

El próximo paso

LA VIDA EXPONENCIAL

Cuando los robots gobiernen la Tierra: el legado humano

ROBIN HANSON



Imagen de apertura:
Zaha Hadid Architects
Interior del Heydar Aliyev Centre (2013)
Baku, Azerbaiyán



Robin Hanson

George Mason University, Fairfax, Vancouver, USA
Future of Humanity Institute of Oxford University, Oxford, Reino Unido

Robin Hanson es profesor asociado de Economía en la George Mason University e investigador asociado en el Future of Humanity Institute de la Oxford University. Tiene un doctorado en Ciencias Sociales por Caltech y sendos másteres en Física y Filosofía por la University of Chicago. Durante nueve años trabajó como investigador de inteligencia artificial en la empresa aeroespacial Lockheed y la NASA, y cuenta en su currículum con 3.175 citas, 60 publicaciones académicas, 500 menciones en medios de comunicación, 200 charlas como conferenciante invitado y ocho millones de visitas a su blog OvercomingBias.com. Oxford University Press publicó en junio de 2016 su libro *The Age of Em: Work, Love and Life when Robots Rule the Earth*, y en la primavera de 2017 publicará *The Elephant in the Brain*, escrito en colaboración con Kevin Simler.

CONOCE MÁS DEL AUTOR [+]

Las tres grandes disrupciones en la historia han sido la llegada de los humanos, la agricultura y la industria. En algún momento del próximo siglo, de la inteligencia artificial podría surgir una disrupción similar en forma de emulaciones de cerebros o *ems*. A partir del consenso académico en muchas disciplinas, perfiló un escenario básico de un mundo inmediatamente posterior a un periodo de transición dominado por *ems*. Hablo de arquitectura, energía, enfriamiento, velocidad mental, tamaño corporal, estrategias de seguridad, realidad virtual, mercados de mano de obra, gestión, formación profesional, salarios, identidad, jubilación, ciclos vitales, reproducción, apareamiento, conversación, desigualdad, ciudades, tasa de crecimiento, políticas de coalición, gobernanza, derecho y guerra en un mundo automatizado.

INTRODUCCIÓN

Durante muchas décadas hemos tenido numerosos tipos de robots (y otros ayudantes basados en computadoras), pero adolecían de unas capacidades limitadas. Por ejemplo, si valoramos los trabajos en función de lo que pagamos por hacerlo, debemos deducir que los humanos siguen siendo mucho más valiosos que los robots, puesto que se les remunera más.

Sin embargo, son ya muchas décadas de progresión de las capacidades de los robots; una progresión lenta, pero mucho más rápida que el avance de las capacidades humanas. Por tanto, aunque tarde siglos en ocurrir, cabe esperar que los robots lleguen a superar a los humanos en prácticamente todas las tareas. Entonces, casi toda la remuneración por tareas irá destinada a los robots (o a sus propietarios). Sería este un mundo «dominado» por los robots, al menos en términos de quién hace la mayor parte del trabajo y quién toma la mayor parte de decisiones concretas que requieren una comprensión detallada del contexto. En cuanto a los humanos, quizás seguirían decidiendo sobre las grandes políticas abstractas, simbólicas y de alto nivel.

¿Qué podemos hacer hoy para influir en los acontecimientos de un futuro dominado por los robots? En primer lugar, abordo la cuestión general de cómo influir en el porvenir, para después centrarme en dos tipos de robots del futuro.



LA DIFICULTAD DE DEJAR UNA HERENCIA

Cuando la gente habla de un futuro lejano, suele incidir en lo que le gustaría que ocurriera. Entre los muchos escenarios futuros imaginables, ¿cuál prefiere? Se habla mucho de sus valores básicos, poco de las limitaciones prácticas que restringirían los escenarios factibles y menos aún de cómo podemos influir hoy en los resultados del mañana. Sin embargo, esta última cuestión, que a menudo se omite, parece clave: ¿cómo influir en el futuro? ¿Para qué pensar en qué queremos si no podemos cambiarlo?

Imagina que estás en la orilla de un río que desemboca en el mar, a muchos kilómetros de su origen. Desde el río, ves a una persona a quien aprecias que está en ese mismo río, pero cerca de la desembocadura en el mar, y quieres hacer algo en tu trecho de río para influir, de un modo u otro, en la persona que está corriente abajo. Aunque lo ideal sería conseguir un efecto benigno, para empezar te conformas con tener una influencia significativa.

Pero resulta bastante difícil. Puedes salpicar con el agua de alrededor, calentarla con una antorcha o tirar piedras al río. Pero los ríos tienden a ser sistemas estables que absorben estas distorsiones y enseguida recuperan su forma anterior. Incluso la construcción de una presa podría suponer solo un cambio temporal, revertido una vez que la presa se llenara y desbordara. Quién sabe si se podría lanzar río abajo una botella lo suficientemente resistente como para que no se rompiera en los rápidos del río. También podrías intentar desviar el río por un nuevo curso que no desemboque en el mismo punto del mar. Pero ninguna de estas soluciones es sencilla.

Tratar de influir en el futuro lejano se parece mucho a influir en un río corriente abajo, cuando está lejos de su fuente. Muchos aspectos de nuestro mundo constituyen sistemas estables a escala local que absorben las pequeñas distorsiones. Por ejemplo, si construyes un montículo, es posible que la lluvia se lo lleve. Del mismo modo, si abres una tienda de bocadillos en tu ciudad, puede que cierre otra tienda, con lo que el número de establecimientos quedaría inalterado.

En cualquier caso, lo cierto es que en el mundo hay muchos sistemas estables posibles, así que cabría la esperanza de dar a un sistema estable un «empujoncito» para que cambie y se estabilice con otra configuración, como en la analogía del cambio de curso del río. Por ejemplo, cabe la posibilidad de que, si hay bastantes clientes que comen en más sitios de bocadillos durante un tiempo, se aficionen al producto y lo pongan de moda, con lo cual habría mercado y lugar para nuevas tiendas de bocadillos. No obstante, las modas pueden ser volátiles, de manera que otra moda puede, a su vez, desplazar a los bocadillos en beneficio de otros productos de comida. Asimismo, puede resultar difícil identificar los sistemas estables y, todavía más, encontrarse cerca de un punto de inflexión en el cual un pequeño esfuerzo pueda comportar un gran cambio.

En política, la alianza de un conjunto de grupos puede ser, en cierta medida, estable. Por consiguiente, puedes intentar propiciar la formación de un nuevo conjunto de coaliciones que compartan tus tendencias políticas. También existe la opción de unirse a un movimiento social para luchar por la priorización de determinados valores en contextos sociales y políticos. Sin embargo, si estos valores responden a las circunstancias, pueden también formar parte de sistemas estables y, por tanto, resistirse al cambio. Sí, es probable que hayas asistido a cambios recientes en los valores expresados políticamente, pero quizá no son tanto la consecuencia de verdaderas transformaciones en los valores sino respuestas a cambios temporales de las modas y las circunstancias.





Algunas cosas se acumulan por naturaleza. Es el caso de muchas innovaciones (como los avances académicos y las decisiones de diseño técnico u organizativo) que son lo suficientemente generales y robustas como para conservar su valor más allá de los contextos sociales actuales. Si dichas innovaciones son también lo bastante grandes y sencillas, es posible que el mundo las acabe acumulando en gran cantidad. En tal caso, uno podría influir en el futuro añadiendo una innovación a la lista. Pero podría ocurrir que, de no haber hecho el descubrimiento, otra persona lo hubiera logrado un poco más tarde en nuestro lugar. En ese contexto, el futuro lejano no se presentaría muy distinto como resultado de nuestra aportación.

La economía mundial en su conjunto acumula (es decir, crece) sobre todo a través de la innovación. Podrías intentar ayudar a algunos de sus componentes a crecer más rápido que otros, pero, si existe un equilibrio natural entre componentes, las fuerzas correctoras podrían revertir nuestros cambios. Podrías intentar influir en la tasa global de crecimiento (por ejemplo, ahorrando más), aunque de ese modo quizás solo conseguirías que el mismo futuro llegara un poco antes.

Otra opción sería intentar ahorrar recursos y dedicarlos a un plan futuro. Esta podría ser una opción atractiva, en especial si tenemos en cuenta que, como casi siempre ha sucedido a lo largo de la historia, las tasas de rendimiento de la inversión han sido superiores a las tasas de crecimiento económico. En este caso, sin que deje de ser ínfima, tu influencia puede ser mayor en el futuro que en el presente, al menos si dispones de un modo fiable para influir en cómo se gastará en el futuro el rendimiento de nuestras inversiones.

Por ejemplo, podrías ahorrar e intentar vivir una larga vida; o elaborar un plan y aleccionar a tus hijos y nietos para que lo ejecuten. Podrías crear y financiar una organización longeva comprometida con unos fines específicos. De un modo análogo, podrías tratar de asumir el control de una institución de larga duración ya existente, como una iglesia o un gobierno, para conseguir que se comprometa con tu plan.

Si crees que hay muchas otras personas interesadas por lo mismos resultados que tú, pero que están luchando por otros objetivos empujados por incentivos individuales, tal vez puedas intentar coordinarte con ellos mediante contratos o instituciones más amplias. Por otra parte, si crees que la mayor parte del mundo está de acuerdo contigo, podrías incluso tratar de crear un gobierno más fuerte a escala mundial y hacer que se implique en tu plan. Aunque acuerdos como estos quizás sean difíciles de organizar.

Algunas cosas, como las piedras, los edificios o las constituciones, tienden a ser longevas por naturaleza. Cabría pues pensar que transformar algo así comportaría cambios a largo plazo. De entre las cosas que nos caracterizan, de las más duraderas son las maneras de coordinarnos los unos con los otros. Por ejemplo, nos coordinamos para vivir cerca de los mismos lugares, hablar las mismas lenguas y compartir las mismas leyes y los mismos gobiernos. Dado que modificar

estas cosas es difícil, los cambios logrados pueden durar más. Por la misma razón, no obstante, sería muy improbable que te encontraras en situación de influir de manera significativa en ese tipo de cuestiones.

En la biosfera, durante miles de millones de años el modo más común (con mucho) de influir en el futuro lejano ha consistido en esforzarse por tener hijos que crezcan y, a su vez, tengan más hijos. ¿Cómo se ha conseguido? Sobreviviendo, debilitando a los rivales, recolectando recursos, haciendo gala de buenas aptitudes para atraer parejas y criando hijos. Ese comportamiento ha sido la principal estrategia humana durante miles de años.

Este predominio abrumador de las estrategias biológicas habituales indica que, de hecho, constituyen vías de relativa eficacia para influir en un futuro lejano. Parece ser que las fuerzas correctivas que absorben distorsiones de este tipo son, en promedio, relativamente débiles. Tiene sentido si lo que hace nuestro complejo mundo está siempre tratando de coordinarse para corresponder a la complejidad que alberga en cada momento. En tal caso, cuanto más puedas llenar el mundo de elementos como tú, más intentará el mundo adaptarse para corresponderse con dichos elementos y, por tanto, más espacio dará, en un futuro lejano, a elementos que son como tú.

Destruir tiende a ser más fácil que crear, lo que nos empuja a buscar modos de alcanzar objetivos a largo plazo mediante la destrucción. Por suerte, los enfoques destructivos entran en conflicto, en parte, con los enfoques basados en «crear más elementos semejantes a nosotros». Sí, intentamos matar a nuestros rivales y en ocasiones las sociedades van a la guerra, pero en general la destrucción indiscriminada rara vez ayuda a que los individuos tengan descendencia.

Llegados aquí y habiendo revisado algunas cuestiones básicas sobre cómo influir en el futuro, ¿qué podemos decir sobre hacerlo en un futuro de robots?

ROBOTS TRADICIONALES

Las máquinas llevan varios siglos reemplazando a humanos en tareas laborales y, desde hace setenta años, muchas de ellas están controladas por ordenadores. Pese a que las capacidades brutas de estos ordenadores han mejorado a un ritmo exponencial de muchos órdenes de magnitud, el ritmo de sustitución de los empleos humanos ha permanecido reducido y relativamente constante. Es plausible que esta tendencia sea el producto de la enorme variación de los empleos humanos en cuanto al poder de computación necesario para desempeñarlos de un modo adecuado. Esto indica que la tasa de sustitución futura de empleos podría permanecer baja y relativamente constante incluso si la potencia de los ordenadores continúa mejorando de un modo exponencial de muchos más órdenes de magnitud.





La inteligencia artificial (IA) es un área de la investigación en computación cuyos expertos intentan entrenar a ordenadores para que lleven a cabo tareas que antes solo podían hacer humanos. Cuando los investigadores en IA han dado un paso más allá para presentar en público previsiones sobre el ritmo general de la investigación futura en esta materia (como promedio de todas las subáreas del IA), su estimación media predecía que se alcanzarían unas capacidades de nivel humano en alrededor de tres décadas. Esta estimación de treinta años ha permanecido constante durante más de cinco décadas, de modo que, a día de hoy, podemos decir que, en sus primeros veinte años, los pronósticos de este tipo estaban bastante errados. Por su parte, los investigadores, que en lugar de dar ese paso más allá se han limitado a responder encuestas, han añadido diez años a la estimación (Armstrong y Sotala 2012; Grace 2014).

No obstante, los expertos en IA son mucho menos optimistas cuando se les pregunta por aquellos temas sobre los que deberían saber más: el progreso reciente en la subárea de la IA que más dominan. En mis nueve años (1984-1993) de investigador profesional en IA he mantenido encuentros informales con otros expertos experimentados en IA y les he preguntado qué progresión habían notado en su subárea en particular durante los últimos veinte años. Suelen decir que solo han constatado entre el 5% y el 10% del progreso necesario para alcanzar capacidades de nivel humano en su subárea. También es habitual que no hayan observado una aceleración significativa de dicho progreso en este periodo (Hanson 2012). A este ritmo, cabe esperar que la subárea mediana de la IA tarde entre dos y cuatro siglos en alcanzar capacidades humanas. Me inclino más bien a creer en esta última estimación (en lugar de la habitual y presuntuosa de cuarenta años), puesto que se basa más directamente en información que estas personas dominan más.

Con todo, aunque tarden muchos siglos, es plausible que los robots acaben por hacer la práctica totalidad de trabajos necesarios. En ese momento, la tasa de crecimiento total de la economía podría ser muy superior: el volumen de la economía podría más o menos duplicarse cada mes, en lugar de cada quince años, como ocurre en nuestra era. Llegados a tal punto, los ingresos de los seres humanos no podrían proceder de su capacidad de trabajo, sino de otros activos, como las acciones, las patentes y la propiedad inmobiliaria. Pese a que el valor de los activos se debería duplicar tan rápido como la economía, los humanos que carecieron de los activos suficientes, de un seguro o de derecho a prestaciones públicas podrían morir de hambre.

En principio, a partir de un mundo dominado por los humanos podría evolucionar, de un modo gradual, un mundo futuro dominado por los robots. La lenta sustitución del trabajo humano por máquinas no requiere un cambio excesivo de la naturaleza, las divisiones y la distribución básicas de las ciudades, los países, las industrias, las profesiones y las empresas. Es decir, las máquinas podrían encajar en las casillas sociales antes ocupadas por los humanos. Huelga decir que, cuando menos, las industrias que antes educaban y formaban a los humanos serían reemplazadas por nuevas industrias que diseñarían, mantendrían y fabricarían robots.

Sin embargo, podría también haber cambios mucho mayores en la organización de una sociedad robótica si, como parece plausible, las máquinas fueran lo bastante diferentes de los humanos, en términos de costes relativos o productividad, como para volver eficientes unos sistemas sustancialmente distintos. Un modo razonable de estimar los costes, la productividad y las estructuras a gran escala de una sociedad de robots es observar la distribución de características similares en el *software* que hemos creado y usado durante muchas décadas. Aunque es posible que el *software* del mañana parezca muy distinto del existente hasta ahora,

a falta de buenas razones para esperar determinados tipos de cambios, puede que el *software* conocido sea aún la mejor pista del *software* por conocer. Por consiguiente, resulta razonable esperar que la estructura de una sociedad de robots se asemeje a la estructura de nuestro sistemas de *software* más grandes; en especial, aquellos extendidos por muchas empresas de muchos sectores distintos.

¿Cómo se puede intentar influir en un futuro robótico así? Un enfoque directo consistiría en acumular recursos y confiárselos a las organizaciones adecuadas. Por ejemplo, si solo quisieras influir en el futuro para conseguir comodidad y felicidad para ti y tus descendientes, podrías intentar vivir mucho tiempo y guardarte tus recursos, o bien dárselos a tus descendientes para que los gasten como quieran.

Tal vez, te desagrada la naturaleza o la estructura generales que, probablemente, tendría una sociedad robótica en un mundo descentralizado con escasa coordinación mundial. En este caso, si compartes este sentimiento intenso con un número de personas suficiente, puedes intentar promover unas instituciones políticas a gran escala y alentarlas a adoptar unas normas lo suficientemente restrictivas. Con una supervisión lo bastante detallada de las vulneraciones de la ley y unos castigos lo bastante duros para aquellos juzgados culpables, dichas normas podrían forzar los cambios deseados. Si los beneficios que las organizaciones pudieran obtener de sistemas más descentralizados tuvieran la magnitud suficiente, podría ser necesaria una regulación de ámbito mundial.

Sería plausible que las estructuras de la sociedad robótica del futuro surgieran de una evolución gradual en el tiempo, a partir de las estructuras más automatizadas de nuestra sociedad actual. En este contexto, cabría esperar que pudiéramos influir en las estructuras futuras mediante nuestras elecciones actuales de las estructuras de nuestra sociedad que hacen un uso más intensivo de ordenadores. Por ejemplo, si se prefiere que la sociedad robótica futura tenga unos mecanismos de seguridad relativamente descentralizados, se podría fomentar el desarrollo y la adopción, en la actualidad, de mecanismos de seguridad relativamente descentralizados. Asimismo, quien tema unos niveles elevados de concentración empresarial en un sector determinado de la futura sociedad robótica, puede intentar promover desde ahora unos niveles bajos de concentración de empresas en dicho sector.

ROBOTS BASADOS EN LA EMULACIÓN

Como explicaba, es posible que un mundo futuro esté repleto de robots similares a los que llevamos décadas fabricando. Sin embargo, también es posible, al menos durante un tiempo, llenar el futuro de robots muy distintos: las emulaciones cerebrales.

Las emulaciones cerebrales, conocidas también como «copias» o «ems», han sido una constante tanto de la ciencia ficción como del futurismo tecnológico desde hace décadas. Para hacer una emulación cerebral, se toma un cerebro humano, se escanea para registrar sus características y conexiones celulares concretas y a continuación se elabora un modelo informático que procesa las señales de acuerdo con esas mismas características y conexiones. Una emulación lo bastante buena cuenta con un comportamiento general muy próximo al de su original humano en términos de señales de entrada-salida. Podríamos hablar con ella y convencerla de que hiciera trabajos útiles.



Al igual que los humanos, las emulaciones recordarían un pasado, serían conscientes de un presente e imaginarían un futuro. Las emulaciones pueden ser o estar felices o tristes, entusiastas o cansadas, temerosas o esperanzadas, orgullosas o avergonzadas, creativas o faltas de originalidad, compasivas o frías. Pueden aprender y tener amigos, amantes, jefes y colegas. Aunque sus características psicológicas pueden apartarse de la media humana, suelen estar cerca de su rango de variación.

Parece probable que las tres tecnologías necesarias para crear emulaciones —es decir, la computación, el escaneo y la modelización celular— estarán listas en el plazo aproximado de un siglo, mucho antes de la estimación anteriormente citada de dos a cuatro siglos para que los robots comunes hagan casi todos los trabajos. Por lo tanto, las emulaciones podrían aparecer en un momento con alta demanda de trabajadores humanos y, por tanto, con alta demanda de emulaciones que los sustituyan.

Recientemente publiqué un libro, *The Age of Em: Work, Love, and Life when Robots Rule the Earth* (La era Em. Trabajo, amor y vida cotidiana cuando los robots gobiernen el mundo, Oxford University Press, 2016), que ofrece una perspectiva pormenorizada de un mundo dominado por emulaciones, al menos en su forma inicial, la que haría aparición poco después de la transición a un mundo de emulaciones. A continuación voy a resumir algunos de esos pormenores.

Mi análisis de una era temprana de emulación plantea un panorama que para muchos resultará perturbador y desconocido. La población de emulaciones no tardaría en dispararse hacia cifras de billones, lo que reduciría los sueldos a niveles cercanos a la subsistencia, con las emulaciones dedicando casi todas sus horas de vigilia a trabajar y a la economía, duplicando su producción a intervalos de un mes o menos. La mayoría de emulaciones serían copias hechas a partir de solo un millar de humanos muy inteligentes, concienzudos y productivos. La mayoría de las emulaciones tendría una edad de productividad máxima de cincuenta o más años. Por añadidura, la mayor parte de emulaciones serían también copias fabricadas para efectuar una tarea a corto plazo y desaparecer al terminarla.

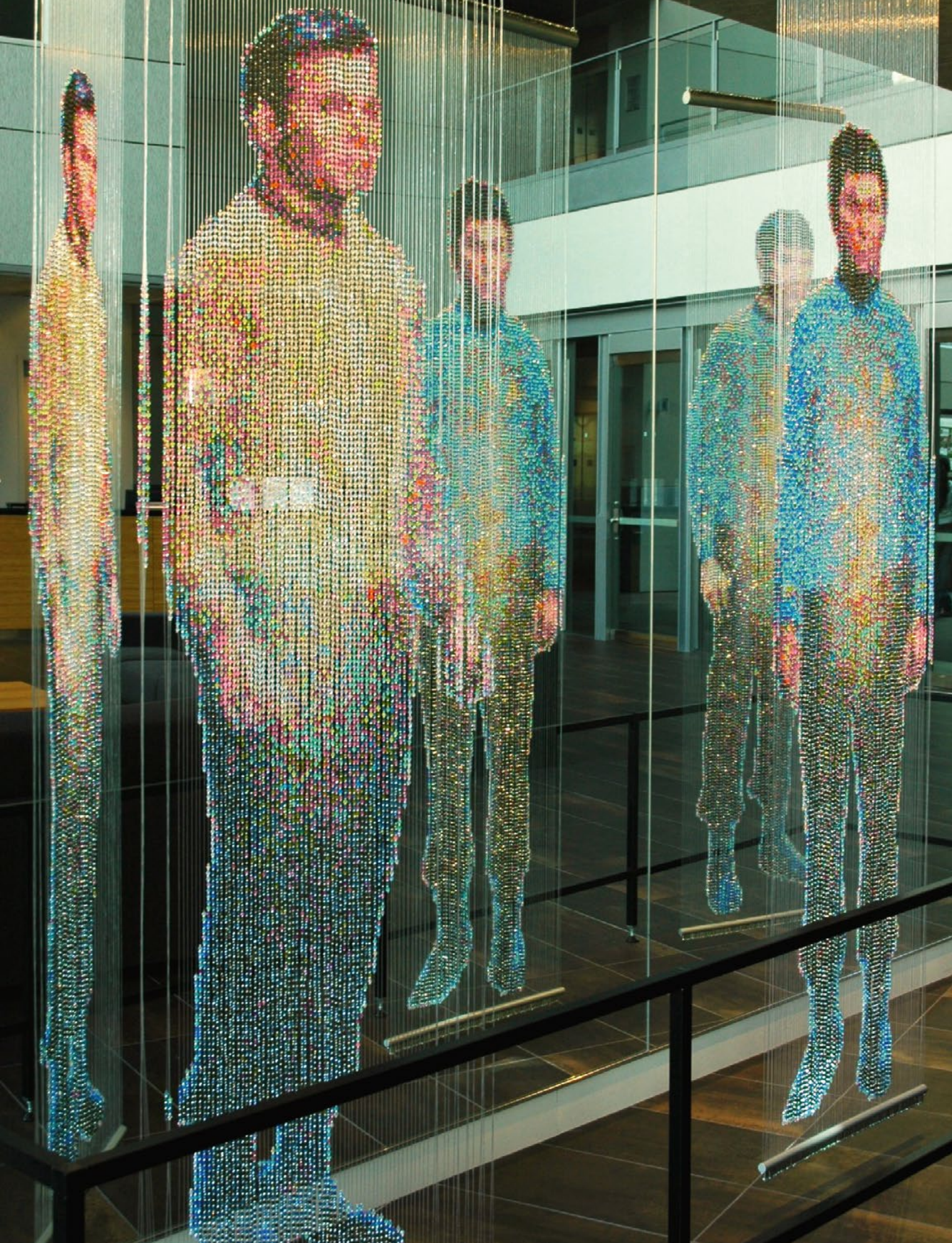
Las emulaciones se hacinarían en algunas ciudades de gran altura llenas a rebosar de *hardware* informático muy caliente y pasarían su tiempo libre en la realidad virtual, que para la mayoría sería también el lugar de trabajo. La realidad virtual de las emulaciones tendría una calidad espectacular y contarían con unos preciosos cuerpos virtuales siempre libres de hambre, frío, suciedad, dolor y enfermedades. Dado que la emulación típica funcionaría cerca



«Segunda estrella a la derecha, todo recto hacia el mañana.»

Kirk, uno de los personajes de la saga *Star Trek*, en *Star Trek VI. Aquel país desconocido* (1991)

La instalación de Devorah Sperber, que se encuentra en las oficinas del Studio D de Microsoft en Redmond, Washington, recrea imágenes del episodio «Espejo, espejito» de la serie *Star Trek. La serie original* (1967)





de mil veces más rápida que los humanos, su mundo les parecería más estable a ellas que el nuestro a nosotros.

Las emulaciones crearían copias propias que efectuarían tareas a corto plazo y desaparecerían después. Después de una carrera subjetiva de, quizás, un siglo o dos de duración, las mentes de las emulaciones habrían perdido flexibilidad y ya no competirían bien con las mentes más jóvenes. Estas emulaciones se podrían jubilar para dedicarse por tiempo indefinido a una vida de ocio a una velocidad más baja.

La facilidad para fabricar copias de emulaciones facilitaría la preparación. Una de ellas podría concebir un *software* o un diseño o visión artísticos y, acto seguido, multiplicarse en un ejército de emulaciones para ponerlos en práctica. Sería mucho más frecuente que, a pesar de superar el presupuesto previsto, los grandes proyectos se completaran dentro de su plazo, gracias a la aceleración de aquellas emulaciones encargadas de las partes más atrasadas de la obra. Se podría entrenar a una emulación para hacer un trabajo y, a partir de la emulación entrenada, hacer muchas copias. Por tanto, los mercados laborales de emulaciones se parecerían más a nuestros mercados de productos de hoy, dominados por algunos grandes proveedores.

Entre las emulaciones habría más desigualdad de la que existe en la actualidad, tanto por la variabilidad de sus velocidades como porque su mayor vida útil permitiría que se acumulara su desigualdad en resultados. Las emulaciones se dividirían entre castas en función de su velocidad, y las más veloces gozarían de un estatus más alto. Por otra parte, es probable que las democracias de emulaciones se sirvieran del voto ponderado según la velocidad y que sus dirigentes funcionaran a una velocidad mayor que sus subordinados, lo que facilitaría la coordinación de organizaciones de mayor magnitud. De igual manera, las organizaciones de emulaciones podrían también aplicar nuevos métodos de gobierno, como los mercados de decisión y las subastas combinatorias.

Cada emulación se sentiría muy apegada a su clan de copias, todas ellas descendientes del mismo humano original. Los clanes de emulaciones podrían autogobernarse y negociar con otros clanes las normas jurídicas a aplicar en caso de disputa entre sí. De igual manera, dichos clanes podrían brindar a los suyos un asesoramiento continuo basado en las experiencias vitales de miembros del clan similares.

Para permitir relaciones románticas en un contexto de demanda desigual de emulaciones trabajadoras macho y hembra, el género menos demandado podría funcionar a menor velocidad y acelerarse periódicamente para conocer a compañeros más rápidos. Las emulaciones rápidas con un cuerpo físico de tipo robótico tendrían cuerpos proporcionalmente más pequeños. Una emulación con una velocidad típica de mil veces la del ser humano tendría dos milímetros de altura y, desde su punto de vista, la Tierra parecería mucho más grande. En cuanto a los viajes físicos a larga distancia, en su mayor parte consistirían en viajes electrónicos mediante teletransportador, como en *Star Trek*, con el debido cuidado para evitar el robo de mentes.

Las ciudades de emulaciones probablemente serían poco habitables para los humanos ordinarios, quienes, al controlar la mayor parte del territorio restante del planeta, vivirían en su mayoría una cómoda jubilación gracias a las ganancias de sus inversiones en la economía de las emulaciones. Pese a que las emulaciones mismas gozarían de sobrada capacidad para comprar el resto de la Tierra, ni siquiera se lo plantearían, más allá de garantizar el suministro de energía, materias primas y refrigeración para las ciudades de emulaciones. Del mismo modo que los



humanos raramente ejecutamos a nuestros jubilados y les arrebatamos sus posesiones, es razonable esperar que las emulaciones tal vez dejaran en paz a la humanidad una vez inactiva.

Es probable que, a largo plazo, el principal peligro tanto para los humanos como para la naturaleza fuera la inestabilidad de la civilización de las emulaciones, en forma de guerras o revoluciones. Las emulaciones, que funcionarían mil veces más rápido que los humanos, podrían reproducir unos cuantos milenios de historia en unos pocos años naturales. Como las lentas emulaciones jubiladas afrontarían unos riesgos similares, serían aliados que ayudarían a los humanos a promover la estabilidad en la nueva civilización.

EL LEGADO EN LOS ROBOTS EMULADORES

Hasta ahora, he abordado con brevedad algunas de las principales vías para influir en un futuro robótico en general. ¿En qué cambia esta situación en cuanto a los robots basados en emulaciones?

La diferencia más evidente es que, dado que cada emulación es el resultado de escanear un humano en concreto, los humanos pueden tener la expectativa de ejercer una gran influencia sobre la emulación resultante de su propio escaneo. Los padres y abuelos de dichos humanos podrían, asimismo, guardar también esperanzas en este sentido. Este tipo de influencia es bastante similar a la resultante de la estrategia típica y hasta ahora muy efectiva en la biosfera: promover la existencia de futuros seres que compartan muchas de nuestras características.

Otra gran diferencia sería que, dado que las emulaciones se asemejan mucho a los humanos, su inserción en las diversas posiciones sociales de la sociedad humana anterior podría ser mucho más fácil y directa. Cuando lleguen las emulaciones, hay razones para esperar menos cambios inmediatos y a gran escala en la naturaleza, las divisiones y la distribución básicas de las ciudades, los países, las industrias, las profesiones y las empresas. En consecuencia, resulta plausible que las inversiones actuales para influenciar ese tipo de instituciones sociales puedan gozar de una mayor longevidad durante la era de las emulaciones. Sin embargo, en el mundo de las emulaciones, los sistemas y las instituciones sociales también serían susceptibles de cambiar con el paso del tiempo, del mismo modo que habrían cambiado con el tiempo si los humanos hubieran seguido dominando la Tierra.

Cabe esperar, pues, que durante la era de las emulaciones, los robots emuladores sigan desarrollando las capacidades de los robots tradicionales, no basados en emulaciones. Al final, estos últimos podrían alcanzar capacidades superiores a las de las emulaciones en casi todos los trabajos. Esto marcaría, quizás, el fin de la era de las emulaciones. Que los robots tradicionales acabaran por sustituir a las emulaciones es menos obvio que la sustitución de los humanos; no en vano, las emulaciones tienen

más posibilidades de progresar con el tiempo que nosotros. En todo caso, la sustitución de las emulaciones por parte de los robots tradicionales parece un escenario a considerar.

En comparación con una rápida sustitución de los humanos por los robots tradicionales, un escenario en el cual las emulaciones sustituyeran primero a los humanos para, después, ser a su vez sustituidas por los robots clásicos parecería más proclive a otorgar a los humanos de nuestra era un mayor impacto sobre el futuro lejano. Esto se debe a que el primer escenario contiene una transición más traumática, con una mayor probabilidad de que los sistemas sociales existentes sean reemplazados de A a Z por otros más adecuados para los robots tradicionales. Por el contrario, el segundo escenario parece más susceptible de un cambio gradual que herede más estructuras procedentes de los sistemas de hoy.

A medida que avance la economía de las emulaciones, es probable que tenga cada vez más éxito a la hora de encontrar modificaciones mayores y más útiles de los cerebros emulados. Sin embargo, las modificaciones que se buscarían serían, sobre todo, aquellas que pudieran incrementar la productividad de las emulaciones en sus trabajos y las relaciones ya existentes. Asimismo, dado que es probable que la mayoría de modificaciones sean pequeñas, las mentes de las emulaciones, sus empleos y otras relaciones sociales evolucionarían de un modo gradual hacia nuevos sistemas.

En cuanto a las grandes modificaciones cerebrales de las emulaciones, es muy posible que fueran acompañadas tanto de una capacidad mejorada para separar con éxito las distintas partes de los cerebros de las emulaciones como de mejores teorías sobre su funcionamiento. Los desarrolladores de los robots tradicionales se verían, pues, alentados a incluir en estos más subsistemas semejantes a partes del cerebro de las emulaciones, que encajarían de un modo más fácil y natural en la economía de las emulaciones.

En definitiva, una transición desde las emulaciones hasta los robots tradicionales podría resultar menos traumática y más progresiva, al incluir y dar continuidad a más elementos de las mentes de las emulaciones y a sistemas más amplios propios de su sociedad. Puesto que, a su vez, una sociedad de emulaciones habría dado continuidad a más sistemas propios de la sociedad humana anterior, los humanos de hoy y nuestros sistemas nos enfrentaríamos a una trayectoria más progresiva de cambio futuro. De este modo, gozaríamos de más vías para influir en el mañana y daríamos pie a un futuro más conforme a las personas e instituciones de hoy.





BIBLIOGRAFÍA

- Armstrong, S. y Sotala, K., «How We're Predicting AI—or Failing to», en Romportl, J., Ircing, P., Zackova, E., Polak, M. y Schuster, R. (eds.), *Beyond AI: Artificial Dreams*, Pilsen, Universidad de Bohemia Occidental, 2012, pp. 52-75.
- Grace, K., «MIRI AI Predictions Dataset», en AI Impacts, 20 de mayo de 2014. <http://aiimpacts.org/miri-ai-predictions-dataset/>
- Hanson, R., «AI Progress Estimate», en el blog *Overcoming Bias*, 27 de agosto de 2012. <http://www.overcomingbias.com/2012/08/ai-progress-estimate.html>



CANAL OPENMIND



LEE EL LIBRO COMPLETO

El próximo paso: la vida exponencial

[+]

ARTÍCULOS RELACIONADOS

Inteligencia artificial de beneficios probados

[+]

**Innovación: cambiar el rostro
de la discapacidad**

[+]

**Cuestiones éticas derivadas
del mejoramiento humano**

[+]

OTROS LIBROS

