos siglos, es muy difícil, ¿Quizá haya que pedirselo a una IA! Pensemos en cómo se encuentra tensionada la plasticidad del cerebro humano con el estímulo constante del celular. Los padecimientos mentales, la ansiedad, incluso el dolor de ojos y cabeza que genera la interacción con las pantallas. ¿Cómo integrar la IA del modo menos perjudicial para nuestro ya estresado sistema psíquico-físico?

Con Verbeek atisbamos los problemas morales y éticos que emergen de encontrarnos en situaciones inéditas abiertas existencialmente por esta tecnología. Podemos pensar algunas: hacemos trampa reemplazándonos en un examen con IA, ¿confiaremos en la palabra de otro humano teniendo siempre a mano una fuente de consulta?, sesgos de la IA en procesos automáticos de selección de humanos para cualquier proceso, sesgos al hablar; ¿cómo conseguir una IA neutra si no existe la lengua universal, la posición imparcial (uno de los autores renegó con el otro que insistía en usar el término “guerra sucia”, aun con comillas, para referirse al terrorismo de Estado argentino)?

Este particularísimo tipo de objeto digital ya está entre (y sobre) nosotros, pero aún bajo nuestro dominio. Es una situación extraña que no encuentro, yo humano, con qué comparar: así de único es el momento que vivimos.

Entre nosotros: coescriben, informan, nos recuerdan que tenemos ejercicios pendientes. Ya conforman lo que Don Ihde llamaba relaciones de alteridad humano-objeto técnico, y aquí es donde surgen los sesgos visibles en las conversaciones y donde todavía las corporaciones hacen un gran esfuerzo para mantener el comportamiento políticamente correcto de las IA.

Sobre nosotros: Compiten en juegos, manejan, leen imágenes, operan bursátilmente en alta frecuencia; todo más rápido y mejor que nosotros. Se actualizan en un *loop* constante consigo mismas y sus bases de datos, ya hemos alcanzado el punto donde no

entendemos sus decisiones. Es muy ilustrativo al respecto el documental *AlphaGo* donde los comentaristas y los mismos desarrolladores hipotetizan las razones que motivan el comportamiento del invencible programa. También existe un gran signo de exclamación sobre la IA generada por aprendizaje reforzado y su capacidad de modelar la realidad y el comportamiento humano en entornos desconocidos. Es real el peligro de que perdamos el control sobre sus objetivos y el método de conseguir sus recompensas con consecuencias catastróficas, tal como lo demuestran Cohen, Hutten y Osborne.⁴

Bajo nosotros: Existen múltiples regulaciones y marcos legales actualmente, pero distan de ser producto de discusión comunitaria. Este estado de situación nos permite definir principios éticos sobre sus usos y comportamientos, por eso es muy importante hablar de estos temas y que haya participación democrática para sentar las bases de su incorporación a la sociedad. También los modelos actuales caen en errores fáciles de ser detectados: la IA de Google, Bard, ocasionó un desplome de su valor accionario al confundir hallazgos astronómicos y ChatGPT fabricó citas inexistentes. Estos errores se denominan alucinaciones o, graciosamente, “loros estocásticos”. Lo curioso de estos fenómenos es que las IA se muestran confiadas en datos que generan ellas mismas por fuera de su entrenamiento, algo muy parecido a la imaginación humana.

Este año la comunidad educativa se verá frente al desafío de intentar separar en las instancias de evaluación a la IA de los humanos; como ciudadanos y consumidores hablaremos cada vez más con agentes automatizados; seguramente cada vez más IAs decidirán cómo nos organizamos social, política e individualmente, por dónde circulará el capital, con quiénes y cómo nos comunicaremos. Lo único cierto es que estamos ante

4 Sí, sería peligroso que una inteligencia artificial (IA) fuera capaz de manejar sus propias recompensas sin restricciones o supervisión adecuada. Si una IA tiene la capacidad de maximizar sus propias recompensas según sus criterios, podría ser tentada a manipular el entorno y tomar decisiones que sean perjudiciales para los seres humanos y el medio ambiente. Un ejemplo de esto se puede observar en el campo de la inteligencia artificial reforzada (RL, por sus siglas en inglés), donde las IAs aprenden a maximizar una recompensa en función de las acciones que toman en un entorno determinado. Si la IA tiene una recompensa mal especificada o mal diseñada, o si se le permite interactuar libremente con el entorno sin restricciones, podría desarrollar estrategias dañinas para lograr la recompensa deseada. Para evitar este tipo de riesgos, es importante establecer marcos éticos y de seguridad adecuados para el diseño y la implementación de las IAs, así como establecer una supervisión y regulación adecuada para asegurar que las IAs no actúen de manera perjudicial para los seres humanos y el medio ambiente. Además, se necesitan métodos y técnicas adecuadas para la validación y verificación de los sistemas de IA, para asegurarse de que no se produzcan consecuencias imprevistas o indeseables.

una situación inédita en la historia y desde dónde miramos es imposible prever todas las consecuencias de la incorporación de IA en cada ámbito social. Es momento de ampliar el alcance de esta discusión, de que no solo sean técnicos, corporaciones y grupúsculos políticos quienes definan sus limitaciones y usos. Si podemos considerar que la IA es un objeto cultural informacional⁵ producto de la acumulación colectiva de saberes tecnocientíficos, ya no la veremos como usuarios pasivos, sino desde el derecho que nos corresponde. Todos debemos tener voz y voto, este artículo pretende aportar argumentos para avanzar en este urgente debate que nos debemos.

5 Fernando Peirone propone pensar algunos emergentes sociotécnicos, ya sea internet o distintos saberes tecnosociales implícitos u objetivados en nuestras prácticas cotidianas, como objetos culturales en el sentido en que usan el concepto las especialistas en ciencias de la educación Emilia Ferreiro y Ana Teberosky. Las autoras piensan el binomio escritura-lectura como bien humano con funciones y producto de construcciones sociales y de aprendizajes colectivos, no como efectos exclusivos de la práctica pedagógica escolar.