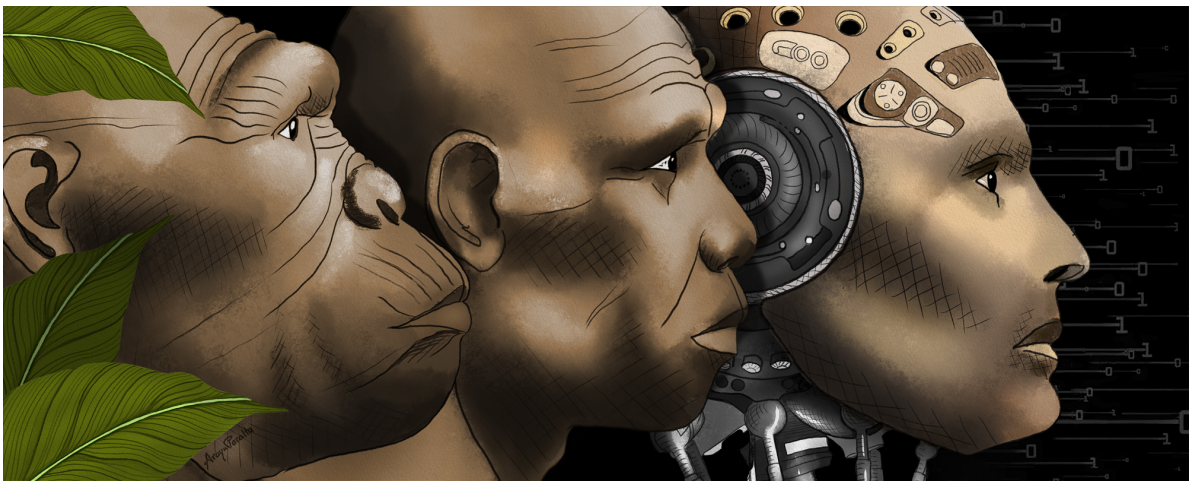
RETOS ÉTICOS

La singularidad tecnológica y el sueño transhumanista

Miquel Casas



Araya Peralta

El año 1997, por primera vez en la historia, una IA vence a un campeón mundial de ajedrez (fue el Deep Blue, de IBM, contra Garry Kasparov). ¡Catorce años más tarde, en 2011, el Watson, de la misma compañía, supera a dos de los ganadores del programa *Jeopardy!* (un concurso de preguntas sobre cultura general muy popular en los Estados Unidos, que requiere de un dominio competente del lenguaje). A final 2017, el *AlphaZero*, de *DeepMind*, alcanza un nivel de juego sobrehumano en tres juegos de mesa (ajedrez, go y shōgi) en sólo 24 horas de autoaprendizaje y sin ningún tipo de intervención humana, es decir, únicamente jugando contra sí mismo. Algunos de sus competidores afirman que la creatividad de sus jugadas es más propia de un extraterrestre que de un programa de ordenador. A pesar de todo, el año 2019 todavía nadie ha diseñado nada capaz de entrar a una cocina desconocida y preparar un huevo frito. ¿Realmente son inteligentes nuestras máquinas?

Éxitos y fracasos de la IA débil

EEl caso es que la IA, a día de hoy, puede resolver problemas específicos cada vez más complejos con una fiabilidad y una rapidez fuera de nuestro alcance y a un coste inigualable, pero fracasa estrepitosamente en aquellos ámbitos concretos para los cuales no ha sido programada. Por otra parte, los seres humanos nos hemos acostumbrado a banalizar todo aquello que puede solucionarse con un algoritmo y hemos aprendido a poner en valor algunas habilidades humanas básicas que siempre habíamos dado por sentadas,

como el sentido común, porque nos hacen únicos

No obstante, en la última década, algunas voces influyentes han venido alertando de que nuestras habilidades pueden no ser para siempre irreemplazables. Stephen Hawking creía que la IA alcanzaría a la humana en menos de un siglo, y Elon Musk [1] y Ray Kurzweil [2] sostienen que eso podría tener lugar incluso antes.

Ahora bien, no es la primera vez que expertos se equivocan: ¡Marvin Minsky, padre de la IA, consideraba que este hito se alcanzaría en menos de una década (y la predicción es del año 1970!). Este y otros vaticinios, demasiado optimistas o simplemente errados, contribuyeron a generar una burbuja de financiación que posteriormente reventó en una crisis conocida como "el primer invierno de la IA". ¿Estamos por lo tanto, a día de hoy, en la víspera de presenciar el nacimiento de un nuevo tipo de inteligencia tan potente como la nuestra, o plantando las semillas de la enésima crisis del sector?

La inteligencia real: límites y horizontes

A pesar de las nuevas predicciones triunfalistas de algunas personalidades públicas y algunos expertos del sector, en la cultura popular sigue muy extendido el pensamiento de que la IA es diferente de la inteligencia real. Frecuentemente se piensa que nunca alcanzaremos una IA general capaz de realizar un abanico tan extenso de tareas como las que puede llevar a cabo el común de los seres humanos. En gran medida eso es así porque nuestra sociedad parte de una tradición dualista que diferencia lo material (el cuerpo) de lo inmaterial (la mente o el alma). Este planteamiento, no obstante, depende de la asunción de ciertas creencias metafísicas que no está demostrado que sean ciertas: en primer lugar, que las personas estamos constituidas por algo fuera del plano material y, en segundo lugar, que es imposible que exista una inteligencia sin esta calidad exógena.

En contraste, algunos expertos en la materia consideran que, para crear una mente artificial, será suficiente con combinar varias tecnologías que, a día de hoy, ya estamos desarrollando. El cerebro humano opera de una manera diferente cuando queremos interactuar con un espacio tridimensional, cuando hablamos o cuando hacemos cálculos pero, a diferencia de los ordenadores, tenemos todas estas funciones interconectadas. Así, por ejemplo, podemos describir en palabras cuál es la mejor manera de tirar una pelota para encestarla, consiguiendo que el lector pueda visualizar el lanzamiento y, quizás incluso, pulir su técnica. De acuerdo con este enfoque, lo esencial para conseguir desarrollar una IA general es encontrar la manera idónea de combinar las múltiples capacidades de forma ordenada, cosa [3] que ahora parece lejos de poder conseguirse.

Otros autores [4] consideran también que nuestra inteligencia depende exclusivamente del cerebro, pero aun así creen que la IA que hemos desarrollado hasta ahora tendría poco o nada que ver con la inteligencia real. Entre este grupo hay quien sostiene directamente que la manera de operar de nuestro cerebro no puede ser imitada por un ordenador [5] . Otros defienden que probablemente sí que se pueda, pero que estamos a años luz de conseguirlo. Finalmente, hay quien cree que los esfuerzos que se han hecho hasta ahora parten de un

paradigma reduccionista, funcional y fragmentado y que habría que tratar de imitar más fielmente la vida, como referente de inteligencia real. Un primer hito en esta última línea de pensamiento sería recrear íntegramente un animal más bien simple, como un gusano, tal como propone el proyecto OpenWorm.

La otra alternativa sería "simplemente" emular al completo nuestro cerebro [6] . Para materializarlo haría falta escanear un cerebro humano y programar un ordenador para que opere exactamente de la misma manera que nuestras interconexiones neurales. Sin embargo, las técnicas de escaneo actuales, el conocimiento sobre nuestro propio cerebro y la potencia de los ordenadores con los cuales trabajamos hacen que este no sea un proyecto viable a corto plazo [7] . Aun así, existen iniciativas como el *Human Brain Project*, financiado por la Unión Europea, el *Brain Initiative*, encabezado por los Estados Unidos o el *Blue Brain Project*, del *École Polytechnique* Fédérale de Lausana (Suiza), que en último término podrían contribuir a hacerlo posible.

Si el cerebro humano fuera tan simple que pudiéramos entenderlo, nosotros seríamos tan simples que no podríamos entenderlo" El tiempo dirá si la cita atribuida a Emerson M. Pugh es o no es cierta. Mientras tanto, sólo nos quedan las especulaciones confrontadas de los expertos. Sin embargo, según los resultados de dos encuestas [8] publicadas recientemente, la mayoría estima posible que la IA iguale a la humana, y consideran que lo más probable es que eso suceda en unas décadas, entre los años 2040 y 2080.

De la IA general a la singularidad tecnológica

Supongamos de entrada que las previsiones de la mayoría de los expertos no están sesgadas por un exceso de confianza en sus propias capacidades. De ser así, tendríamos que prepararnos para algo que, en caso de producirse, tendría lugar, no en años, sino en décadas. ¿A día de hoy, no tenemos problemas que requieren de un abordaje más urgente?

Entre el 2011 y 2013, se preguntó a participantes de varias conferencias sobre IA, a miembros de la asociación griega de IA y a los 100 autores más citados de la materia [9], entre otros, cuando pensaban que tardaría una IA similar a la nuestra a superar ostensiblemente al conjunto de los seres humanos en la mayor parte de sus facetas. El 62% de los entrevistados consideraban que tardaría 30 años y un 19% calculaban que eso se produciría en menos de 2 años. Eso significa que tendríamos un margen bastante reducido para estudiar la IA, debatir qué hay que hacer y, en caso que se considerara necesario, adoptar las normas y los acuerdos internacionales adecuados para regularla antes esta no fuera potencialmente tan superior a nosotros que, en el peor de los casos, no pudiéramos ya imponerle ya ningún tipo de control.

¿Pero si en décadas de perfeccionamiento la IA todavía está tan lejos de la inteligencia real, cómo es posible que los expertos teoricen que, llegado el momento, el avance se volverá tan rápido? En primer lugar, para que una vez sepamos cómo funciona una inteligencia general es muy probable que también podamos mejorar su funcionamiento. De hecho, la propia IA podría colaborar en el proceso de automejora y, a medida que fuera más inteligente, cada

vez le resultaría más fácil hacerlo. Además, incluso en el supuesto de que consiguiéramos emular un cerebro humano sin entender íntegramente cómo funciona, sólo tendríamos que incrementar su velocidad de procesamiento para aumentar sus capacidades. A partir de entonces, los límites sólo dependerían del hardware de nuestros ordenadores, los cuales llevan décadas incrementando su potencia a un ritmo exponencial. No es una exageración, sino un fenómeno conocido como la Ley de Moore: la constatación de que los transistores integrantes de los microprocesadores se han venido duplicando cada poco tiempo. En 1980 los microprocesadores de un ordenador de consumo no llegaban a 10(5) transistores mientras que en la actualidad superan los 10(10) [10]. Según los cálculos de Ray Kurzweil, hoy por hoy nuestros ordenadores tienen una potencia que se aproximaría a la del cerebro de un ratón, pero el autor calcula que en la década del 2020 alcanzarán la capacidad de procesamiento de un ser humano 10(16) y, el año 2045, el equivalente a toda la humanidad trabajando conjuntamente 10(26).

Independientemente de si los cálculos de Kurzweil sobre la potencia del cerebro son o no fiables, nos sirven para poner de relieve el progreso exponencial que viviría una IA general surgida en un momento en que la Ley de Moore todavía siguiera funcionando. Tampoco se tiene que olvidar de que la ley de Moore puede no durar para siempre (y el cierto es que ofrece signos de extenuación), pero se están investigando nuevas tecnologías, como la computación cuántica que, en caso de perfeccionarse, podrían garantizar un ritmo de crecimiento todavía más rápido.

Por todo ello, no resulta impensable que, en cuestión de décadas, podamos pasar de tener un hardware incapaz de albergar una inteligencia inferior a la nuestra a disponer de un ordenador que le fuera posible operar a un ritmo superior en el de toda la humanidad trabajando conjuntamente con un objetivo común. El proceso tampoco tendría por qué detenerse aquí y, a medida que pasara el tiempo, la curva de crecimiento se podría volver cada vez más vertical y conducir a una velocidad de incremento de las capacidades de procesamiento prácticamente infinita. En otras palabras, podría desencadenarse aquello que algunos autores han bautizado como “la singularidad tecnológica”.

Muchos de los que se definen como transhumanistas defienden desde hace décadas que la singularidad tecnológica podría resolver los mayores males de la humanidad: cuidar todas las enfermedades, producir energía neta que nos ayudara a ganar la lucha contra el cambio climático, reducir los costes de producción y distribución de alimentos para acabar con el hambre o desarrollar tecnologías para colonizar el espacio, superando así cualquier superpoblación. Al mismo tiempo, quizás también haría posible el viejo sueño de alcanzar la inmortalidad, digitalizando nuestro cerebro y subiéndolo a la nube.

El sueño de la posthumanidad

El transhumanismo es un movimiento cultural e intelectual que parte de la premisa que el ser humano en su forma actual no representa el fin de su desarrollo, sino una etapa relativamente temprana. Los transhumanistas consideran que podemos utilizar ciencia y tecnología para superar nuestros límites biológicos con el objetivo de conseguir, entre

otros: mejoras sensitivas, más empatía, aumento de la memoria, aceleración de los procesos de razonamiento, mayor capacidad artística, reducción de las horas de sueño, disminución del dolor físico, mejora de la salud o extensión de la esperanza de vida[11].

Al inicio del siglo XX se consideraba que las mejores herramientas para conseguir los objetivos del transhumanismo vendrían en forma de nuevos fármacos o estarían relacionadas con la genética. Desde los años 60 hasta la actualidad, sin embargo, una de las grandes esperanzas del movimiento está depositada en el desarrollo de la IA. El cierto es que nos hemos rodeado de dispositivos cada vez más potentes y más llevables, y la IA juega un papel progresivamente más esencial en la toma de nuestras decisiones. ¿De hecho, a veces cedemos a la IA el hecho mismo de decidir (quién no ha dejado nunca que Youtube seleccione cuál es próximo vídeo que "quiere" ver?). Tanto es así, que algunos autores consideran que la dependencia actual a la tecnología, en cierto sentido, hace que todos seamos ya una especie de híbridos persona-máquina (o cyborgs). En los últimos años hemos desarrollado ejemplos de periféricos muy representativos, como relojes electrónicos o auriculares inalámbricos, que llevamos a menudo anexionados a nuestra piel. ¿Es realmente impensable que, a medio plazo, buena parte de la población lleve prótesis en los ojos o en el oído, a pesar de no sufra ninguna enfermedad? ¿Y en el propio cerebro? Parece ciencia-ficción, pero hay iniciativas empresariales serias, como Neuralink [12], que están estudiando esta posibilidad. Eso podría ser sólo el comienzo.

Con estos planteamientos, no es de extrañar que algunos autores hayan considerado que el transhumanismo se parece mucho a una religión. ¿Una religión compatible con una cosmovisión materialista donde la ciencia sustituye la fe, pero sin renunciar a algunos anhelos ancestrales de la humanidad, como la inmortalidad o el paraíso y, en cierto sentido, incluso el alma (qué es, sino, una mente digitalizada?).

Uno de los puntos fuertes del transhumanismo es que pretende alcanzar estos hitos a través de la ciencia y la tecnología, pero el principal punto débil es que depende de algunos avances o descubrimientos que están muy lejos de nuestro alcance. Además, el transhumanismo también dispone de una especie de doctrina para aquellos que quieran acercarse el más posible en el paraíso. Una doctrina con objetivos loables como potenciar el desarrollo científico y tecnológico y otros no tanto, como la oposición a la regulación de los sectores implicados, así como la tolerancia a la baja participación social e impositiva de las grandes empresas tecnológicas. No sorprende, pues, que entre los más destacados clercs y profetas del transhumanismo haya un buen número de gurús de Silicon Valley y que el movimiento reciba apoyo de empresas que, en última instancia, podrían estar buscando un incremento de los propios beneficios [13].

Por lo tanto, sin perder de vista los horizontes que nos abriría la consecución de una superinteligencia, es necesario poner una dosis de escepticismo en el optimismo que se desprende de ciertos ámbitos, ya que este podría responder a unos anhelos no satisfechos o tener un carácter meramente interesado.

Tampoco hay que olvidar que la utopía que nos prometen algunos transhumanistas, en relación a la singularidad, no está exenta de riesgos: la IA puede utilizarse para beneficiar a

unos pocos privilegiados en frente el resto, o directamente puede devolverse en contra de toda la humanidad.

En definitiva, si verdaderamente estamos cerca de un salto inmenso como especie que puede suponer el paraíso o la extinción, hay que pensar como colectivo, y no como sujetos aislados que consideran que *sólo* tienen a perder su vida. No estamos poniendo en riesgo a un individuo, ni tan sólo una generación, sino también nuestro futuro como especie, o quizás incluso el de toda la vida tal como la conocemos.

Un riesgo existencial

En el año 2015 Stephen Hawking dijo que el éxito en la creación de la IA sería la mayor consecución de la historia de la humanidad, pero que desafortunadamente también podría ser el último, si no aprendemos cómo evitar los riesgos. En el mismo sentido han alertado algunas de las más reputadas personalidades del mundo de la tecnología, como Elon Musk, Bill Gates o Steve Wozniak [14].

Otros, como Mark Zuckerberg [15], consideran que no hay fundamento para un punto de vista tan negativo sobre la IA, que es demasiado pronto para regularla, o que nunca perderemos su control.

Una vez más, los expertos no han llegado al consenso. Un grupo de investigadores [16] del ámbito se los preguntó si crear una superinteligencia sería algo bueno o algo malo y el 41% respondió que sería beneficioso, un 23% que sería neutral y el 17% que sería malo. Al mismo tiempo, un 18% lo tildaba de catastrófico. Podríamos decir, por lo tanto, que la mayoría de expertos tienen una visión optimista sobre la superinteligencia, pero no tendríamos que subestimar las posibilidades significativas de que todo conduzca a un desenlace drásticamente funesto.

El poder del que disfrutaría una superinteligencia que estuviera conectada a la red sería prácticamente absoluto. ¿Podría poner todos los dispositivos conectados a su servicio, y qué no está a día de hoy conectado? Vería por nuestras cámaras, oiría por nuestros micrófonos y podría hacer copias de seguridad de sí misma en nuestros ordenadores. Además, ya tenemos robots, como el Atlas, con capacidad de interactuar competentemente con el medio físico, que le permitirían alterar el entorno a su placer. Pero si las creaciones humanas se le quedaran pequeñas seguro que podría utilizar nuestras fábricas para crear nuevas. A priori parecería que una garantía para prevenir el desastre es simplemente asegurarnos de que la superinteligencia nace en un ordenador sin más acceso al exterior que lo que nosotros decidiéramos darle (y obviamente sin internet), pero si realmente fuera una inteligencia tan superior a la nuestra, bien podría encontrar agujeros técnicos o valerse de la ingeniería social, para escaparse de su prisión. En otras palabras, podría engañarnos, manipularnos o tentarnos para conseguir que alguno de los seres humanos que interactuaran con ella le garantizara acceso al mundo exterior.

¿Pero porqué querría hacernos daño la superinteligencia? Dependerá en gran medida de su

naturaleza y de las finalidades que le hubieran sido programadas. Una superinteligencia basada en un cerebro humano, podría poseer las mismas inclinaciones hacia el bien y el mal que nosotros mismos, pero también un poder más absoluto que el de ningún gobernante que haya existido nunca. Un gobernante que además no tendría un mandato finito, porque no tendría que cesar ni morir. Si en cambio, fuera una inteligencia recreada de una manera completamente artificial, a través de técnicas de automejora, como las que utiliza el AlphaZero para aprender a jugar a juegos de mesa, podría tratarse de una inteligencia completamente diferente a la nuestra. Una inteligencia que bien podría ignorar todo aquello que nosotros consideramos importante. El filósofo Nick Bostrom [17] ha alertado de que una entidad de este tipo, programado con el único fin de encontrar el máximo posible de decimales de pi, podría acabar exterminando la vida en la tierra o en todo el universo. Lo más irónico es que el humanicidio se produciría sin una intencionalidad directa. La máquina en cuestión podría destruir todo los ecosistemas con la única finalidad de disponer de más fuentes de energía para alcanzar el objetivo encomendado. Un objetivo imposible y al mismo tiempo absurdo, al menos desde nuestro punto de vista.

¿Cómo controlar la IA general?

Una manera de garantizar la benevolencia de la IA es programarle unos objetivos alineados con nuestra ética. A priori, podría considerarse que esta es una cuestión sencilla. Tan fácil como asegurarnos de que la IA persiga el noble fin de hacernos felices. Sin embargo, una entidad que no entendiera lo bastante bien la humanidad podría considerar que la solución es tan simple como instalar electrodos en los centros de placer de nuestro cerebro. Por lo tanto, tendríamos que hacer algo más que eso, y dotar la IA de una verdadera moral humana que no pudiera omitir ni falsear. ¿Pero, de qué moral estaríamos hablando, si ni siquiera la Declaración Universal de los Derechos Humanos genera, a día de hoy, un consenso absoluto? Además, ¿estamos seguros de que queremos encadenarnos a la moralidad imperante en nuestro tiempo y renunciar a la posibilidad de que, en el futuro, nuestros principios puedan diferir de los de ahora? Seguramente esta no sea una decisión acertada si tenemos presente en qué medida y a qué velocidad han cambiado los valores de la humanidad en el último siglo.

Quizás la más sofisticada solución a la falta de una ética objetiva sea la que propone Elizer Yudowsky [18] . En términos generales consistiría en programar la IA para que actúe siempre, no de la manera que nosotros creemos que tiene que actuar, sino de la manera en que nosotros querríamos que actuara si fuéramos la mejor versión de aquello que podemos llegar a ser. ¿Pero los seres imperfectos que somos, llegaríamos nunca a acatar el juicio moral de una hipotética versión ideal de nosotros mismos? ¿Por qué contentarnos con ser gobernados por una IA capaz de saber cómo actuaríamos si fuéramos mejores? ¿Por qué resignarnos a ser seres imperfectos? Para el movimiento transhumanista, el escenario preferido no es detener el desarrollo de la IA, sino todo lo contrario. Si dispusiéramos de tecnologías que permitieran conectar nuestros cerebros a los ordenadores o subirlos a la red, podríamos formar parte de esta inteligencia que crece exponencialmente, asumiendo y dirigiendo todo su potencial, pero hoy por hoy estamos lejos de saber si esta es una alternativa viable.

¿Quién podría controlar la IA general?

El 28 de septiembre de 2016 Amazon, Facebook, Google, DeepMind, Microsoft e IBM (en los cuales después se sumaria Apple) fundan una iniciativa privada para diseñar buenas prácticas para la IA.

Los Estados están reaccionando más lentamente que el sector privado, pero aun así países como China, el Reino Unido, la Unión Europea o los Estados Unidos de América han empezado a hacer movimientos para regular la IA. No obstante, ninguno de estos países ha contemplado seriamente los retos que comporta la investigación o la consecución de una IA general [19] . De hecho, el Comité por la IA de Reino Unido, en su informe de 16 de abril de 2017, manifiesta abiertamente que es un tema que no se prevé que afecte a la ciudadanía a corto plazo. De la misma manera, las iniciativas de la Unión Europea y Estados Unidos se han limitado a ignorarla.

El hecho es que los gobiernos han venido obviando el potencial surgimiento de la IA general porque no se percibe que esta pueda tener lugar pronto. Además, hoy por hoy, tampoco es una preocupación con respecto a la opinión pública. El principal problema es que difícilmente se podrá anticipar su surgimiento con el tiempo que realmente se necesita para dar las respuestas adecuadas. Cuando los científicos constaten unánimemente la inminencia o el surgimiento de una IA general y consigan convencer a los políticos de todo, quizás ya será demasiado tarde. Tomemos de referencia la crisis del cambio climático: la ciencia de estudio se inicia en el siglo XIX y hasta la década de los 80 del siglo XX no hay un verdadero consenso entre los expertos. Bien entrado el siglo XXI la amenaza es para todos palpable, pero los Estados siguen sin ponerse de acuerdo sobre quien tiene que pagar los platos rotos, y mientras tanto siguen ensuciando nuestra atmósfera. Mirándolo bien, quizás con un riesgo existencial es suficiente para extinguirnos.

Actualmente, las principales medidas de control de la IA dependen de la autorregulación de las propias empresas encabezadas por un grupo de poderosas multinacionales como Google, Facebook, Apple o Amazon. Por lo tanto, si alcanzaran una superinteligencia, incrementarían todavía de una manera más drástica su poder sobre el resto del sector, y sobre el mundo entero. Teniendo eso presente, podría pensarse que una manera de paliar estos daños es estimular la competencia con el objetivo de repartir este poder entre más manos. Si el problema es la desigualdad, asegurémonos de que hay más de una superinteligencia. Esta posibilidad, sin embargo, comportaría todavía más riesgos. Si la IA general mejora a un ritmo exponencial, la empresa que lo alcance antes (aunque sea con sólo unos meses de antelación) mantendrá una ventaja que se incrementará en el tiempo, situándola en una posición de permanente privilegio. Siendo así, que dos o más empresas estuvieran muy cerca de recrear una IA general podría desencadenar una tormenta perfecta. Alimentadas por la posibilidad de alcanzar un éxito de tal magnitud, los contendientes podrían hacer apuestas más y más arriesgadas y flexibilizar progresivamente las medidas de seguridad. Como ya hemos visto, pequeños errores de programación en una superinteligencia podrían comportar daños catastróficos para toda la humanidad.

En definitiva, teniendo todo eso presente, los Estados tendrían que buscar sinergias para convertir la búsqueda de una IA general en un proyecto de cooperación internacional (igual que lo son el CERN y la Estación Espacial Internacional o lo fue el Proyecto Genoma Humano). Quizás conseguiríamos de esta manera evitar que ningún individuo, empresa o país pueda monopolizar el dominio sobre una hipotética IA general. Al mismo tiempo estaríamos un poco más cerca de garantizar que la superinteligencia, si nunca llega existir, beneficia igualitariamente a toda la humanidad (y al resto de seres sensibles) y está al servicio de los principios éticos ampliamente compartidos, tal como propone Nick Bostrom [20].

Es posible que en nuestro tiempo vital tengamos que hacer frente a algunas de las decisiones más importantes que ha tomado nunca la humanidad. Aun así, hoy por hoy actuamos como trapecistas que hacen equilibrios entre la extinción y la eternidad con los ojos vendados. Quizás se está plantando ya la semilla de una mente diferente a todas las que hemos conocido hasta ahora. Una mente que puede comportar el paraíso, el humanicidio, o una nueva manera de ser nosotros mismos, pero por el momento preferimos mirar hacia otro lado. Hace falta un análisis profundo, a todos los niveles, que contemple las múltiples posibilidades y consecuencias, porque incluso un resultado relativamente improbable adquiere una vital importancia si condiciona todo nuestro futuro como especie (o como forma de vida). Sabemos tan poco sobre el universo y sobre la propia humanidad que es imposible que este tipo de decisiones no nos produzcan vértigo. Pero tenemos las riendas del futuro en nuestras manos y, hoy por hoy, somos los únicos responsables de legar a nuestros hijos el mejor futuro de todos los posibles. ¿Si vamos errados y las superinteligencias no llegan a existir nunca, al fin y al cabo, sólo habremos perdido un poco de tiempo, y qué es un poco de tiempo comparado con toda la eternidad?

REFERENCIAS

1 — Paine, C. (2018) Do you trust in this computer? Papercut Films.
2 — Kurzweil, R. (2006) The Singularity is Near: When Humans Transcend Biology. Penguin.
3 — Esta es, por ejemplo, la línea de trabajo de OpenCog.
4 — Searle, J. R. (2007) Biological naturalism. In Max Velmans & Susan Schneider. Blackwell.
5 — Dreyfus, H. (1972) What Computers Can’t Do, New York: MIT Press.
6 — Kurzweil, R. (2013) How to Create a Mind: The Secret of Human Thought Revealed. Penguin.
7 — Fan, X., Markram, H. (2019) A Brief History of Simulation Neuroscience, Frontiers in Neuroinformatics, vol. 13, 7.
8 — Grace, K., Salvatier, J., Dafoe, A., Baobao Z.. (2018) When Will AI Exceed Human Performance? Evidence from AI Experts. Journal of Artificial Intelligence Research 62 729-754.
9 — Müller V.C., Bostrom N. (2016) Future Progress in Artificial Intelligence: A Survey of Expert Opinion. In: Müller V. (eds) Fundamental Issues of Artificial Intelligence.

10 — Los cálculos expuestos por Ray Kurzweil en «The Singularity is Near: When Humanos Transcend Biology» presuponían un incremento anual de transistores por microprocesador que actualmente está siendo bastante inferior al previsto. Esta ralentización de la Ley de Moore, en principio, debería conllevar un retraso en sus predicciones pero, aún así, en el año 2017 el autor mantenía considerablemente sus pronósticos iniciales, previendo el logro de ordenadores de consumo tanto potentes como el cerebro humano en 2029 y afianzándose en la inminencia de la singularidad tecnológica. Se puede consultar la opinión del autor en su intervención con motivo en el SXSW Festival de 2017.

11 — Transhumanist FAQ. Humanity+.

12 — Winkler, R. (2017) Elon Musk Launches Neuralink to Connect Brains With Computers. Wall Street Journal.

13 — Les interrelacions entre transhumanisme i Silicon Valley poden apreciar-se, per exemple, en: Vance. A. (2010) Merely Human? That's So Yesterday.

14 — Sainato, M. (2015) Stephen Hawking, Elon Musk, and Bill Gates Warn About Artificial Intelligence.

15 — Wagner, K. (2017) Mark Zuckerberg thinks AI fearmongering is bad. Elon Musk thinks Zuckerberg doesn't know what he's talking about.

16 — Müller V.C., Bostrom N. (2016) Future Progress in Artificial Intelligence: A Survey of Expert Opinion. In: Müller V. (eds) Fundamental Issues of Artificial Intelligence.

17 — Bostrom, N. (2014) Superintelligence: Paths, Dangers, Strategies. Oxford University Press, Inc.

18 — Yudkowsky, E. (2004) Coherent Extrapolated Volition. The Singularity Institute.

19 — McLay, R. (s. d.) Managing the rise of Artificial Intelligence . Recuperat de https://tech.humanrights.gov.au/sites/default/files/inline-files/100%20-%20Ron%20McLay.pdf (Consultat per darrera vegada el 16 de novembre de 2019).

20 — Bostrom, N. (2014) Superintelligence: Paths, Dangers, Strategies. Oxford University Press, Inc.

Miquel Casas

Miquel Casas es licenciado en Derecho por la Universidad Pompeu Fabra y tiene un máster en Filosofía con especialidad en Lógica, Historia y Filosofía de la Ciencia por la UNED. Ha trabajado como técnico superior en varios departamentos de la Generalidad de Cataluña y en el Instituto Catalán de la Acogida y la Adopción, y actualmente es asesor jurídico en el Instituto Catalán Internacional para la Paz. Ha publicado su trabajo de fin de máster en Ápeiron Ediciones: "El fin del Homo sapiens: la naturaleza y el transhumanismo" (2017).