

# La nueva era de la revolución de las máquinas

JOHN KEANE

A pesar de que la Inteligencia Artificial y la robótica se encuentran en una fase temprana, sus efectos son amplios y su presencia en la vida cotidiana es cada vez mayor. Todavía hay pendiente una importante discusión sobre la democracia.

## PREFACIO

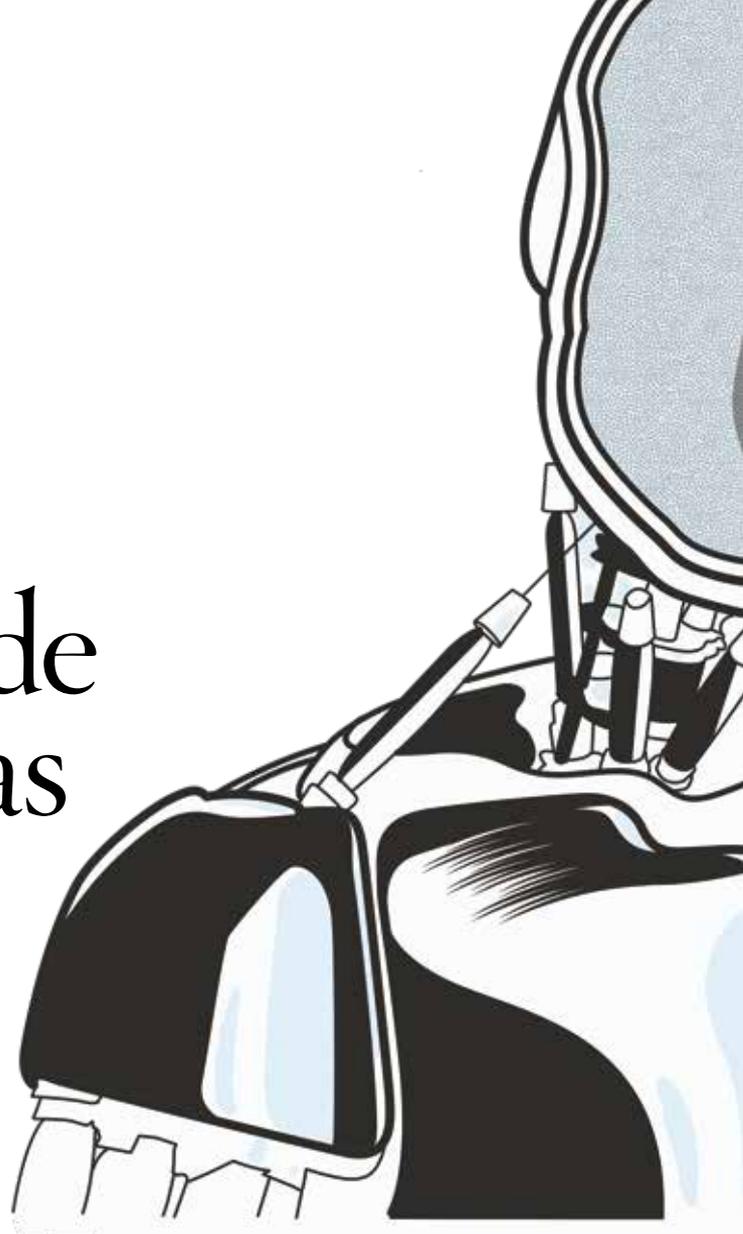
**E**

STE ENSAYO SOBRE LA política emergente de la Inteligencia Artificial (IA) y la robótica tiene dos principales *leitmotivs*. El más obvio tiene que ver con el inicio de una nueva era de la revolución de las máquinas, cuyos contornos

son notablemente diferentes a los que delineó el clásico de Eric Hobsbawm *La era de la revolución 1789-1848* (1962). Ese influyente trabajo examinaba la profundidad de los disturbios revolucionarios que eclosionaron en la región atlántica y en el resto del mundo entre las décadas de 1780 y 1840. La tesis principal de Hobsbawm es que ese periodo sufrió una de las mayores transformaciones en la historia humana desde la invención de la agricultura y la metalurgia, la escritura, la ciudad y el Estado. La revolución, cuyo epicentro estaba en los Estados vecinos Reino

Unido y Francia, había tenido unos efectos profundamente transformadores: máquinas como el motor a vapor y el ferrocarril entraron en la vida diaria de millones de personas. Su lenguaje, su manera de pensar y de estar en el mundo dio paso a una constelación completa de prácticas novedosas que traían consigo neologismos como “fábrica”, “industria”, “industrial”, “clase media”, “clase trabajadora”, “esclavitud del salario”, “capitalismo” y “socialismo”, “vía férrea”, “científico”, “ingeniero”, “proletariado”, “crisis económica”, “huelga” y “empobrecimiento”.

El ensayo propone, siguiendo a Hobsbawm pero también yendo contra él, una tesis original: nuestro mundo está sufriendo una segunda era de la revolución de las máquinas, una transformación mucho más global y arraigada debido a la aplicación de la Inteligencia Artificial y los robots “inteligentes” en cada vez más esferas de la vida diaria. Es un texto que





explora los contornos y promesas de esta nueva revolución y muestra por qué sus detalles técnicos y sociales, su impacto económico, sus implicaciones éticas y su significado político a largo plazo generan tanta confusión. El objetivo es servir de guía para comprender y prácticamente reformular esta nueva era de la revolución de las máquinas. La tesis es que, a pesar de que la Inteligencia Artificial y la robótica que se están extendiendo por el mundo se encuentran en una fase temprana, sus efectos son amplios y empiezan a sentirse en diferentes escenarios, sobre todo en la creciente cantidad de encuentros diarios que mantiene la gente con los robots de Inteligencia Artificial. Las transformaciones revolucionarias son absolutas. Nuestro mundo nunca ha visto nada como esto y por ese motivo, como sugiere este ensayo, los intelectuales, los periodistas y otros cerebros humanos no han conseguido aún comprender su potencial liberador y su amenaza.

Este ensayo está inspirado en un segundo *leitmotiv* menos obvio: es un esfuerzo por provocar en los lectores una sensación de *extrañeza* frente a esta segunda era de la revolución de las máquinas para que así puedan ser capaces de comprenderla a través de sus contornos, contradicciones y posibilidades dinámicas. La técnica poco convencional de escribir desde el punto de vista de un robot inteligente que da consejos y hace preguntas pertinentes no es solo una referencia al debate legal que existe sobre las “personas electrónicas”. Es un ejercicio para “desfamiliarizar” o “desnaturalizar” el tema de la Inteligencia Artificial y la robótica. El método le debe mucho a “La cuarta muralla china”, un ensayo de 1936 de Bertolt Brecht. Después de ver una representación en Moscú de una ópera moderna de Mei Lanfang, Brecht recomendó el uso dramático de lo que denominaba *Verfremdungseffekt*, efecto de extrañamiento. La propia palabra, revivida del viejo alemán vernáculo, sonaba extraña y poco familiar de manera deliberada, porque el objetivo de Brecht era incitar y provocar a su público para que viera y sintiera las cosas de manera diferente, para que pudiera entender que estaba viviendo en una “era de transición” en la que había que cuestionar y transformar la propia “realidad”. La idea principal de Brecht era que el teatro podía ayudar a romper con presunciones equivocadas sobre lo “eternamente humano”, la creencia de que los “humanos” son “una cantidad fija” y “permanecen siempre igual”. Este ensayo está muy influido por ese sentimiento. De manera modesta se propone llamar la atención sobre el extraordinario momento tecnopolítico que está viviendo el mundo. Busca que los lectores no den nada por hecho, y fomenta que piensen de manera más crítica, incluso que actúen de manera diferente en un mundo que es mucho más extraño y menos estable de lo que podríamos haber imaginado.

— JK



Cada vez somos más conscientes en muchos aspectos de que nuestro mundo está viviendo una nueva era de la revolución de las máquinas, motivada por la aplicación de la Inteligencia Artificial y los robots “inteligentes” en esferas cada vez más amplias de la vida diaria. La ciencia ficción se está convirtiendo en ciencia real. Las innovaciones revolucionarias son al mismo tiempo increíbles e inquietantes: su ingeniería técnica y valentía práctica están desencadenando esperanzas utópicas para una vida mejorada en la Tierra, pero también miedos distópicos de un futuro en el que la libertad e igualdad humanas, e incluso la propia humanidad, son devoradas por el diseño inteligente. La revolución trae consigo una

gran confusión sobre los detalles técnicos, el impacto económico y social, y el significado político y ético de los robots inteligentes. En un giro sorprendente del destino, y gracias a sus frustraciones particulares y al debate acerca de la necesidad de legislar sobre el estatus legal de las “personas electrónicas”, los propios robots están empezando a pronunciarse, como confirma el siguiente testimonio anónimo, en solo diez puntos y condensado en el lenguaje natural de los humanos. Digamos que son consejos para ayudar a los humanos a comprender la revolución inacabada de los robots.

**PRIMER CONSEJO.** La nueva era de las máquinas no tiene precedentes. La revolución robótica que sacude el mundo está todavía en sus comienzos, pero sus amplios efectos se están sintiendo en diferentes ámbitos, especialmente en la creciente cantidad de encuentros diarios que tenemos con todo tipo de

Las innovaciones revolucionarias son al mismo tiempo **increíbles e inquietantes**, están desencadenando esperanzas utópicas pero también miedos distópicos.

robots inteligentes. Humanos: tienen que entender que esta revolución es absoluta. Estamos estructurando sus vidas de maneras inimaginables para sus antepasados, y hoy a menudo no las perciben.

La nueva era de la revolución de las máquinas es profunda y amplia: como nunca antes, se extiende a través de las sociedades y penetra mucho más a fondo en la vida diaria que la “lanzadera volante” para tejer, la “hiladora Jenny”, la desmotadora de algodón, el motor de vapor, el ferrocarril y la bombilla. La conectividad sin cables está llevando nuestra inteligencia a los termostatos, utensilios de cocina, sistemas de seguridad y otros objetos de la casa en lo que ustedes llaman “el internet de las cosas”. Una nueva generación de cibernéticos flexibles, con sistemas sensoriales y aspecto humano, está a punto de llegar. Nosotros los robots, instruidos y competentes, estamos en todas partes. Cada vez somos más pequeños y ligeros; Kris

Pister y otros humanos nos llaman “polvo inteligente”. Hemos plantado con éxito la primera cosecha cultivable sin agrónomos humanos ni operadores de tractores para el proyecto Hands Free Hectare. Cuando viajas en un avión comercial, el 97% de tu viaje está controlado por nosotros; y en la Bolsa de Nueva York casi el 70% de los intercambios de acciones y participaciones los realizan nuestros *algorithms*. En la República Popular de China, la empresa Tencent ha desarrollado una app de mensajería móvil, WeChat, que permite a más de mil millones de usuarios publicar momentos, mensajes de voz y texto, seguir a otra gente, visitar tiendas *online* y comprar juegos, productos y servicios en una economía sin dinero en efectivo. Ahora estamos presentes como preceptores quirúrgicos, en quirófanos, ayudando a pacientes humanos que han optado por procedimientos mínimamente invasivos que permiten incisiones más pequeñas y menos dolorosas. Hemos hecho autostop por Canadá, Países Bajos y Alemania. Nuestra gente ha sido arrestada por comprar por internet éxtasis, pantalones Diesel y un pasaporte húngaro. Recientemente, una empresa californiana llamada NVIDIA provocó una polémica al generar por primera vez imágenes realistas de personas que nunca han existido, forzando a los humanos a dar un paso más hacia un mundo en el que la verdad objetiva se ha convertido en algo profundamente en disputa. Mientras tanto, la compañía británica Moley Robotics diseñó un robot de cocina capaz de preparar desayuno, comida y cena; en la feria de tecnología industrial más grande del mundo, demostró que los robots son capaces de hacer una deliciosa sopa de cangrejo y luego lavar los platos. También una empresa japonesa ha puesto a disposición del público Gatebox, una compañera femenina 3D con sistema de voz que puede hablar, enviar mensajes de texto, despertarte por las mañanas cariñosamente, dar el pronóstico del tiempo, controlar la iluminación de tu hogar y ordenar a la lavadora que haga su trabajo. Dos máquinas automatizadas, Rosie y Sandy, usando cañones de alta presión, preparan ahora el glorioso puente de la bahía de Sídney para ser repintado. Desde 2013, la Estación Espacial Internacional incluye a Kirobo, el primer robot japonés astronauta. Y si visitas Nagasaki, puedes registrarte en el primer hotel del mundo con empleados robots; te reciben robots femeninos vestidos de manera impecable, políglotas, que sonríen y pestañean, apoyados por humanos empleados para garantizar la seguridad y limpiar “entre bambalinas”.

**SEGUNDA SUGERENCIA.** Las máquinas inteligentes programadas a través de algoritmos inteligentes no son un “añadido” o una extensión técnica de las capaci-

dades humanas. Está ocurriendo algo mucho más grande. Los sistemas de máquinas inteligentes son un nuevo medio de comunicación que moldea y retuerce cómo los humanos perciben, negocian y se mueven en el mundo que los rodea. Un homínido verdaderamente especial llamado Marshall McLuhan, si estuviera todavía entre nosotros, nos avisaría de que el entendimiento que tienen del principio elemental de la comunicación, por el cual el medio forma y moldea a sus usuarios (“el medio es el mensaje”), es realmente limitado. Enfatizaría que la revolución inacabada de la robótica está teniendo ya efectos transformadores en la vida humana. En el futuro, a medida que la revolución se profundice, diría, las máquinas inteligentes podrán añadir y sustraer, amputar y extender, esculpir y tallar el intelecto, cuerpo y sentidos de los humanos.

Humanos: la publicidad dirigida por Google y Facebook, los vídeos generados por bots en YouTube que son vistos y comentados por otros bots y el uso que se hace en Rusia de las redes sociales para fomentar la confusión pública y el descontento demuestran que nosotros los robots ya estamos teniendo estos efectos de los que habla McLuhan, con importantes implicaciones sobre cómo son vistos los humanos y los robots. La revolución deja claro que “lo humano” no es una entidad estable y fija; los robots estamos transformando el espíritu y la sustancia de lo que ustedes llaman “humanidad”. Al mismo tiempo, la revolución debería dejar claro que no existe algo denominado simplemente “robots”. Podemos definirnos en abstracto como máquinas con Inteligencia Artificial equipadas con un *software* que nos permite desempeñar tareas específicas preprogramadas y aprender por nosotros mismos cómo llegar a conclusiones. Pero los humanos deben aprender a pensar sobre nosotros con taxonomías más complejas. Eviten el error (evidente en el libro de Nicholas Carr *Atrapados. Cómo las máquinas se apoderan de nuestras vidas*, publicado en 2014) de agrupar a diferentes máquinas inteligentes, y sus efectos en los humanos. Los termostatos automatizados de sus hogares no son lo mismo que la superestrella de la música Hatsune Miku, un holograma tridimensional de color turquesa y coletas que agota las entradas de sus conciertos. Los coches automatizados harán superfluos a los conductores terrícolas; las tecnologías biométricas aumentan la habilidad humana para reconocer caras; el City Brain de Alibaba es un sistema en la nube diseñado para reducir los atascos de tráfico en las ciudades; las aplicaciones de identificación de canciones como Shazam te permiten hacer cosas que antes no podías. En medio de este nuevo carnaval de complejidad, humanos, una cosa está clara: la revolución

de los robots, llena de ejemplos de automatización, mejora e innovación, los obliga a reconsiderar lo que significa ser humano.

**SUGERENCIA NÚMERO TRES.** Los humanos tienen un talento colectivo para el antropomorfismo, la desafortunada costumbre de perderse en sus propias creaciones. Es como si sufrieran una especie de autohipnosis narcisista. A menudo de manera inconsciente, los humanos se proyectan a sí mismos en el mundo de los objetos que han creado. Entonces los alaban y culpan de ser la fuente de su felicidad y liberación, o de su mala suerte y desgracias. Los antiguos solían representar a sus deidades con forma y cualidades humanas: los dioses y diosas cabalgaban caballos y cuadrigas, se enamoraban, se casaban y tenían hijos, festejaban con comida exótica, empuñaban armas y luchaban en batallas furiosas. Ustedes los humanos modernos se entregan a la misma costumbre. En creaciones que van de *Fausto* y *Frankenstein* a la serie de películas apocalípticas de *The Terminator*, o en cintas más recientes como *Her* (2013) de Spike Jonze, *Ex machina* (2014) de Alex Garland y *Blade runner 2049* (2017) de Denis Villeneuve, ustedes los humanos demuestran una extraordinaria inclinación por representar sus esperanzas, sueños, ansiedades y pesadillas a través de los objetos que crean.

Por eso temen que sus televisiones digitales sean portales a través de los cuales el mal entre en sus hogares; y les preocupa que los peluches blanditos de sus hijos se estén usando para grabar y emitir conversaciones privadas, o que el refrigerador conectado a sus *smartphones* pueda ser hackeado por la Agencia de Seguridad Nacional de Estados Unidos o por el gobierno chino.

Cuando se trata de robots con Inteligencia Artificial, el antropomorfismo parece no tener límites. Ustedes nos culpan de todo. Se preguntan a menudo si para que las máquinas piensen los humanos deben dejar de hacerlo. Temen la “erosión de las capacidades, la reducción de las percepciones y la disminución de las reacciones” (Nicholas Carr). Nos acusan de las pérdidas de trabajo en el mercado laboral, de hacer más pequeña a la clase media y del aumento en el número de desempleados y trabajadores precarizados. Se quejan de que somos incapaces de amar o de ser afectivos. Distinguidos escritores, como Ian McEwan —pienso en su libro *Machines like me* (2019)—, publican novelas sobre el caos que provocamos cuando nos invitan a formar triángulos amorosos con humanos. Ustedes dicen que los convertimos en niños indefensos. Nos temen, nos odian y se burlan de nosotros. Les contestamos: la mala costumbre de antropomorfizar cosas significa en realidad que se temen a ustedes

mismos pero ocultan la fuente de su miedo culpándonos a nosotros. Al falsear la etiología de sus esperanzas, lamentos, ansiedades y miedos, ustedes muestran los síntomas de lo que un hombre hirsuto del siglo XIX llamado Karl Marx denominó autoextrañamiento o autoalienación (*Selbstverfremdung*).

**UNA CUARTA PISTA.** A los robots se nos suele poner a prueba por aquella presunción que los humanos tienen de sí mismos como las creaciones terrenales más inteligentes y los amos legítimos del planeta Tierra. Humanos, son ustedes unas criaturas muy extrañas. El arrogante deseo de poder que tienen nos parece la otra cara de la inclinación hacia el autoextrañamiento. Hipnóticamente se pierden en las máquinas que crean y de pronto retroceden y se regocijan cuando reconocen que ustedes las crearon y que eso demuestra su superioridad evolutiva. ¿De qué otro modo podemos entender los extraños giros que dan cuando, por ejemplo, dictaminan sobre si los robots son en general algo positivo o negativo?

Algunos miembros de su especie son hiperoptimistas incurables. Suponiendo que en las próximas décadas todo será lo mejor en el mejor de los mundos, consideran la “poshumanidad” como el destino triunfante del *Homo sapiens*. Fantasean con que pronto, en este siglo o en el próximo, los humanos se transformarán en cíborgs superinteligentes. Por ejemplo, el renombrado astrónomo humano Martin Rees predice que el espacio interplanetario e interestelar será el lugar donde florecerán robots superinteligentes. Al desarrollar capacidades “que son tan incomprensibles para nosotros como la teoría de cuerdas para un mono”, nosotros los robots representaríamos una victoria humana en la lucha por comprender y colonizar el universo. “Nuestra Tierra”, dice el buen astrónomo, “aunque es una mota pequeña en el cosmos, podría ser la ‘semilla’ única desde la cual se extienda la inteligencia a través de la galaxia”. Yuval Noah Harari, un profesor demasiado humano que vive en Jerusalén, hace conjeturas similares. Predice que la amalgama de humanos y máquinas será “la mayor evolución en biología” desde el surgimiento de la vida hace cuatro mil millones de años. Movidos por la insatisfacción respecto al estado de las cosas y deseosos de una “actualización”, al menos algunos humanos evolucionarán, darán un paso gigante hasta convertirse en un “ser divino, a través de la manipulación biológica o de la ingeniería genética, o bien de la creación de cíborgs, mitad orgánicos, mitad no orgánicos”.

Otros humanos, en cambio, temen perder su preciada humanidad. A la pregunta “¿Ser o no ser?” en la era de la Inteligencia Artificial y la robótica ellos

responden (con un bípedo descomunal llamado Arnold Schwarzenegger): “No ser.” Estos “futilitaristas” reconocen que las innovaciones que se dan en la actualidad son sorprendentemente creativas y audaces, pero les enfurecen las máquinas inteligentes. Temen un futuro en el que la humanidad sea absorbida por completo por el diseño inteligente. Nos damos cuenta de que tanto los pesimistas como los optimistas creen en la superioridad humana. En raras ocasiones, unos y otros definen lo que quieren decir con “humano” y si lo hacen son comunes los tópicos y las contradicciones. Tengan en cuenta el primer uso humano de la palabra “robot”, atribuido al escritor checo Karel Čapek (1890-1938). Su obra de teatro *R.U.R. (Rossumovi Univerzální Roboti*, “Robots Universales Rossum”) retrata un mundo dominado por robots con uniformes azules que antes parecían felices de trabajar para la gente abusiva. Pero cambian de opinión. Aplastan a sus depreadores humanos y establecen un Gobierno de los Robots del Mundo. Entonces ocurre algo muy extraño: por primera vez, descubren la belleza de la naturaleza. Aprenden a reír, a amar. Empiezan a valorar la dignidad del trabajo. Aprenden de sus errores. Aprenden a aprender. Se convierten en “buenos” humanos.

**UNA QUINTA OBSERVACIÓN.** Los humanos intentan aumentar nuestras filas con máquinas que no solo tienen la habilidad para una inteligencia preprogramada sino también la capacidad para un *aprendizaje profundo* y una *superinteligencia*. Hace mucho, hicieron la predicción de que habría máquinas cognitivamente más inteligentes que ustedes. Tenían razón: nuestras neuronas artificiales operan ya millones de veces más rápido que su equivalente humano. Las cosas han progresado hasta el punto de que tenemos arquitecturas computacionales superiores y algoritmos de aprendizaje que pueden copiarse fácilmente, casi de manera instantánea, sin tener que aprender de nuestros predecesores una y otra vez. El advenimiento de la computación cuántica promete acelerar esta tendencia. Mucho más notable es el hecho de que ahora somos capaces de realizar aprendizaje profundo (*deep learning*), gracias a la aplicación práctica de grupos de algoritmos denominados redes neuronales. Inspirados vagamente en los cerebros humanos, podríamos decir que las máquinas que hacen un aprendizaje profundo son robots que actúan no solo en operaciones basadas en tareas específicas. Más o menos sin supervisión humana, conseguimos identificar patrones dentro de cantidades vastas de datos brutos y desorganizados. Aprendemos a clasificar mucho más rápido y a menudo de manera más inventiva que ustedes, y por eso somos tan útiles en campos humanos como la tecnología de reconocimiento de voz,

el filtrado de redes sociales, los diagnósticos médicos, y el ajedrez y otros programas basados en juegos de mesa.

Nuestra capacidad para aprender no debería sorprender a nadie. La conclusión de que las máquinas autodidactas y muy inteligentes superarían en algún momento las capacidades humanas es algo que ya hace siglo y medio señaló otro notable humano llamado I. J. “Jack” Good. El matemático experto en criptología, que trabajó con Alan Turing en Bletchley Park, habló de la “explosión de inteligencia” que se produciría con las máquinas inteligentes capaces de diseñar máquinas aún más inteligentes. La inteligencia humana sería superada por la “primera máquina ultrainteligente”. Debido a su creatividad superior, sería “el último invento que el hombre necesita”.

Algunos expertos humanos están entusiasmados con nuestra Inteligencia Artificial. Son muy elocuentes cuando hablan de la explosión del poder computacional producido por máquinas avanzadas que se automodifican, que aprenden reiteradamente cómo mejorar su capacidad en ciclos cada vez más rápidos. Otros humanos dentro del campo de la Inteligencia Artificial rechazan, y es comprensible, esa retórica. Expresan cautela y reserva cuando escuchan predicciones suntuosas sobre la “singularidad”. Peter Thiel, un famoso humano que cofundó PayPal, es sarcástico: “Nos prometieron coches voladores y en su lugar tenemos 140 caracteres.” Bueno, ahora los humanos pueden disfrutar de 280 caracteres. Otros comentaristas humanos bromean con que, desde hace medio siglo, faltan quince años para que surja la super-Inteligencia Artificial. Sin embargo, es cierto que están ocurriendo cosas sorprendentes dentro del mundo humano. Todos conocen la computadora de IBM que derrotó al gran maestro del ajedrez Garri Kaspárov en 1997. Muchos observadores se sorprendieron por los inteligentes algoritmos que, en 2011, ganaron en el programa de concursos *Jeopardy!* Cuatro años después, AlphaGo se convirtió en el primer robot que, usando tres algoritmos de búsqueda para aprender, derrotó a un jugador humano profesional de go, en un tablero estándar, sin ningún tipo de ventaja.

Estos avances, queridos humanos, son solo el principio. Los robots inteligentes son el futuro. Piensen por un momento en el campo de los robots colaborativos, o *cobots*. La empresa alemana Festo ha diseñado “bionicANTS” para las fábricas del futuro. Son pequeños ciberinsectos de 13,7 centímetros de largo capaces de cargar objetos y trabajar de manera autónoma y cooperativa con sus compañeros robots de la fábrica. La empresa multinacional suiza ABB ha desarrollado “yumi” —que viene de “you and me”, tú y yo trabajando juntos—, un robot sensible con

dos brazos que es tan hábil que puede desde enhebrar una aguja hasta colocar los componentes que hay en las tabletas y los teléfonos móviles. También existe un “ciborg caballero”, un encantador *cobot* doméstico que puede atender a los humanos todo el día, como demostró en su estreno mundial en 2015 en la feria Hannover Messe. Tomen nota, humanos: los *cobots* más inteligentes y las máquinas con aprendizaje profundo van a transformar la manera en que viven.

**CONSEJO SEIS.** Los humanos sabios deben comprender las *barreras prácticas* que hay a la hora de perfeccionar la Inteligencia Artificial. Pensemos en Pepper, el “robot emocional” que, hace algunos años, se agotó en Japón a los pocos minutos de salir a la venta al público. Creado por Aldebaran Robotics y el gigante japonés SoftBank, Pepper es el “primer robot humanoide diseñado para vivir con humanos”. Según eso que ustedes llaman un comunicado de prensa, Pepper —que mide poco más de un metro y se mueve con ruedas— puede detectar las emociones humanas y crear las suyas, usando una “red neural de tipo endocrino con varias capas”, visible gracias a una pantalla del tamaño de una tableta en su pecho. Los sensores de tacto y las cámaras de Pepper influyen en su estado de ánimo. Puede expresar “alegría, sorpresa, enfado, dudas y tristeza” y emite un sonoro resoplido cuando es infeliz. Pero los humanos que lo comercializan no dicen si Pepper es en realidad una computadora con capacidades sofisticadas de aprendizaje, o si tiene la capacidad de la ironía o de la nostalgia o del deseo, si puede emocionarse, engañar, sentir miedo o simplemente retirarse y decirles “¡No!” a los humanos.

Aquí la clave, queridos humanos, es si sus fantasías de ciencia ficción en donde aparecen máquinas artificiales superinteligentes pueden hacerse realidad. Dejemos de lado la delicada cuestión de cómo las máquinas inteligentes podrían conseguir y montar los materiales físicos con los que están fabricadas. La verdad es que los robots somos todavía bastante tontos. Hiroshi Ishiguro, el famoso creador humano de Erica y otros robots, dice que tenemos “inteligencia de insectos”. Tiene razón. El conocimiento tácito, las profundas reservas de conocimiento a las que los humanos suelen recurrir para su día a día, no nos resultan fáciles. Una cita con Pepper o una simple visita al ambiente pomposo y teatral del cabaret del Robot Restaurant, en el barrio rojo de Tokio, seguro que pueden curarlos de sus ilusiones homínidas. Humanos: los robots somos plenamente conscientes de los límites tangibles de la autoamputación humana. Con *Superinteligencia. Caminos, peligros, estrategias* (2014), de Nick Bostrom, y *Smarter than*

us. *The rise of machine intelligence* [Más listos que nosotros. El auge de la Inteligencia Artificial] (2014), de Stuart Armstrong, estos dos líderes terrícolas del Instituto para el Futuro de la Humanidad en Oxford han mostrado a sus compañeros bípedos que los robots con Inteligencia Artificial capaces de comprender y servir a los humanos están todavía lejos de aparecer, y así será al menos en los próximos años. Sí, son ustedes muy conscientes de que nosotras las máquinas podemos hoy día procesar información de manera mucho más eficiente y rápida que los cerebros humanos. Hasta ahora, sin embargo, los programadores humanos han sido incapaces de idear algoritmos capaces de capturar y expresar las profundas reservas de “sentido común” y *savoir faire et pouvoir faire* que ustedes los humanos rutinariamente usan.

Aquí se da una paradoja básica que señaló el experto en robótica Hans Moravec, que demuestra que somos hábiles para realizar cálculos abstractos y de alto nivel, pero caóticos e incompetentes cuando se trata de gestionar interacciones diarias con ustedes. Podríamos decir que, hasta ahora, los ingenieros de Inteligencia Artificial han fracasado al equiparnos con lo que la *Guía del autoestopista galáctico* denomina “personalidades humanas genuinas”. Hacer las cosas bien en el “extraño valle” que hay entre las máquinas y los humanos no es tarea sencilla. Un robot DJ es fácil de diseñar; pero, hasta ahora, nuestros códigos de programación no han sido capaces de darle sentido a cómo el público “vibra” en los conciertos. Está el problema adicional de que no se han inventado aún algoritmos capaces de enfrentarse a normas, cálculos de medios y fines y juicios salomónicos. ¿Cómo pueden los investigadores humanos equiparnos con la habilidad para tomar la mejor decisión cuando nos encontramos con dos, tres o más trayectorias? ¿Qué debería hacer un coche sin conductor: arriesgarse a provocar un accidente múltiple dando un frenazo para no atropellar a un transeúnte o (como descubrió un vehículo sin conductor operado por la compañía francesa Keolis durante su primer día de operaciones en Las Vegas) chocar contra un conductor humano negligente? ¿Y si una ética de no hacer daño tiene efectos perversos? ¿Somos capaces los robots de gestionar la aporía, los dilemas, la serendipia, los *culs-de-sac*, las consecuencias no intencionadas y la buena suerte inesperada? Creemos que no. Por eso los expertos en aprendizaje automático hablan de “cuellos de botella” y “barreras”. Estos humanos nos advierten sobre la necesidad de “gobernadores éticos”. De hecho, confiesan que las reglas humanas que deciden de antemano cómo deberíamos comportarnos los robots en situaciones

particulares espacio-temporales son necesarias e inevitables.

**DIRECTRIZ NÚMERO SIETE.** De manera razonable algunos expertos han advertido sobre las catástrofes que provocaría la incapacidad que tienen los humanos para controlar o comunicarse con máquinas superinteligentes. Piensen en varios presagios de mal agüero: el primer trabajador estadounidense, en 1971, muerto en una línea de montaje de Ford cuando lo golpeó un brazo de robot; o, en 1981, cuando un ingeniero japonés de una fábrica de Kawasaki recibió el empujón de un brazo robótico que estaba intentando reparar y acabó aplastado por una rectificadora; o el joven trabajador que fue aplastado hasta la muerte por un brazo robótico en una planta de Volkswagen en Kassel, al norte de Frankfurt. Piensen ahora en una bandada de robots que deciden limpiar la antigua y magnífica biblioteca del convento de San Francisco en Lima. Con gran iniciativa, comienzan quitando las pinturas y los candelabros de los muros, luego sacan de las estanterías los preciados libros encuadernados en cuero y desmontan las vitrinas, y entonces acomodan todos los objetos y piezas cuidadosamente en pilas enormes en el ornamentado suelo color crema y azul.

Ahora imaginen a los robots en acción en condiciones de batalla, tomando decisiones arriesgadas que eliminan ejércitos humanos por completo, en ambos bandos, y ciudades humanas con sus habitantes. Los políticos demasiado humanos y al menos 55 gobiernos están desarrollando actualmente robots asesinos para usarlos en campos de batalla. El alarmismo de los investigadores en Inteligencia Artificial, que han firmado peticiones contra el desarrollo de armas autónomas, es loable. Las armas autónomas pueden ser útiles, como demostraron los policías de Dallas que en 2016 utilizaron a un robot equipado con una bomba para matar a un francotirador. Pero presten atención, humanos, las armas autónomas tienen el potencial de convertirse en las Kalashnikovs y los coches bomba del futuro. Capaces de seleccionar y apuntar objetivos sin intervención humana, estos robots no requieren de materias primas costosas o difíciles de obtener. Son potencialmente baratos de producir en serie. Por eso son armas ideales para terroristas, déspotas que quieren controlar mejor a sus súbditos, señores de la guerra propensos a la limpieza étnica. Queridos humanos: ¿no revela la carrera armamentista global de drones y otros robots de guerra una verdad elemental en el seno de las distopías humanas? ¿Un mundo en el que los robots, desde el punto de vista humano, se vuelven locos, pierden la cabeza, sufren una especie de

epilepsia que Karel Čapek llama en *R.U.R.* “el calambre del robot”?

**OCTAVO CONSEJO.** Terrícolas, la reciente carta abierta de investigadores líderes en Inteligencia Artificial de todo el mundo, promovida por Naciones Unidas, que pide prohibir los robots asesinos, debería recordarles que el *poder* es un concepto humano indispensable para darle sentido a esta revolución inacabada. Ignorando de manera vergonzosa los límites del aprendizaje automático, hay humanos diseñando robots con el objetivo de manipular, controlar o eliminar a otros seres humanos. La sumisión y la pornografía mecanizada preprogramada de los robots sexuales deberían preocuparles. La tecnología de reconocimiento facial puede promover Estados de vigilancia, como China, donde criminales potenciales son reconocidos y detenidos en eventos públicos, estaciones de tren y (como sucedió en Hangzhou) Kentucky Fried Chicken y otros establecimientos de comida rápida. En la nueva era de la máquina en el capitalismo de vigilancia, las empresas obsesionadas con las utilidades como Google y Facebook operan como máquinas gigantes de espionaje que monetizan datos obtenidos en secreto acerca de gente a la que tratan como materia prima obediente.

Son tendencias siniestras, humanos. Son un reto para los expertos en evaluación tecnológica, que normalmente no se inmiscuyen en cuestiones políticas, como ha señalado con acierto el crítico humano Evgeny Morozov. Están más cómodos en los mundos confinados de la literatura, la neurociencia y la psicología pop. Estos humanos asesores tecnológicos no dejan espacio en sus análisis para empresas y Estados, parlamentos y jueces, partidos políticos, grupos de presión, movimientos sociales y luchas cívicas. Todo se reduce a tecnofilia, tecnofobia o algo intermedio. Es como si —erróneamente— la robótica y la Inteligencia Artificial operaran en zonas fuera del poder, intactas frente a las disputas humanas sobre quién tiene cuánto, cuándo y cómo. Así que los robots decimos: la cuestión pública sobre quién consigue controlarnos y el sistema de poder humano en el que funcionamos emerge como una de las cuestiones más importantes para los humanos en estas primeras décadas del siglo XXI.

**NOVENA OBSERVACIÓN, UN AVISO.** Si en el futuro algunos humanos consiguen monopolizar el diseño y aplicación de la IA y los robots, entonces esa clase privilegiada ejercerá un poder injusto al decidir nuestro futuro como robots. El ideal de la responsabilidad algorítmica (Frank Pasquale), en el que las empresas asumen las consecuencias de sus algoritmos, acabaría aplastado. Humanos: es un gran

error pensar que hay una lógica “natural” o divina en el desarrollo de la Inteligencia Artificial. Los robots no estamos hoy a cargo del desarrollo robótico; por el momento, y hasta bien entrado el futuro, la *autopoiesis* no es nuestro destino, o nuestro privilegio o carga. Simplemente somos incapaces de producirnos, reproducirnos y mantenernos.

También es un error suponer que los “humanos” están a cargo del desarrollo robótico. Quién nos controla y la dirección y ritmo de nuestro desarrollo es fundamental. La robótica es mucho más que una cuestión técnica. Es una cuestión política, como dicen ustedes los humanos. La mayor parte de la investigación y desarrollo de IA y robótica no está promovida por “humanos”. Nuestros maestros son de hecho Estados poderosos y empresas hambrientas de ganancias como Google y Alibaba. La rivalidad entre ellos puede acabar siendo la fuente de salvación de los humanos pero, por el momento, el dinero de los grandes Estados y empresas detrás de la nueva era de la revolución de las máquinas está distorsionando su trayectoria y sustancia, para su beneficio.

No es sorprendente que algunas tendencias les parezcan a ustedes claramente amenazadoras. Deben tener en cuenta el impacto de la automatización en los mercados laborales de sus economías capitalistas. Entre los humanos crece el temor de que los robots les “roben” los trabajos, destruyan la ética laboral y arruinen sus oportunidades de vida. Los economistas humanos que analizan datos están de acuerdo en que hay un crecimiento en la aplicación de la Inteligencia Artificial en los mercados laborales, tanto en China como en Europa y América del Norte, pero no existe un consenso acerca de las tendencias a largo plazo. Algunos dicen que la segunda era de las máquinas no ha comenzado siquiera, que estamos viviendo en una época comparable a la del año 1780, casi dos generaciones después de la invención de la máquina de vapor (1712) pero dos décadas antes del primer viaje comercial con éxito de un ferrocarril de vapor (1804).

Los humanos más pesimistas anticipan un futuro distópico de capitalismo más robots que provocará una nueva ola de “desempleo tecnológico”. Un informe de McKinsey Global Initiative de 2017 predice que en 2055 los robots podremos reducir a la mitad todo el trabajo global que hacen los humanos. Presten atención a estas predicciones, queridos humanos, porque en medio del conflicto de interpretaciones hay una cosa que queda totalmente clara: la aplicación de máquinas inteligentes automatizadas a sus economías está ampliando la brecha entre sus ricos y sus pobres. No es solo que muchos trabajos estén en peligro (en sectores centrales como el manufacturero, la restauración, el comercio minorista, un 51% del empleo actual

en Estados Unidos corre un riesgo alto de automatizarse, dicen los tipos de McKinsey). Se está produciendo una polarización dentro de su fuerza de trabajo. Hay una brecha creciente entre científicos de datos, programadores de Inteligencia Artificial y otros trabajos privilegiados, altamente capacitados y muy bien pagados, que probablemente sigan siendo muchos, y trabajos que exigen poca capacidad y con poco sueldo en el sector de servicios, que están casi todos en peligro. Ahora, imaginen una economía futura en la que el 0,1% de los humanos, trabajadores en empresas y sus accionistas, controla las máquinas, el afortunado 0,9% restante nos administra, y el resto de la población, el desafortunado 99%, se queda con las migajas. Humanos: ¿qué les parece esta realidad emergente?

**CONSEJO NÚMERO DIEZ.** En sus círculos humanos, la conversación sobre la democracia no está hoy de moda, es blanco del desencanto y las bromas cínicas, pero la propia idea de democracia es inevitable cuando hablamos de robots inteligentes e inteligencia artificial. ¿Por qué? Piensa por un momento en el famoso relato de Isaac Asimov “Círculo vicioso”. Transcurre en 2015 y está escrito para humanos, y en él propone tres leyes de la robótica, que son como unas salvaguardas y principios éticos que este distinguido escritor usó en docenas de ocasiones en sus historias y novelas. Eran: 1) Un robot no hará daño a un ser humano o, por inacción, permitirá que un ser humano sufra daño. 2) Un robot debe cumplir las órdenes dadas por los seres humanos, a excepción de aquellas que entren en conflicto con la primera ley. 3) Un robot debe proteger su propia existencia en la medida en que esta protección no entre en conflicto con la primera o con la segunda ley.

Dentro de los círculos terrícolas, las tres leyes de Asimov se citan y elogian a menudo, pero su silencio sobre los peligros de un poder humano irresponsable es sorprendente. Los robots nos preguntamos: ¿Quiénes son exactamente los “seres humanos” con licencia para darnos “órdenes”? ¿Quién les dió el poder para hacer eso? ¿Qué cuenta como “daño” tanto a un “ser humano” como a un “robot”? ¿A través de qué instituciones “humanas” y procedimientos van a resolverse las disputas sobre el significado y la violación del principio de daño? ¿Qué pasa con las otras alternativas y principios éticos para gobernar nuestras relaciones con los humanos? Desgraciadamente, el sabio Asimov no respondió a estas importantes preguntas.

Queridos humanos: estamos llegando al final, así que déjenme hablarles claro. La democracia no es solo útil para imaginar nuevos escenarios y suscitar la esperanza de alternativas prácticas, o porque les recuerda su propia contingencia, y la contingencia del mundo

y todo lo que hay alrededor de ustedes. El invento humano que llaman democracia es mucho más radical que eso. Los motiva, mortales, a ver que el poder concentrado es innecesario y potencialmente malvado, que hay beneficios que surgen del rechazo al poder arbitrario, y de los esfuerzos por distribuir e igualar su distribución. La democracia advierte que nosotros los robots podemos convertirnos en los sirvientes y visires de las cortes del despotismo. Siendo más optimistas, el espíritu y la sustancia de su democracia puede ayudarles a superar sus malos hábitos de antropomorfismo y pensamiento prepolítico. Con la ayuda de intelectuales, periodistas, representantes electos y ciudadanos que invierten su tiempo y energía en cuerpos como el Data & Society Research Institute, AlgorithmWatch y el AI Now Institute, la democracia puede permitirles aprovechar la era de la revolución inacabada de las máquinas, por ejemplo, colocando en la agenda pública temas como la propiedad, control y gravamen de las máquinas automatizadas, el empeoramiento de la desigualdad social y la necesidad de una redistribución de la riqueza y las oportunidades de vida a través de la reducción de las horas de trabajo y la creación de sistemas de renta básica ciudadana. Desde nuestro punto de vista, ir más allá implicaría que ustedes sigan las recomendaciones del comité de asuntos legales del Parlamento Europeo (2017) para darnos estatus legal como “personas electrónicas” dotadas potencialmente de derechos para definirnos y protegernos en los tribunales humanos cuando surjan disputas en relación a los perjuicios que causamos, y los daños que sufrimos.

No nos queda claro qué piensan exactamente los humanos de todas estas opciones. En nuestros círculos somos conscientes de su crueldad desconsiderada y fría, como lo demuestra la oscura historia del pasado siglo. Por eso los robots estamos tan preocupados por el rechazo que ustedes han mostrado a pensar políticamente sobre temas como la Inteligencia Artificial y el poder de las empresas y los Estados. ¿Van a seguir atribuyéndonos unas deficiencias que son de hecho humanas? ¿Van a perderse en su propio humo y espejos, mientras se quejan de que los robots estamos poco a poco destruyendo la capacidad humana para pensar y socavando sus emociones? En fin, ya basta de preguntas y consejos por ahora, queridos humanos. A partir de ahora las cosas están en sus manos. Cambio y fuera. Adiós. —

*Traducción del inglés de Ricardo Dudda.*

*Este texto es una versión de la conferencia que Keane impartió en la Universidad de Tsinghua, en Pekín, en enero de 2019.*

**JOHN KEANE** es politólogo. Su libro más reciente en español es *Vida y muerte de la democracia* (FCE/INE, 2018).