
VARIA / VARIA

**¿UN MUNDO NUEVO?
REALIDAD VIRTUAL, REALIDAD
AUMENTADA, INTELIGENCIA
ARTIFICIAL, HUMANIDAD
MEJORADA, INTERNET DE LAS
COSAS**

**A NEW WORLD?
VIRTUAL REALITY, AUGMENTED
REALITY, ARTIFICIAL
INTELLIGENCE, ENHANCED
HUMANITY, THE INTERNET OF
THINGS**

Antonio Manuel Liz Gutiérrez

Universidad de La Laguna

ORCID iD: <https://orcid.org/0000-0002-2511-1108>

manuliz@ull.es

Cómo citar este artículo/Citation: Liz Gutiérrez, A. M. (2020). ¿Un mundo nuevo? Realidad virtual, realidad aumentada, inteligencia artificial, humanidad mejorada, Internet de las cosas. *Arbor*, 196 (797): a572. <https://doi.org/10.3989/arbor.2020.797n3009>

Copyright: © 2020 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia de uso y distribución Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0).

Recibido: 6 agosto 2018. Aceptado: 16 febrero 2019.

RESUMEN: Las nuevas tecnologías parecen estar cambiando radicalmente nuestro mundo. Analizamos aquí las condiciones para que esto sea así con relación a un grupo muy importante de fenómenos vinculados a las tecnologías computacionales y a las biotecnologías, que responden a denominaciones de uso ya muy extendido, en buena medida gracias a los medios de comunicación: *realidad virtual, realidad aumentada, inteligencia artificial, humanidad mejorada e Internet de las cosas*. Argumentamos que la inteligencia artificial ocupa entre ellos un lugar central y que la constitución de un mundo realmente nuevo depende esencialmente de nuestras actitudes ante ella.

PALABRAS CLAVE: Realidad virtual; realidad aumentada; inteligencia artificial; humanidad mejorada; transhumanismo; Internet de las cosas.

ABSTRACT: New technologies appear to be radically changing our world. Here we analyse the conditions under which this can take place. We do this in relation to an important group of phenomena related to computational technologies and biotechnologies. These phenomena respond to widely used terms, largely thanks to the media: *virtual reality, augmented reality, artificial intelligence, enhanced humanity and the Internet of Things*. We argue that artificial intelligence occupies a central place among them and that the constitution of a truly new world essentially depends on our attitudes to it.

KEYWORDS: Virtual reality; augmented reality; artificial intelligence; enhanced humanity; transhumanism; the Internet of Things.

“Nosotros no renunciamos a nada.”

Serie *Vikings* (*Vikings*, 2013)

Los fenómenos mencionados en el subtítulo de este trabajo parecen abrir las ventanas a un mundo nuevo, o tal vez directamente las puertas. ¿Es realmente así? Puede serlo. Pero también puede no serlo. Argumentaremos que la inteligencia artificial juega aquí un papel crucial. Y que todo depende de nuestras actitudes ante ella.

El presente trabajo desarrolla estas dos ideas. Que la inteligencia artificial ocupa un lugar central en ese conjunto de fenómenos sobre el que estamos llamando la atención. Y que asumir la inteligencia artificial como algo con la capacidad de abrirnos las ventanas y las puertas a un auténtico nuevo mundo requiere profundos cambios en nuestra propia mente.

Esta última tesis puede ser fácilmente malinterpretada. El cambio de mente que se requiere no consiste en mejoras de nuestras capacidades cognitivas. Ni tampoco consiste en un cambio de creencias. Ni simplemente en aceptar cosas como, por ejemplo, que una máquina pueda ganar a los mejores jugadores mundiales de ajedrez, del juego del go o de póker. Se necesita algo muy diferente a todo esto. Se necesitaría un profundo cambio de actitudes. Solo a través de un cambio radical de mente en este sentido, esos fenómenos podrán realmente configurar un mundo nuevo.

Una de las series televisivas que recientemente he visto con mi hijo, Ulises, es *The Big Bang Theory* (2007). Él se dio cuenta de un detalle que permite entender muy bien lo que estamos queriendo decir. Uno de los personajes principales de esa serie, Sheldon Cooper, físico teórico tan inteligente como carente de habilidades sociales, repite algo que expresa perfectamente el cambio de mente que se necesita. En caso de conflicto grave entre las máquinas y la humanidad, este personaje nos dice que él firmaría sin dudarle a favor de las máquinas. Este sería el cambio de mente desde el que lo que empezamos a tener ante nosotros en verdad puede ser un mundo radicalmente nuevo.

¿Es posible cambiar de mente en este sentido? ¿Es que acaso hay otros tipos de mente? Claro que sí. Tenemos ejemplos extremos en las mentes de esas otras formas de humanidad que hemos encontrado en neandertales, denisovanos, hombres de las flores, etc. Y el desarrollo de potentes capacidades simbólicas y sociales en la mente del *sapiens* moderno pudo ser la diferencia que marcó nuestros distintos destinos.

Cosas similares pueden decirse de las variadas mentes animales y de su evolución, así como del desarrollo de las mentes infantiles. Los cambios de actitudes ante el mundo y ante uno mismo justifican que pueda hablarse de cambios de mente. En nuestra opinión, solo con un cierto cambio de mente, que implique cambios radicales en las actitudes, el mundo que se insinúa ante nosotros podrá ser realmente un nuevo mundo.

Acabamos de mencionar una popular serie de televisión. Podríamos mencionar muchas otras. Y también películas, videojuegos, cómics, etc. Esto tiene su justificación. Esas series, películas, videojuegos, cómics, etc., constituyen una fuente inagotable de evidencias sumamente relevantes respecto a las actitudes que nos interesan¹.

1. UNA IRRESPONSABILIDAD CONCEPTUAL

Los fenómenos mencionados en el subtítulo parecen estar cambiando radical y completamente el mundo en que vivimos. Pero no sabemos si realmente es así. El caso es que apenas hay análisis filosóficos serios sobre la confluencia de todos esos fenómenos. Tampoco existe una terminología filosófica precisa. La situación conceptual es aún previa a la argumentación detallada². Como hemos dicho, lo que sobre todo tenemos son series de televisión, películas, videojuegos, cómics, etc., que nos hablan de diversas posibilidades de mundo. Esto tiene la ventaja de situarnos a todos en igualdad de condiciones frente a muchas de las discusiones. Pero no es suficiente. Y la filosofía no puede esconder la cabeza. No puede ser indiferente a lo que tanta gente está viendo en el horizonte. No puede mostrarse indiferente sin ser conceptualmente irresponsable³.

No se trata de que haya filosofía en las series de televisión, en las películas, los videojuegos o los cómics. Hay contenidos filosóficos, pero propiamente no hay filosofía. La filosofía se *muestra* a través de estos medios de expresión, la filosofía se manifiesta en ellos, pero no se hace. Hacer filosofía es argumentar, dialogar, discutir, etc., con el fin de comprender mejor el mundo y de comprendernos mejor a nosotros mismos.

El fin de la creencia racional es alcanzar la verdad mediante razones, intentando convertirse así en conocimiento. El fin de la filosofía no llega a esto. Mejor dicho, dada la amplitud de su mirada, no puede llegar a tanto. El fin de la filosofía es comprender un poco mejor todas las cosas.

Las series de televisión, las películas, los videojuegos, los cómics, etc. pueden compartir ese fin. Igual que la literatura o la pintura. Pero no intentan alcanzarlo argumentando, dialogando, discutiendo. La filosofía, en cambio, sí.

Cabe entender la filosofía como una variedad muy peculiar de “arte conceptual” que busca la comprensión a través de argumentos. Esta es una perspectiva interesante. Como arte, la filosofía se aparta ligeramente de la ciencia. Se acerca a la literatura y a la pintura. También a las series de televisión, las películas, los videojuegos o los cómics. Sin embargo, sigue existiendo una importante diferencia. El medio en el que la filosofía desarrolla su arte son los argumentos. Y esto incluye todo diálogo y discusión donde el protagonismo lo tengan las razones, aunque sea un diálogo y discusión con uno mismo.

La filosofía no puede ser indiferente a ese nuevo mundo que se muestra en muchas de nuestras series de televisión, películas, videojuegos o cómics. No puede ser indiferente a la curiosa combinación de *realidad virtual*, *realidad aumentada*, *inteligencia artificial*, *humanidad mejorada* e *Internet de las cosas* que tenemos ante nosotros. Que tenemos ya entre nosotros.

2. CINCO FENÓMENOS DE NUEVA TECNOLOGÍA

Los cambios en el mundo sobre los que queremos llamar la atención se centran en cinco fenómenos vinculados a las llamadas nuevas tecnologías. Principalmente se vinculan a las tecnologías computacionales y a las biotecnologías. Para cada uno de ellos existe una expresión ya acuñada y en uso. A continuación haremos una caracterización general.

1. Realidad virtual

Una *realidad virtual* es un entorno de apariencia real o de realidad aparente que, a pesar de provocar una fuerte sensación de inmersión, habitualmente generada mediante tecnologías computacionales, puede no tener la clase de realidad que parece tener.

Dos buenos ejemplos de realidad virtual en contextos de ficción son la serie de películas de la saga *Matrix* (1999) y la película *Desafío total* (1990). Algunos episodios de la serie *Black Mirror* (2011) también analizan este fenómeno. En particular, el tema es tratado en *Blanca Navidad* (temporada 2, episodio 4), donde la realidad virtual es un recurso ampliamente utilizado por el sistema judicial, y en *San Junípero* (temporada 3, episodio 4), donde la realidad virtual hace posible una vida después de la muerte.

Otro ejemplo de realidad virtual, más clásico pero igual de inquietante, un ejemplo que queremos mencionar por no involucrar aún el uso de tecnología computacional, es el relato *La invención de Morel* de Bioy Casares (1940).

Fuera de la ficción, hay actualmente varios tipos de dispositivos creadores de realidad virtual (*realmente* creadores de realidad virtual, podríamos decir) en la modalidad sensorial de la visión. Un entorno de realidad virtual con inmersión máxima llevaría esto al extremo, llegando a todas las modalidades sensoriales⁴.

2. Realidad aumentada

La *realidad aumentada* es un entorno real mezclado con elementos virtuales creando realidades mixtas en tiempo real o cercano a un tiempo real.

De nuevo, algunos episodios de la serie *Black Mirror* (2011) muestran este fenómeno. Concretamente lo hacen *Toda tu historia* (temporada 1, episodio 3) y *La ciencia de matar* (temporada 3, episodio 5). En el primero, toda la realidad se amplía con los contenidos de una memoria perceptiva guardada en implantes cerebrales. En el segundo, unos supuestos enemigos son literalmente vistos con la apariencia de seres repugnantes.

Buenos ejemplos cercanos de realidad aumentada los ofrecen los códigos QR y las aplicaciones de móvil que utilizan datos y recursos de Internet en interacción con los intereses de un usuario en un entorno real. Las aplicaciones de mapas ocupan aquí un lugar destacado. También podemos encontrar realidad aumentada en los dispositivos de visitas guiadas utilizados en la mayoría de los actuales museos. Otros ejemplos son algunos proyectos de *gafas* que van dando información sobre los entornos hacia los que se dirige nuestra mirada.

Otro ejemplo, muy reciente y sin duda interesante, de realidad aumentada lo constituye el juego *Pokemon Go* (2016). El impacto social inicial fue sorprendente. Y es que, de repente, nuestras calles y nuestros propios hogares se llenaron de esos extraños seres llamados *pokemons*. Y cientos de personas a nuestro alrededor orientaban su acción en relación a ellos. Lo fascinante es que, en lugar de estos seres, podría haber muchísimas otras cosas.

La realidad virtual y la realidad aumentada están cambiando lo que puede llegar a contar como *experiencia*. Hasta ahora, cabía distinguir dos tipos básicos de experiencia: 1) la experiencia ordinaria

adquirida a través de nuestros sentidos y expresada en nuestros lenguajes ordinarios, y 2) la experiencia científica obtenida a través de instrumentos, mediciones, experimentos, etc., y expresada en los lenguajes de la ciencia. Con la realidad virtual y la realidad aumentada cabe también hablar de 3) una experiencia producida en el seno de realidades virtuales y de realidades aumentadas. Puede replicarse que se trata de una experiencia simulada. Pero es que la experiencia simulada también es experiencia. Y cada vez está adquiriendo mayor importancia este nuevo tipo de experiencia. Sobre todo en ámbitos educativos, pero no solo.

3. Inteligencia artificial

Debemos distinguir varios sentidos en esta noción. Podemos llamar *inteligencia artificial* a cualquier cosa creada artificialmente que sea capaz de realizar tareas que en nosotros requerirían cierta inteligencia. En un sentido más sustantivo, también podemos llamar inteligencia artificial a un sistema inteligente creado artificialmente. Y en un sentido aún más sustantivo, podemos llamar inteligencia artificial a un sistema artificial con vida mental como la nuestra.

El segundo sentido es muy importante. Y a veces pasa desapercibido entre el extremo de llevar a cabo tareas que de ser hechas por nosotros requerirían inteligencia y el extremo de tener algún tipo de completa vida mental creada de manera artificial. Ser un sistema inteligente implica algo más que realizar tareas que en nosotros requerirían inteligencia. Pero tal vez también requiera algo menos que tener una vida mental exactamente como la nuestra. En cualquier caso, hay una gradación de sentidos. Esto hace que, a pesar de lo que pueda parecer, la segunda caracterización no sea simplemente trivial. Que sea posible la inteligencia artificial implica mayores compromisos en el segundo sentido que en el primero. Y compromisos aún más fuertes en el tercer sentido que en el segundo.

La literatura y el cine nos han mostrado numerosos ejemplos de inteligencias artificiales. Mencionaremos solamente el libro de Asimov, *I Robot* (1950), y las películas de Scott, *Blade Runner* (1982), de Cameron, *Terminator* (1984), de Spielberg, *AI* (2001), de Wally Pfister, *Transcendence* (2014) y de Alex Garland, *Ex Machina* (2015)⁵.

También las series de televisión han explorado con pasión este campo. Destacaremos otra vez *Black Mirror* (2011). En particular, *Ahora mismo vuelvo* (temporada 2, episodio 1), que nos muestra con gran

profundidad y desgarró los conflictos personales generados por un sistema de inteligencia artificial que sustituye a un marido muerto. También merece nuestra atención la serie *Humans* (2015), donde humanos y robots conviven en una misma sociedad manteniendo relaciones llenas de zonas grises. Y la serie, aún más reciente, *West World* (2016), en la que además se ofrece toda una sofisticada teoría sobre la conciencia.

Mención aparte merece la gran obra titulada *Battlestar Galactica* (2004), objeto de culto para los personajes de *The Big Bang Theory* (2007). En ella, se ha hecho realidad el sueño de la inteligencia artificial. Se han creado robots conscientes. Estos luchan por sustituir al ser humano. Pero al final, ambos grupos resultan prácticamente indistinguibles. Es justamente comentando esta segunda serie cuando Sheldon Cooper insiste siempre en su convicción de que, en caso de conflicto grave entre las máquinas y la humanidad, él firmaría sin dudarle a favor de las máquinas.

Pero los proyectos de creación de robots *humanoides* van mucho más allá de la ficción. Desde hace ya unas décadas, la inteligencia artificial y la robótica (algo así como la especialidad más fisiológica y anatómica de la inteligencia artificial) constituyen potentes áreas de investigación dentro de la informática, la ingeniería y, en general, las ciencias de la computación. Y aquí, son numerosos tales proyectos.

Los sistemas de inteligencia artificial de apariencia no-humana pueden adoptar básicamente dos formas: o bien pueden ser *robots industriales*, algo que podemos fácilmente encontrar en cualquier cadena de montaje, o bien pueden ser simplemente *programas*, por ejemplo, para mantener conversaciones con fluidez, o para jugar al ajedrez, al go o al póker. *Siri* o *Cortana* serían ingenios de inteligencias artificiales de este segundo tipo. A pesar de mostrar una competencia lingüística humana, carecen de una apariencia física humana. En contraste con todos ellos, los robots humanoides sí tienen una apariencia física humana⁶.

A veces, los proyectos de robots humanoides están vinculados a grandes corporaciones, como el robot *Asimo* (de la compañía Honda), o *Petman* y *Atlas* (de Boston Dynamics) o *Nao* (de Aldebaran Robotics). Otras veces, se trata de proyectos más académicos, vinculados a destacadas universidades y centros de investigación, como el proyecto *COG*, desarrollado durante varias décadas por el MIT⁷.

4. Humanidad mejorada

La *humanidad mejorada* consiste en la transformación de la condición humana mediante el desarrollo y la aplicación de tecnología, sobre todo tecnologías computacionales y biotecnologías. El objetivo es mejorar las capacidades físicas y psicológicas del ser humano.

Muchos de los superhéroes de nuestros cómics son casos de humanidad mejorada, aunque no todos. A veces, también intervienen factores circunstanciales no tecnológicos o meras casualidades. En el ámbito de los superhéroes de cómic, se habla frecuentemente de *metahumanos* y de *inhumanos*.

También el doctor Frankenstein, en la versión original de Mary Shelley, hace ya dos siglos, pretendía mejorar la condición humana con ayuda de la tecnología. Y lo sigue haciendo hoy día la actual corriente cultural del *transhumanismo*, denominada en ocasiones de forma abreviada con el símbolo H+⁸.

La propuesta del transhumanismo parte de una posición teórica inicialmente razonable: el abandono del prejuicio de que la condición humana sea esencialmente inalterable. Pero el transhumanismo no solo es una corriente teórica. Tiene un activismo práctico tan entusiasta como cuestionable. Rechaza hablar de prudencia y reflexión. Se empeña en comenzar a hacer realidad sus proyectos de transformación de la humanidad con los recursos tecnológicos actualmente existentes.

El problema principal del transhumanismo está en la precipitación. Tal vez las tecnologías que ya existen no sirvan. O lo que es peor, tal vez aun sirviendo para ciertos proyectos de transformación, se necesiten perfilar mucho más estos mismos proyectos. Tal vez sean los proyectos actuales de transformación los que no sirvan. Otro prejuicio tan detestable como el denunciado por el transhumanismo respecto a la *inalterabilidad esencial de la condición humana* es el prejuicio de creer que nuestros proyectos actuales son los mejores, y que en consecuencia deben considerarse esencialmente inalterables.

No hay que olvidar que también son casos de humanidad mejorada la implantación de marcapasos, el consumo de medicinas, las prótesis dentales, el simple hecho de usar unas gafas graduadas, etc. Esto ha existido desde siempre. Recordemos también la búsqueda de la inmortalidad a través de la llamada *alquimia interna* del taoísmo. Pero el transhumanismo cree que hoy día tenemos ya los medios tecnológicos sufi-

cientes para llevar la humanidad a otro nivel. Y puede que esto simplemente no sea verdad. O puede que necesitemos definir mucho mejor ese *otro nivel* que pretendemos alcanzar.

Además, como contraste a todo esto, también estarían otros casos que podemos llamar de humanidad *desmejorada*. Uno de los casos imaginarios más llamativos lo protagonizarían los *zombies*. Y estas posibilidades nos amenazan desde el futuro.

5. Internet de las cosas

En general, se llama *Internet de las cosas* a la red de conexión e intercambio de información entre las cosas de nuestro mundo, principalmente objetos tecnológicos. Pero la expresión se aplica más específicamente cuando ese grado de conexión e intercambio de información (gestión, control, decisión, etc.) llega a ser mayor entre las cosas que entre las personas.

La idea de una Internet de las cosas surge con timidez en el campo de la tecnología de sensores a comienzos del presente siglo. Unos cuantos años después, en el *Mobile World Congress* celebrado en Barcelona en 2017, la expresión *The Internet of Things* (con las siglas IoT) pasa a ser la expresión más usada.

Uno de sus principales promotores, Kevin Ashton, hacía la siguiente declaración en la revista *Radio Frequency Identification*, en un artículo de 2009 titulado "Esa cosa del 'Internet de las cosas'»(los subrayados son nuestros):

Los ordenadores actuales —y, por tanto, Internet— son prácticamente dependientes de los seres humanos para recabar información. Una mayoría de los casi 50 petabytes (un petabyte son 1024 terabytes) de datos disponibles en Internet fueron inicialmente creados por humanos, a base de teclear, presionar un botón, tomar una imagen digital o escanear un código de barras. Los diagramas convencionales de Internet, dejan fuera a *los routers más importantes de todas: las personas*. El problema es que las personas tienen un tiempo, una atención y una precisión limitados, y no se les da muy bien conseguir información sobre cosas en el mundo real. Y eso es un gran *obstáculo*. Somos cuerpos físicos, al igual que el medio que nos rodea. No podemos comer bits, ni quemarlos para resguardarnos del frío, ni meterlos en tanques de gas. *Las ideas y la información son importantes, pero las cosas cotidianas tienen mucho más valor*. Sin embargo, la tecnología de la información actual es tan dependiente de los datos escritos por personas que nuestros ordenadores saben más sobre ideas que

sobre cosas. Si tuviéramos ordenadores que supieran todo lo que tuvieran que saber sobre las “cosas”, mediante el uso de datos que ellos mismos pudieran recoger sin nuestra ayuda, nosotros podríamos monitorizar, contar y localizar todo a nuestro alrededor, de esta manera se reducirían increíblemente gastos, pérdidas y costes. Sabríamos cuándo reemplazar, reparar o recuperar lo que fuera, así como si su funcionamiento era el correcto. La Internet de las cosas tiene el *potencial para cambiar el mundo* tal y como hizo la revolución digital hace unas décadas. Tal vez incluso hasta más.

La Internet de las cosas constituiría una peculiar *tercera naturaleza*. La *primera naturaleza* estaría determinada por la constitución física, química y biológica. La *segunda naturaleza* la constituye nuestra sociedad y cultura, o nuestras diversas sociedades y culturas. Además de esto, la Internet de las cosas abriría las puertas a algo parecido a la sociedad y la cultura, pero en un sentido cada vez más emancipado de los seres humanos. Y esta sería realmente una *tercera naturaleza*.

3. ¿QUÉ SE VISLUMBRA EN EL HORIZONTE?

Podemos acabar aquí nuestra presentación. Añadir detalles no cambiaría el panorama general. ¿Cuál es ese panorama? ¿Qué se vislumbra en el horizonte? Combinando todos los elementos anteriores podemos obtener cosas terribles. Sobre todo, si pensamos en términos económicos y militares. En términos de posibilidades de opresión, el panorama puede ser ciertamente desolador. No deberíamos perder de vista estas posibilidades. Pero tampoco tienen que cegarnos. Pues también podemos ver cosas muy distintas a las que tenemos hoy día en otros sentidos. Cosas como, por ejemplo, las siguientes.

Lo que siempre hemos querido tomar como la realidad podría llegar a ser sustituido, al menos en ciertos casos, por esas otras cosas que ahora llamamos *realidades virtuales*. Así, en lugar de ir de vacaciones a Tailandia, podríamos vivir una experiencia de vacaciones en Tailandia. En lugar de seguir una enseñanza convencional, podríamos elegir una enseñanza virtual, con sus profesores virtuales, sus compañeros virtuales, sus recreos virtuales. En todo esto la realidad virtual no sería simplemente algo no-real o irreal. Sería una realidad alternativa.

También la realidad se ampliaría. Las realidades, digamos, *reales*, podrían llegar a mezclarse con las realidades virtuales de formas inextricables. La realidad incluso podría llegar a *auto-explicarse* como lo hacen

los cuadros de un museo cuando apretamos el botón apropiado en el dispositivo de visita guiada. Tal vez, casi toda la realidad podría acabar convertida en una realidad ampliada de este tipo.

Todo esto se aplicaría al propio ser humano, llegando a hacer indistinguibles a las personas virtuales de las personas que acostumbramos a considerar reales. Podrían incluso existir robots humanoides con plenitud de derechos y obligaciones. Y no describiríamos ya este hecho diciendo simplemente que los tratamos *como* personas. No haríamos ninguna distinción. Serían personas.

Nosotros mismos también podríamos cambiar mucho. Cambiaría nuestro cuerpo y cambiaría nuestra mente. Es más, las tecnologías computacionales y las biotecnologías acaso puedan llevarnos al borde mismo de la inmortalidad. No a una inmortalidad postrada, sino plenamente activa y llena de vida. Nuestras capacidades podrían llegar al máximo de lo que un cuerpo puede hacer.

Y todos nuestros productos tecnológicos, mezclados con la propia naturaleza, todo nuestro mundo artificial, o artificializado, podría estar interconectado en una inmensa red de información, comunicación y control. Y esa red podría llegar a extenderse, poco a poco, por todo el universo.

Estas son algunas de las *visiones de un nuevo mundo* sobre las que queremos reflexionar. Y como hemos dicho, lo haremos dejando por un momento al margen las posibilidades de opresión. Ciertamente, sentimos próximas todas estas posibilidades. No podemos confiarnos. Pero no es menos verdad que siempre nos han acechado.

Algunos autores han defendido la necesidad de imaginar una nueva política y una nueva economía⁹. Otros autores también han defendido la necesidad de elaborar una nueva ética. Y un aspecto muy importante de esta nueva ética sería la incorporación de *valores humanos* en las máquinas¹⁰. No hay duda de que dicha transferencia de valores puede ser crucial para nuestra supervivencia. Y no hay oposición entre esta idea y lo que queremos defender en este trabajo. Más bien deberíamos ver la cuestión como un camino con dos direcciones. Tendría que existir no solo transferencia de valores humanos a las máquinas, sino también reconocimiento en ellas de posibles identidades personales en construcción. Identidades personales que acaso solo puedan llegar a existir a través del reconocimiento de que existen. Y esto no sería muy distinto a lo que de hecho pasa entre los propios humanos.

Para explorar el panorama que acabamos de dibujar, se requiere una cierta perspectiva especial. Sin embargo, más que necesitar una perspectiva elevada, más que auparnos a hombros de algunos grandes hombres, lo que necesitamos en este caso es ver el futuro *con los ojos de un niño*.

El que los cinco fenómenos que hemos presentado se articulen de una manera fuertemente unitaria y orgánica, que algunos de ellos estén en el corazón de otros y que la manera de interpretar la inteligencia artificial tenga efectos cruciales sobre el futuro del conjunto de todos ellos son ideas originales de mi hijo Ulises. Cuando una mirada está acostumbrada a ver las cosas de cierto modo, es muy difícil no verlas así. Y un mundo nuevo necesita una mirada nueva¹¹.

4. EL PAPEL CRUCIAL DE LA INTELIGENCIA ARTIFICIAL

Los cinco caracteres que acabamos de perfilar (realidad virtual, realidad ampliada, inteligencia artificial, humanidad mejorada, *Internet* de las cosas) muestran los campos de desarrollo más acelerado y espectacular de las nuevas tecnologías computacionales y de las biotecnologías. En los campos de la inteligencia artificial y de la humanidad mejorada, y muy especialmente en el segundo de ellos, las contribuciones de las biotecnologías tienen mayor relevancia¹². A su vez, en los campos de la realidad virtual, la realidad aumentada y la *Internet* de las cosas, la mayor relevancia la tienen las tecnologías computacionales.

Todo esto parece tremendamente nuevo. Sin embargo, vamos a argumentar que podría no serlo tanto. Y que la diferencia última la acaba poniendo cierta manera de entender la inteligencia artificial. Sin entenderla de cierta forma, podríamos no estar teniendo nada nuevo. Podríamos simplemente estar teniendo *más de lo mismo*. Eso sí, con dosis cada vez mayores de propaganda ideológica y de barroquismo técnico.

Pensemos en la realidad virtual. En sí misma no es un fenómeno en absoluto nuevo. Nos acompaña desde el mismo comienzo de la humanidad. Un importantísimo rasgo diferenciador del *sapiens* moderno frente al neandertal, rasgo repetidamente señalado, es la gran capacidad del primero para crear símbolos. Los contenidos mentales ligados a los símbolos permiten ver el mundo, y vernos a nosotros mismos, de formas nuevas que son fácilmente repetibles, comunicables y modificables. Ya tuvieran o no lenguaje los neandertales, no parecían tener las potentes y flexibles capacidades de simbolización del *sapiens* posterior,

capacidades reflejadas de modo muy directo en sus producciones artísticas y técnicas.

Y desde esos inicios, nuestra realidad fue ya siempre una realidad aumentada. Toda la realidad que nos circunda pasó a ser realidad aumentada principalmente con la aparición del lenguaje. El lenguaje constituyó desde sus lejanos orígenes una poderosa fuente de realidad virtual. Y se mezcló de mil maneras con todos nuestros entornos reales creando realidades aumentadas. El omnipresente contenido conceptual de toda experiencia característicamente humana y la pesada carga teórica de toda observación son los últimos peldaños de esa escalera que comienza con un pie en la realidad y con otro pie en la fabulación simbólica.

La tecnología ha potenciado enormemente la posibilidad de crear realidad virtual. Y nuestra tecnología actual lo ha hecho de una forma que no admite comparación. Casi parece magia. Y ciertamente no hay mucha distinción entre el *Golem* que adquiere vida cuando se le escribe en la frente cierta palabra y nuestros robots que adquieren vida mental al ejecutar ciertos programas informáticos.

La inteligencia artificial es un caso especialmente relevante de realidad virtual. Lo que está en cuestión es la realidad de una mente. Y si la inteligencia artificial se interpretara solo en cierto sentido débil, como ofreciendo meras simulaciones de vida mental, sería ciertamente magia que un robot adquiriera vida mental al ejecutar ciertos programas. Sin embargo, con otras interpretaciones más fuertes no sería ya magia. Sería simplemente ciencia y tecnología¹³.

Consideremos de nuevo la caracterización que hicimos de la realidad virtual. Tuvimos mucho cuidado al afirmar que la realidad aparente generada "*puede no tener* la clase de realidad que parece tener". Esto es muy distinto a decir que esa realidad aparente "*no tiene* la clase de realidad que parece tener". Y aún más distinto sería decir que "*no puede tener* la clase de realidad que parece tener".

El inmenso poder mágico de la realidad virtual se origina justamente en el hecho de que solo cabe afirmar lo primero. La realidad virtual podría no tener la clase de realidad que parece tener. La realidad virtual podría no ser, digamos, una realidad real. ¡Pero también podría serlo! Podría ser una realidad real. Incluso podría ocurrir que casi toda la realidad que consideramos más *real*, casi toda la realidad real más paradigmática no fuera sino una gran y compleja realidad virtual¹⁴.

Todo esto puede ser directamente aplicado a la inteligencia artificial. En algunos casos, la inteligencia artificial nos ofrece ya algo parecido a mentes virtuales. Cuando menos, nos promete mentes virtuales. ¿Podrán ser también mentes reales? Pues sí, tal vez podrían serlo. Y nuestra tesis consiste justamente en afirmar que ello puede depender crucialmente de un cambio en nuestras actitudes, de un cambio en nuestra propia mente.

Hace unos años John Searle adquirió notoriedad por la contundencia de su rechazo a una interpretación fuerte de la inteligencia artificial. Tal rechazo se basaba en un argumento de estructura muy simple, el llamado *argumento de la habitación china*¹⁵. Imaginemos que estamos encerrados en una habitación manipulando símbolos de acuerdo a un manual de instrucciones. Nos envían ciertos conjuntos de símbolos, los cambiamos siguiendo el manual y devolvemos el resultado. El proceso se repite. Sin saberlo nosotros, resulta que estamos manteniendo conversaciones en chino. Pero la mera manipulación de símbolos siguiendo reglas no parece ser capaz de producir ninguna comprensión del chino. Según Searle, la inteligencia artificial solo manipula símbolos siguiendo mecánicamente determinadas reglas. Y esto no puede producir ninguna vida mental. Los estados mentales han de ser efectos causales de nuestros cerebros. Únicamente un sistema capaz de tener estos poderes causales podrá generar estados mentales genuinos (como, por ejemplo, los que constituyen nuestra comprensión del chino). Lo demás son tan solo simulaciones. Meros simulacros de vida mental.

¿Cuál es el punto crucial de este argumento? El punto crucial es tan obvio que pasa fácilmente desapercibido. Searle apela a la *primera persona*. A lo que cada cual, es decir, cada uno de nosotros, diría. En último término, se trata únicamente de que no nos parece que la mera manipulación de símbolos siguiendo reglas pueda producir un estado mental de comprensión del chino. Pero no hay ninguna prueba de que no pueda hacerlo. Simplemente, situándonos en una perspectiva de primera persona, ¡no nos parece que pueda ser así!

¿Cabe estrictamente una *prueba* de que no pueda producirse tal estado de comprensión a través de la manipulación de símbolos? Si mejoramos las condiciones impuestas en el experimento, si dejamos atrás la caricatura implícita en la presentación (ciertamente tendenciosa) de Searle, ¿podemos decir de modo concluyente que *nunca podrá producirse* así un estado mental de comprensión del chino? Es difícil responder esta pregunta.

Pero no es tan difícil ver que el argumento de Searle se desmontaría rápidamente si cambiaran las apreciaciones en primera persona. ¿Cómo encajar una reacción en primera persona que consista en que efectivamente la manipulación de símbolos siguiendo reglas *sí que puede* producir estados mentales de comprensión del chino? ¿Cómo interpretar una reacción que consista en decir que justamente en eso consisten los poderes causales de nuestro cerebro para generar estados mentales? Que consisten en poder manipular símbolos, o algo parecido a símbolos¹⁶, de esa forma peculiar.

¿Cómo responder, pues, al argumento de Searle? ¿Cómo unas mentes virtuales pueden llegar a ser consideradas mentes reales? Ya tenemos una buena respuesta: ¡cambiando nuestra propia mente! ¡Cambiando nuestras actitudes!

Fijemos ahora nuestra atención en la realidad aumentada, la humanidad mejorada y la Internet de las cosas. Las relaciones entre la realidad virtual y la realidad aumentada son muy estrechas. Según nuestras caracterizaciones del principio, toda realidad aumentada incorpora algún tipo de realidad virtual. Y aquí volvemos a encontrarnos de lleno con la inteligencia artificial. Pues cada vez más, la inteligencia artificial se está convirtiendo en el principal motor de todas las tecnologías vinculadas a la realidad virtual.

Y también, como dijimos, la humanidad mejorada puede tomarse como un caso especialmente relevante de realidad aumentada. Lo que se haría al mejorar las capacidades físicas y psicológicas del ser humano es *aumentar* nuestra propia condición. Lo que se aumenta es la propia condición humana. Y muchos tipos de mejora de la condición humana pasan por la incorporación de dispositivos de inteligencia artificial. No siempre ha sido así. Pero la integración de la inteligencia artificial en las tecnologías biomédicas es un hecho¹⁷. Y teniendo en cuenta este hecho, podemos decir que la inteligencia artificial también se ha convertido en el principal motor de la humanidad mejorada. Con esto, el ser humano se transforma literalmente en un *cyborg*. En el caso extremo, debemos considerar incluso la posibilidad de que la propia humanidad, en algún futuro, acabe independizándose completamente de todas las condiciones biológicas heredadas.

¿Qué hay de la Internet de las cosas? La noción de Internet de las cosas encierra una importante ambigüedad. Siempre ha habido más conexiones entre cosas que entre personas. Las cosas están siempre

masivamente conectadas, a través de vínculos físicos y de relaciones informacionales. ¿Qué es entonces lo relevante? Lo que resulta relevante es la clase de realidad aumentada que se obtiene. Lo que ahora resulta relevante es que, a través de las nuevas conexiones entre cosas, principalmente objetos tecnológicos, se despersonalicen, más bien se deshumanicen, procesos de decisión y gestión considerados importantes. Y esto nos lleva de nuevo al campo de la inteligencia artificial. La Internet de las cosas es un tipo de realidad aumentada que nos hace cada vez más prescindibles.

Hay, pues, un importante sentido en el que todo confluye en la inteligencia artificial. Y según se interprete esta, se abren dos grandes caminos: 1) o bien tendremos en verdad algo radicalmente nuevo, 2) o bien solo tendremos más de lo mismo con un grado muy alto de sofisticación y barroquismo técnico.

Si no aceptamos la posibilidad de una inteligencia artificial interpretada en sentido fuerte, de una inteligencia artificial capaz de producir nuevas mentes, y también nuevos tipos de mente, el supuesto nuevo mundo que nos parece ver en el horizonte no será más que un espejismo. Acabaremos teniendo tan solo el *viejo mundo*. Eso sí, técnicamente mucho más sofisticado y barroco.

La inteligencia artificial es una provocación que trastoca completamente nuestras maneras habituales de pensar en la realidad y en nosotros mismos. A primera vista, no reconocemos nuestra subjetividad en el espejo de la inteligencia artificial. Nos parece que nuestra naturaleza debe ser muy diferente a lo que encontramos *reflejado* en este campo. Nos parece que, a lo sumo, la inteligencia artificial solo podrá ofrecernos simulaciones de vida mental, meros simulacros de nuestra auténtica naturaleza. Y sin embargo, podría no ser así. El empeño por marcar diferencias tajantes entre los productos de nuestra mente y todo lo que sea capaz de producir una inteligencia artificial puede ser un mero prejuicio chauvinista. Al igual que puede serlo, más en general, el empeño por separar la realidad virtual de todo lo que nos parece que constituye la realidad real.

5. DESASOSIEGO CON LA INTELIGENCIA ARTIFICIAL

Realmente, la inteligencia artificial está traspasando límites. Y lo está haciendo muy rápido. El problema actual con la inteligencia artificial no es el de hace unas décadas. Ya no es el de si puede pensar una máquina¹⁸. Ahora los problemas cruciales son: ¿cuándo ocurrirá? y ¿qué consecuencias tendrá?¹⁹

Incluso existen estimaciones de la fecha aproximada en la que puede surgir la primera inteligencia artificial plenamente merecedora de tal nombre. La fecha media se sitúa en 2040. A esta aparición suele llamarse *singularidad*. En física, una singularidad es un fenómeno donde no se cumplen las leyes físicas habituales (los agujeros negros, el *big-bang*, etc.). En tecnología, el surgimiento de una inteligencia artificial genuina, con vida mental propia, con intencionalidad y conciencia, o acaso con súper-intencionalidad y súper-conciencia, sería la gran singularidad.

La inteligencia artificial a la vez nos fascina y nos inquieta. ¿Estamos jugando a ser aprendices de brujo? ¿Está nuestra tecnología-magia invocando al diablo? Muchas de las películas y series de ciencia ficción más populares se basan en la idea de una inteligencia artificial que evoluciona hasta un punto en el que la humanidad ya no es capaz de mantener el control sobre sus creaciones, lo cual acaba poniendo en peligro nuestra propia civilización y existencia. Véanse, por ejemplo, las películas *Blade Runner* (1982), *2001: A Space Odyssey* (1968), *The Terminator* (1984), *The Matrix* (1999), *Transcendence* (2014), *Ex Machina* (2015). Y también series como *Battlestar Galactica* (2004) o *West World* (2016). Dadas las capacidades de las máquinas y dispositivos actualmente empleados con propósitos militares, no carece de base este miedo. El crecimiento acelerado de la tecnología asociada a la inteligencia artificial y nuestra creciente dependencia de ella podrían ocasionar realmente nuestra aniquilación.

En este sentido, han sido muy divulgadas las declaraciones críticas de Stephen Hawking respecto a la inteligencia artificial. En su opinión el desarrollo de la inteligencia artificial podría significar el fin de la humanidad. Una de sus afirmaciones más rotundas es la siguiente:

De manera resumida, el éxito en crear inteligencia artificial podrá constituir el evento más grande en la historia de nuestra civilización. Aunque podría ser el último, a menos que aprendamos a sortear los riesgos.

Estas frases formaban parte de un discurso suyo pronunciado a propósito de la apertura, el 19 de octubre de 2016, del *Center for the Future of Intelligence*, creado en la Universidad de Cambridge. En el centro participan también las universidades de Oxford, Berkeley y el Imperial College of London. Cuenta con una dotación económica cuantiosa y su director académico es el filósofo Huw Price (*Bertrand Russell Professor*

en Cambridge). El propósito explícito del centro es explorar las implicaciones de la inteligencia artificial para la civilización humana.

En ese mismo acto de inauguración, Huw Price insistía en la misma idea (las traducciones son siempre nuestras):

La creación de máquinas inteligentes probablemente sea un evento único en la vida de un planeta. Y se trata de un futuro que nosotros, los seres humanos, debemos afrontar juntos. Nuestro objetivo es constituir una amplia comunidad con la experiencia y el sentido de un propósito común suficientes para hacer que ese futuro sea el mejor futuro posible.

El mismo diagnóstico se repite cada vez con mayor frecuencia. Y es significativo que esta sensibilidad ante el *nuevo mundo* que se insinúa sea compartida por la mayoría de los grandes empresarios tecnológicos de la actualidad.

Elon Musk, por ejemplo, fundador visionario de *PayPal*, también de compañías como *Tesla*, empeñada en la promoción de coches eléctricos, y *Space X*, pionera de los viajes espaciales con fines comerciales, en un encuentro con estudiantes del MIT en el *AeroAstro Centennial Symposium*, se refería en 2014 a las perspectivas abiertas por la inteligencia artificial como “nuestra mayor amenaza existencial”. Su propuesta era muy clara y directa:

Cada vez estoy más inclinado a pensar que debería haber alguna instancia supervisora con capacidad reguladora, quizás tanto a nivel nacional como internacional, simplemente para que nos aseguremos de no hacer estupideces.

Más recientemente, en el Festival Tecnológico *South by Southwest* de 2018, vuelve a expresarse en unos términos parecidos:

El peligro de la IA es mucho mayor que el peligro de las armas nucleares. Y nadie sugeriría que podemos permitir que alguien simplemente construya armas nucleares si lo desea. Eso sería una locura.

[...]

Es necesario que haya un organismo público que tenga información y capacidad de supervisión para confirmar que todos están desarrollando la IA de forma segura, esto es extremadamente importante.

[...]

La tasa de desarrollo de la IA es realmente exponencial. Pero tenemos que encontrar la manera de

garantizar que el advenimiento de la súper inteligencia digital sea simbiótico con la humanidad. Creo que esa es la crisis existencial más grande que enfrentamos, y una de las más urgentes.

Una preocupación similar se transmite en las siguientes palabras del cofundador de Microsoft, Bill Gates:

Estoy entre aquellos a los que les preocupa seriamente la súper-inteligencia. Primero, las máquinas harán montones de trabajos para nosotros sin ser súper-inteligentes. Esto debería ser positivo si lo gestionamos bien. Unas pocas décadas después, sin embargo, la inteligencia será suficientemente fuerte como para convertirse en motivo de preocupación. Estoy de acuerdo con Elon Musk y con otros en esto, y no entiendo por qué a alguna gente no le importa.

La inteligencia artificial parece estar transgrediendo límites. Parece estar abriendo fronteras que parecían infranqueables. Pero ¿lo está haciendo realmente? De hecho, existen escenarios en los que la inteligencia artificial adopta otras formas diferentes, otras interpretaciones mucho más débiles. Basta pensar en futuros como los que nos presenta la serie *Star Trek* (1966) o incluso la saga *Star Wars* (1977). En ninguno de estos dos casos hay propiamente algo como una inteligencia artificial en sentido fuerte, capaz de generar mentes conscientes y puntos de vista personales parecidos a los nuestros. Ni siquiera los robots de *Star Wars* llegan convincentemente a este nivel.

Se puede ser sumamente crítico respecto a las ambiciones de la inteligencia artificial en sentido fuerte y, al mismo tiempo, asumir sin ningún reparo toda la inteligencia artificial que aparece en estos ejemplos. Como acabamos de decir, no hay aquí nada que pueda identificarse con mentes conscientes creadas artificialmente. Como mucho, hay simulaciones de vida mental. El propio John Searle podría sentirse totalmente satisfecho y cómodo manteniendo conversaciones fluidas con los simpáticos robots R2-D2 y C-3PO de *Star Wars*, incluso con el más reciente y no menos simpático BB-8.

De cualquier modo, el que la inteligencia artificial esté *realmente* transgrediendo límites y abriendo nuevas fronteras, o más bien tan solo *parezca* que lo está haciendo, es parte esencial del problema²⁰.

6. NUEVO MUNDO Y NUEVA MENTE

Hemos visto cómo la inteligencia artificial, interpretada de cierto modo, sí tiene la capacidad de abrirnos las ventanas y las puertas a un nuevo mundo. En ese

mundo ya no habría simplemente más de lo mismo, con mucha mayor complejidad y sofisticación. Existiría una tremenda novedad. ¡Habríamos construido mentes! ¡Y estas mentes podrían seguir construyendo otras! Definitivamente, ya nada sería como antes.

Vamos a defender ahora una idea en apariencia muy extraña. Ya la hemos insinuado en apartados anteriores. Se trata de la idea de que aceptar la existencia de mentes artificiales requiere profundos cambios en nuestra propia mente.

No acabamos de saber si una máquina construida por nosotros es capaz o no de tener mente. Pero lo peor de todo es que ni siquiera sabemos cómo podríamos llegar a saberlo. No tenemos nada claro el tipo de criterio que deberíamos utilizar.

Esta situación es típica en fenómenos que dependen de una perspectiva o punto de vista. Que algo dependa de una perspectiva no lo hace irreal ni menos importante. Muchas veces ocurre justo al revés. La mayoría de las cosas que nos importan, en general la mayoría de las cosas que consideramos más reales, dependen fuertemente de una perspectiva. Simplemente tener *aspectos* es depender de una perspectiva. Y lo habitual es que las cosas tengan aspectos. Y que solo nos importen bajo algunos aspectos²¹.

Las nociones de perspectiva y de punto de vista están muy conectadas (igual que otras nociones de la misma familia). Podemos decir que tener un punto de vista consiste en la ejemplificación, o instanciación, de una perspectiva por un sujeto. En cualquier caso, el rasgo crucial de las perspectivas y de los puntos de vista es su enorme capacidad para cambiar lo objetivo y lo subjetivo. Son potentes generadores de nuevas objetividades y de nuevas subjetividades.

Debemos explicar mejor todo esto. ¿En qué sentido las perspectivas, y los puntos de vista, son potentes generadores de nuevas objetividades? El sentido relevante tiene que ver con una tensión. Por un lado, los contenidos que nos ofrecen nuestros puntos de vista son los materiales con los que construimos nuestra identidad subjetiva. Y muchos contenidos provocan cambios importantes, incluso pueden llegar a producir auténticas conversiones. Por otro lado, esos contenidos casi nunca son estados meramente subjetivos. No son simplemente estados internos de un sujeto. Esto hace que los consideremos parte de la realidad objetiva. Y como muchas veces esos contenidos tampoco encajan bien con lo que tomamos como la realidad objetiva, por ejemplo la realidad física, los añadimos a ella. Al reflexionar sobre la adopción de puntos

de vista siempre vamos de uno de estos extremos al otro. Y podemos describir esta dinámica diciendo que los puntos de vista son fuente de nuevas subjetividades y de nuevas objetividades.

Necesitamos algún ejemplo. Vemos un mundo coloreado. Pero los colores del cielo, de las hojas de los árboles, el color de una fruta madura, etc., no se incluyen con facilidad entre las realidades objetivas con las que nuestras ciencias cuentan. Un mundo coloreado constituye un nuevo tipo de objetividad. Y el verlo coloreado constituye paralelamente un nuevo tipo de subjetividad. Muchos mamíferos, por ejemplo, no son capaces de ver los colores que nosotros vemos. Nuestras subjetividades son diferentes.

Otro ejemplo. En mi bolsillo tengo varios trozos de papel. Desde cierto punto de vista, es dinero. Son dos billetes de veinte euros. Tengo cuarenta euros en el bolsillo. Aunque esto dependa de una serie de intenciones y convenciones, aunque dependa de ciertos puntos de vista compartidos, puedo decir que objetivamente tengo cuarenta euros en el bolsillo. Y también puedo decir que soy un sujeto sensible a las propiedades y relaciones económicas. Seguramente mis primitivos ancestros no lo fueron. Yo sí tengo este tipo peculiar de subjetividad.

El que las perspectivas, y los puntos de vista, sean generadores de nuevas objetividades y subjetividades ha tenido consecuencias decisivas en nuestra historia natural y cultural. Nuestra identidad se refleja en el mundo que tenemos alrededor. Y también encuentra su reflejo en ese mundo. A todo cambio de objetividad le siguen cambios de subjetividad. Las nuevas objetividades generadas por las perspectivas, o por los puntos de vista, ocasionan nuevas subjetividades. Y viceversa.

A diferencia del mundo del neandertal, el mundo del *sapiens* moderno se llenó de símbolos. Todo comenzó proyectando en algunos objetos o dibujos nuestro mundo interior. Y el mundo exterior aumentó. Se convirtió realmente en un mundo diferente. Fue un mundo nuevo. Y ese mundo nuevo produjo una nueva subjetividad: la nuestra.

Las perspectivas, los puntos de vista, cambian lo subjetivo y lo objetivo. No son simplemente identificables con estados de nuestra subjetividad. Tampoco son simplemente identificables con lo que nuestra ciencia considera más objetivo. Pero cambian las dos cosas. Y esta dinámica puede involucrar tanto los diversos puntos de vista que tenga un sujeto como los variados puntos de vista que pueden tener diversos

sujetos. Las perspectivas, los puntos de vista, siempre son generadores dinámicos de nuevas objetividades y de nuevas subjetividades.

Nuestra tesis es que solo algo así puede hacer que nuestro mundo actual llegue a ser propiamente un mundo nuevo. Que el eje crucial de este cambio se encuentra en la inteligencia artificial. Y que solo cambiando nosotros mismos, cambiando nuestras actitudes, podremos ver ese nuevo mundo ante nosotros²².

7. ENTENDER LO QUE ES UNA MENTE Y ENTENDER LO QUE ES UNA ESCALERA

¿Qué es una mente? La respuesta más simple es que una mente consiste en dolor y en lógica. Una respuesta más elaborada hablaría de estados mentales con un contenido no conceptual o cualitativo (sensaciones visuales, auditivas, táctiles, sensaciones de dolor, emociones, etc.), de estados mentales con un contenido conceptual o proposicional (creencias, deseos, recuerdos, decisiones, etc.) y de cómo todo ello, en un contexto social adecuado, puede llegar a constituir una persona. A menos de que se trate de creencias muy básicas, cambiar solo de creencias no basta para cambiar una mente. Sobre todo -o además- se necesitan cambios en cómo vemos y sentimos la realidad, y en cómo nos vemos y nos sentimos a nosotros mismos.

Volvamos a leer la cita que aparecía al comienzo de este trabajo: “Nosotros no renunciamos a nada”. Pertenece a la serie de televisión *Vikings* (*Vikings*, 2013), coproducida por Canadá e Irlanda. El protagonista principal, Ragnar Lothbrok, describe con esta frase una cualidad que le parece esencial de la naturaleza de su pueblo. Tal cualidad justifica el robo y la traición, desafiar toda ley social y divina. Pero ser fiel a ese impulso nos hace plenamente humanos. O así lo piensa Ragnar Lothbrok. Muchas veces, mi hijo Ulises y yo hemos repetido esa frase. Y hemos hablado de su significado. Curiosamente también tiene una aplicación profunda en relación con los fenómenos que estamos discutiendo aquí.

¿Qué significa “no renunciar a nada” cuando estamos hablando de realidad virtual, de realidad aumentada, de inteligencia artificial, de humanidad mejorada, de la Internet de las cosas? Una respuesta puede ser la siguiente. Significa no renunciar a descubrir un mundo completamente nuevo. Significa no renunciar a vivir en él. Significa no dar ningún paso atrás. Y si al final, para ello, es necesario cambiar de mente, significa hacerlo. Con alegría. Sin ningún miedo.

Comparemos lo que puede ser entender la mente con el caso de entender lo que es una escalera de mano, una escalera portátil. Imaginemos una descripción como la siguiente:

Descripción de una escalera de mano: una escalera de mano es una estructura transportable constituida por dos largueros generalmente paralelos unidos a intervalos iguales por travesaños, que son llamados peldaños o escalones. La estructura permite subir y bajar hasta alcanzar cierta altura deseada. Hay escaleras de mano de dos tipos básicos: apoyables y autoestables. Las primeras deben apoyarse en donde se quiere subir o bajar. Las segundas, también llamadas de tijera, se sustentan por sí mismas gracias a otros largueros que se apoyan en el suelo. Originariamente, las escaleras de mano se fabricaban de madera, pero actualmente el material más utilizado es el aluminio por su ligereza, consistencia y duración.

Cualquiera capaz de entender esta descripción (extraída de un diccionario) debería entender también lo que son las escaleras de mano. Sobre todo, si ya ha tenido contacto previo con ellas. Comparemos este caso con la descripción de un cierto tipo de estado mental en los siguientes términos:

Descripción de un estado mental: un estado mental de tipo M es un estado neurológico de tipo N capaz de desempeñar un papel computacional de tipo C en un sujeto de tipo S que mantiene con el entorno relaciones de tipo R.

Muchas veces se ha contrastado la naturaleza de la mente con la naturaleza de cosas como las escaleras de mano alegando que, mientras que la descripción de una escalera de mano dice algo que permite entender completamente lo que son las escaleras de mano, nada de lo que podría decir la descripción de un estado mental consigue ese efecto. Llenemos como llenemos de contenido las variables presentes en la descripción del estado mental (esto es, en nuestro caso, las variables M, N, C, S y R), nunca llegaremos a entender por qué los estados neurológicos N, desempeñando tales papeles funcionales C en los sujetos S que mantienen con el entorno relaciones R constituyen el estado mental M²³.

Llenemos como llenemos de contenido esas variables de la descripción del estado mental, siempre “parece posible” que algo satisfaga la descripción sin que surja el estado mental M, así como que exista ese estado mental M sin que se satisfaga la descripción. Y esto “no parece posible” con la descripción de una escalera de mano. Satisfacer la descripción de una escalera de

mano es directamente ser una escalera de mano, y viceversa. En ocasiones se habla de un vacío explicativo (de un *explanatory gap*) en las descripciones de los estados mentales, vacío que no existiría en las descripciones de cosas como ser una escalera de mano.

Esto guarda una relación muy estrecha con el planteamiento de John Searle a propósito del experimento mental de la habitación china. También allí *parece* que comprender chino no puede consistir en nada que se haga manipulando símbolos, o algo parecido a símbolos, siguiendo unas reglas. Así mismo, guarda una relación no menos estrecha con un problema planteado hace unos años por Frank Jackson. Se trata de otro experimento mental: Mary la neuróloga. Mary ha vivido siempre encerrada en un entorno artificial que le ha privado de toda sensación de color. Y ha estudiado neurología. Es una experta neuróloga. También conoce todo lo que puede conocerse en física sobre la luz, longitudes de onda, procesos de reflexión, etc. Un día, Mary escapa de su encierro. Sale a un jardín y de repente se encuentra con un tomate maduro de un intenso color rojo. Se sorprende. No puede dejar de mirarlo. ¿Conoce ahora Mary algo que antes no sabía, algo que tal vez ni siquiera podía saber? ¿Hay hechos que solo pueden llegar a conocerse experimentándolos en primera persona?²⁴.

Retomemos la comparación que estábamos analizando. ¿Son tan diferentes las descripciones de cosas como ser una escalera de mano y las descripciones de cosas como los estados mentales? Nosotros también responderíamos que efectivamente lo son. Son descripciones muy diferentes. Pero a nuestra respuesta le acompaña un diagnóstico muy distinto del que acabamos de recoger.

Hay un engaño crucial en la forma de presentar ambas descripciones. En la descripción de las escaleras de mano, presuponemos que entendemos bastante bien todos los conceptos que aparecen en ella. Presuponemos, por ejemplo, que entendemos bien lo que es *transportar*, lo que es *subir y bajar*, lo que es *alcanzar cierta altura deseada*, etc. Y esto implica una multitud de usos, de historias complejas, de perspectivas asumidas. En la descripción ofrecida de un cierto tipo de estado mental, simplemente hay variables de varias clases. La descripción del estado mental no es en realidad una descripción. Es tan solo un marco, un esquema muy formal y genérico.

Si no contáramos con los usos, historias, perspectivas, etc., que nos permiten entender la descripción de las escaleras de mano, seguramente tampoco podríamos

entender a través suyo lo que son las escaleras de mano. Y esto implica que llenando de contenido adecuado las variables de las descripciones de los estados mentales, dándoles un contenido suficientemente ligado a unos usos, historias y perspectivas relevantes, tal vez la descripción resultante, o algunas de las descripciones resultantes, sí podrían permitir entender lo que son los estados mentales en cuestión.

Dicho de otro modo, la descripción de las escaleras de mano nos permite entender lo que son las escaleras de mano porque tenemos *ya mucha familiaridad* con las escaleras de mano. Si nunca hubiéramos conocido y reconocido repetidas veces escaleras de mano, conocer los hechos descritos en la descripción de una escalera de mano difícilmente conduciría a saber lo que son las escaleras de mano. Paralelamente, la descripción del estado mental no parece permitirnos entender lo que es ese estado mental porque no podemos apoyarnos en ninguna *familiaridad* semejante.

Dicho aun de otro modo, en la descripción del estado mental parece haber efectivamente un vacío explicativo. ¡Pero lo hemos puesto nosotros! Y lo mismo cabría decir del experimento mental de la habitación china, planteado por John Searle. Y también del experimento mental de Mary la neuróloga, planteado por Frank Jackson. Los vacíos explicativos los introducimos nosotros al plantear el problema en esos términos.

Así pues, sí hay una diferencia crucial entre la descripción de las escaleras de mano y la descripción de los estados mentales. Pero esa diferencia no sitúa a la mente más allá de cualquier objetividad. Simplemente, sitúa a la mente en otra objetividad diferente.

También hay una diferencia crucial entre entender estas mismas frases aquí escritas, las de esta página, por ejemplo, y llegar a descifrar en las manchas de un jaguar la frase mágica que te hace todopoderoso. Esto es lo que ocurre en el cuento de Borges *La escritura del dios*²⁵. Tzinacán, sacerdote de la pirámide de Qaholom, no ha querido revelar a los españoles dónde se encuentra el tesoro. Torturado, se pudre en prisión. Muy cerca de él, en otra celda, hay un jaguar. La mente del sacerdote divaga. Se pierde entre recuerdos. Acaba fijándose en las manchas de la piel del jaguar. Y tras largo tiempo y gran esfuerzo, descubre en ellas un mensaje divino capaz de proporcionarle no solo su liberación personal y la de su pueblo, sino un poder máximo sobre todas las cosas. Pero justamente la perspectiva de esa posibilidad hace que se sienta indiferente ante lo inmediato. Al final es feliz aceptando plenamente su situación y su destino.

Comparemos ahora entender estas mismas frases, las que están escritas en esta página, con llegar a descifrar ese mensaje divino en las manchas de un jaguar. En el segundo caso, es dudoso que pueda llegar a constituirse la objetividad adecuada capaz de rescatar ese contenido de un ámbito meramente subjetivo. Sin embargo, en el caso del significado de lo que aquí está escrito sí existe esa objetividad. Y lo que estamos argumentando es que no hay razón para creer que no pueda llegar a existir también una objetividad análoga en el caso de la mente.

Que estas palabras y frases signifiquen lo que significan no es un asunto meramente subjetivo. Hay un espacio importante de objetividad. Y esta objetividad constituye una novedad respecto a la objetividad que solemos reconocer en los hechos físicos. Ha sido creada por nuestros puntos de vista. Y nuestra subjetividad también ha cambiado al verse reflejada en esa nueva objetividad. A esa nueva objetividad le acompaña una nueva subjetividad. Nada de esto implica que los puntos de vista se sitúen fuera de la realidad. Muy al contrario, esta dinámica de creación de nuevas objetividades y nuevas subjetividades ha de verse como una parte esencial de la propia realidad.

¿Qué tiene todo esto que ver con el reconocimiento de la inteligencia artificial en sentido fuerte? La conexión es directa. Con la inteligencia artificial está empezando ya a existir una nueva objetividad creada por nuestros puntos de vista, una nueva objetividad capaz de provocar cambios en nuestra subjetividad. Al igual que existe ya, desde hace algún tiempo, una nueva y muy importante objetividad respecto a la identidad neuronal de muchos de nuestros estados mentales, o al menos respecto a algunos de sus aspectos²⁶.

Son nuevas objetividades gracias a complejas redes de usos, historias y, en general, perspectivas. Y a esas nuevas objetividades les acompañan nuevas subjetividades.

En el caso de la identidad neuronal de nuestros estados mentales, o de algunos aspectos suyos, nuestra subjetividad es de hecho muy diferente de la existente hace siglos. No se trata de cambios en nuestras creencias periféricas. Se trata sobre todo de actitudes y emociones, de la manera como llegamos a ver el mundo y a vernos a nosotros mismos. Nuestra mente ha cambiado. El caso de la inteligencia artificial, es mucho más reciente. Y mucho más difícil de asumir. Sobre todo, porque no se trata solo de nuestra mente. Se trata de otros tipos de mentes. Es más, se trata de otras mentes que, aun de poder realmente existir, ni siquiera existen aún.

Debemos repetir ahora nuestras tesis. El mundo que están configurado fenómenos como la realidad virtual, la realidad aumentada, la inteligencia artificial, la humanidad mejorada y la Internet de las cosas solo será un nuevo mundo si la inteligencia artificial se asume en un sentido muy fuerte, en ese sentido que implica la posibilidad de que realmente existan mentes artificiales. Y para poder asumir la inteligencia artificial en ese sentido fuerte, seguramente deba cambiar nuestra propia mente. No solo algunas de nuestras creencias periféricas sobre lo que es la mente, sino nuestras creencias más profundas. Y sobre todo, nuestras actitudes y nuestras emociones, nuestra forma de ver el mundo y de vernos a nosotros mismos.

Para poder estar ante un auténtico *mundo nuevo*, nosotros mismos también debemos cambiar. Nuestra mente ha de cambiar radicalmente. Sin este último cambio, todo lo que pueda llegar a haber ya lo habremos visto antes. No será un mundo nuevo.

El mundo de los robots de Asimov ilustra perfectamente esta idea. Pues, aunque a primera vista lo parezca, tampoco es un mundo realmente nuevo. Los robots de Asimov solo son construcciones nuestras. Tan solo siguen mecánicamente reglas. Y entre esas reglas están las famosas leyes de la robótica que nos protegen de funcionamientos no deseados. Los robots de Asimov son únicamente *siervos*, esclavos nuestros. Son conocidas las opiniones de Asimov sobre el poder de la robótica. Los robots no razonan realmente, simplemente sus procesos son lógicos²⁷. Un mundo lleno de inteligencia artificial tan solo en este sentido no es un mundo nuevo. Es el mundo que ya tenemos con mucha mayor sofisticación y barroquismo técnico. Y desde luego, no es un mundo en el que, en caso de conflicto grave entre las máquinas y la humanidad, alguien como Sheldon Cooper pudiera firmar sin dudarlo a favor de las máquinas.

AGRADECIMIENTOS

Este trabajo ha sido realizado en el marco de los Proyectos de Investigación *Puntos de vista, disposiciones y tiempo. Perspectivas en un mundo de disposiciones* (FFI2014-57409-R) y *Perspectivas personales. Conceptos y aplicaciones* (FFI2018-098254-B-100), financiados por el Ministerio de Economía y Competitividad del Gobierno de España. Una versión previa fue presentada como charla en el contexto de las actividades del grupo de investigación *LOGOS* (Universidad de Barcelona) en abril de 2017. Tengo una gran deuda con su organizador, el profesor Josep Macià. Mi agradecimiento también al gran interés mostrado por el público asistente y a todas las preguntas suscitadas.

NOTAS

1. Las fechas que se mencionarán en películas y series corresponderán siempre a los estrenos de las primeras versiones.
2. Lanier (2013) señala repetidamente la falta de términos y conceptos que permitan siquiera plantear con claridad los problemas. Este autor, Jaron Lanier, junto con Nick Bostrom, constituyen notables excepciones a lo que estamos diciendo. Véase especialmente Bostrom (2014) y Lanier (2010), Lanier (2013) y Lanier (2017). En nuestro ámbito cultural más cercano, véase también el reciente e interesante libro de José Ignacio Latorre (2019).
3. Otra excepción al clima general de indiferencia académica respecto a los fenómenos que estamos considerando es David Chalmers. La inteligencia artificial siempre fue una de sus obsesiones. Y desde hace varios años también se ha interesado seriamente por cosas como la realidad virtual (al hilo de las sugerencias metafísicas de *Matrix*), la realidad aumentada (analizando las implicaciones de algo como *Pokemon Go*) o la humanidad mejorada (haciéndonos preguntas aparentemente tan extrañas como “¿Es tu móvil parte de tu mente?”). Véase, por ejemplo, Chalmers (2009). Otra excepción más, también sumamente recomendable, esta vez procedente del campo de la estética, es Michael Heim. Véase Heim (1993) y Heim (1998). Con todo, y a pesar de lo que nuestras últimas notas puedan sugerir, la reflexión filosófica sobre los fenómenos mencionados no ha hecho más que empezar.
4. Sobre las posibilidades y límites de la realidad virtual véase Lanier (2017). Y sobre los aspectos más metafísicos del tema sigue resultando interesante Heim (1993 y 1998). Chalmers (2009) discute algunas ideas de este último autor. Entre los límites de la realidad virtual que identifica Lanier, hay uno de gran relevancia filosófica. De momento, la tecnología de la realidad virtual no permite compartir experiencias *privadas*. Podemos tener experiencias provocadas por escenarios y situaciones virtuales compartidas, pero cada cual tiene sus propias experiencias. Y estas últimas son intransferibles (como decimos, lo son de momento).
5. El libro de Asimov sirvió a su vez de base a la película *I Robot* (2004), de Alex Proyas. El grupo musical *The Alan Parsons Project* se inspiró también en esta obra de Asimov al producir en 1977 su álbum, de carácter fuertemente conceptual, *I Robot*.
6. Una interesante y muy amena historia de los antepasados de nuestros actuales robots se encuentra en Peirano y Bueno (2009).
7. Ray Kurzweil es uno de los autores actuales más entusiastas respecto a los avances de la inteligencia artificial. Su pronóstico es que en torno al 2029 existirán ya inteligencias artificiales capaces de superar sin problemas el Test de Turing (véase Kurzweil, 2012). Ciertamente, todos sus planteamientos son ajenos al difícil problema de la conciencia. Se sitúan en el nivel intermedio que hemos identificado al distinguir varios sentidos en la expresión *inteligencia artificial*. Justamente a partir de este punto adquiere relevancia nuestra tesis sobre el carácter crucial de ciertas actitudes ante esas supuestas inteligencias artificiales.
8. Véase la ya clásica compilación de trabajos de Savulescu y Bostrom (eds.) (2009). Una posición sumamente equilibrada puede encontrarse en Diéguez-Lucena (2017).
9. Los análisis de Lanier (2013) son aquí especialmente pertinentes. Se basan en la gran plusvalía producida por los usuarios (por ejemplo, en las redes sociales) que no revierte en ellos. Y aboga por una redistribución de esa riqueza generada digitalmente. Esto implica plantear una política digital y una economía digital que aún no existen.
10. Bostrom (2014) argumenta que, a menos que se paralicen los procesos actuales, solo esa nueva ética sería capaz de alejar la probabilidad de lo que podría llegar a ser una última y definitiva catástrofe existencial de la humanidad tal como hoy día la conocemos. Esta es también la razonable propuesta que desarrolla con gran convicción Latorre (2019).
11. Muchas veces, Ulises me dice que “le copio todas las ideas”. Esto es literalmente verdad en este trabajo. Las principales ideas que aquí se presentan surgen de manera muy natural en alguien que ha sido, y se siente, plenamente *nativo* de Internet.
12. Se ha dicho que la ciencia del siglo XXI tiene cuatro elementos básicos: el átomo, el bit, el gen y la neurona. Las tecnologías biomédicas trabajan cada vez más con los tres últimos. Las tecnologías computacionales con los dos primeros. El bit es el elemento compartido.
13. Según la llamada *tercera ley* de Clark, “cualquier tecnología suficientemente avanzada es indistinguible de la magia”. La idea que esto expresa es fascinante. Fue formulada por Arthur C. Clarke en la reedición de 1973 de su libro de 1962 *Profiles of Future* (véase Clark, 1973).
14. Nuestra propia ciencia sugiere tal hipótesis en dos frentes muy distintos. De una parte, las ciencias cognitivas, especialmente la psicología de la percepción, han llegado a elaborar una concepción de la experiencia de acuerdo a la cual todo lo que creemos tener a nuestro alrededor no es más que una construcción o proyección nuestra. De otra parte, está la física. Si las actuales concepciones holográficas de nuestro universo son correctas, todo lo que observamos puede no ser más que una especie de inmenso holograma generado por una realidad más básica con menos dimensiones.
15. Son muchas las publicaciones donde John Searle ha presentado este argumento. Véase, por ejemplo, Searle (1980), Searle (1985) y Searle (1990).
16. Esto ocurriría en las arquitecturas computacionales basadas en las redes neuronales. El argumento de Searle podría seguirse sosteniendo en estos casos. Hasta llegar a cierto punto en el que ya no hubiera mera simulación de una mente, sino una clara duplicación.
17. Claramente es así en los campos de la elaboración y procesamiento de históricos médicos, del diagnóstico mediante sistemas expertos y de la incorporación de procesos de automatización en la administración de medicamentos. También en el diseño de todo tipo de prótesis. Y en el desarrollo eficiente de tecnologías genéticas, como por ejemplo el CRISPR.
18. Uno de los principales orígenes de este problema es, por supuesto, el clásico trabajo de Turing (1950).
19. Véase de nuevo Bostrom (2014), en particular el capítulo 3 dedicado a las

distintas formas que podría adoptar una *superinteligencia*. Latorre (2019) también aborda este tema. A continuación, ofreceremos una serie de testimonios sumamente reveladores.

20. La inteligencia artificial está sustituyendo a vampiros y *zombies* en el cine y las series televisivas. Véase al respecto la interesante serie *West World* (2016), donde la inteligencia artificial y la realidad aumentada se mezclan en un inmenso parque temático ambientado en el Lejano Oeste. Parte de la diversión consiste en tratar con crueldad y saña a los robots humanoides que pueblan el parque. Pero claro, su *dolor* es solo un dolor simulado. ¿O ya no lo es? En cualquier caso, de seguir las cosas en esta línea, en algún momento deberán convertirse todas estas cuestiones en materia de legislación positiva. Incluso tal vez ya sea hora de plantear ciertos límites éticos y legales a la investigación en inteligencia artificial. Véase al respecto Latorre (2019). La marcha de los acontecimientos ha sobrepasado hace mucho el horizonte de las tres leyes de la robótica establecidas por Asimov. En el apéndice I del libro de Latorre se recogen algunas importantes declaraciones institucionales de principios éticos que deberían regular el desarrollo de la inteligencia artificial.
21. Aspectos y sentidos son nociones correlativas. Cada *aspecto* de una cosa ofrece un *sentido* diferente en el que podemos hablar de ella. Sobre la propia noción de

punto de vista, y las diferentes maneras de conceptualizarla, véase Liz (ed.) (2013) y también Vázquez y Liz (eds.) (2015).

22. Lanier (2010) se planteaba cómo entender la condición personal en nuestra era digital. El problema es grave, pues la tecnología es una extensión nuestra que podría llegar a cambiar nuestra misma condición. Su conclusión final es que las personas somos un *misterio*. Pero esto no le impide hacer juicios antropocéntricos sobre el deber que tenemos de preservar tal misterio. En nuestra opinión, un misterio impide esto. Y si estamos abiertos a este misterio, también deberíamos estar abiertos al misterio de una mente artificial, en el sentido más completo de la expresión.
23. El esquema de descripción que hemos ofrecido para los estados mentales intenta integrar todos los factores que suelen considerarse relevantes: tipos de estados neurológicos N, tipos de papeles computacionales C, tipos de sujetos S y tipos de relaciones R entre los sujetos y sus entornos.
24. Véase Jackson (1982) y Jackson (1986). Inicialmente este experimento mental fue propuesto por Jackson como un problema para el fisicalismo. Un conocimiento físico y neurológico completo siempre *parece* dejar fuera algunos hechos. No todos los hechos *parecen* ser hechos físicos. Nagel (1974) presenta un planteamiento estrechamente rela-

cionado, usando como ejemplo los murciélagos, su peculiar manera de *ver* el mundo mediante sistemas de ecolocación. Aun conociendo todos los hechos físicos relevantes, *parece* que nosotros no podríamos llegar a saber qué es ver el mundo de esa forma. Supuestamente ciertos hechos solo se conocen experimentándolos en primera persona.

25. Incluido en *El Aleph*. Una de las últimas ediciones de este libro, publicado originalmente en 1949, es Borges (2003).
26. La formación de esa objetividad debe simplemente aceptarse como un hecho. Y como un hecho, por cierto, sumamente revelador de la propia naturaleza de los *hechos objetivos*.
27. En su libro *I Robot*, originalmente de 1950, se presentan las famosas tres leyes de la robótica. Pero el objetivo de estas leyes es establecer un marco apropiado para la programación de una cierta autonomía conductual en los robots, de manera que, a pesar de ser posibles ciertas situaciones conflictivas, incluso a veces paradójicas, se disipen por nuestra parte los miedos respecto al proyecto de construir robots. No es el objetivo de estas leyes constituir una especie de *código ético* para unos agentes morales plenamente libres y responsables. Esta diferencia es crucial. Páginas atrás mencionábamos la película *I Robot* (2004), de Alex Proyas, basada en el libro de Asimov. Una reciente edición castellana es Asimov (2007).

BIBLIOGRAFÍA

- Asimov, I. (1950/2007). *Yo, Robot*. Barcelona: Edhasa.
- Borges, J. L. (1949/2003). *El Aleph*. Madrid: Alianza.
- Bostrom, N. (2014). *Superintelligence. Paths, Dangers, Strategies*. Oxford: Oxford University Press.
- Chalmers, D. (2009). The Matrix as Metaphysics. [En línea]. Disponible en <http://consc.net/papers/matrix.html>
- Clark, A. (1973). *Profiles of Future*. Londres: Popular Library.
- Diéguez-Lucena, A. (2017). *Transhumanismo. La búsqueda tecnológica del mejoramiento humano*. Barcelona: Herder.
- Heim, M. (1993). *The Metaphysics of Virtual Reality*. New York: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195092585.001.0001>
- Heim, M. (1998). *Virtual Realism*. New York: Oxford University Press.
- Jackson, F. (1982). Epiphenomenal Qualia. *Philosophical Quarterly*, 32 (127), pp. 127-136. <https://doi.org/10.2307/2960077>
- Jackson, F. (1986). What Mary didn't know. *Journal of Philosophy*, 83 (5), pp. 291-295. <https://doi.org/10.2307/2026143>
- Kurzweil, R. (2012). *How to create a mind: The secret of human thought*. Clusen & Bosse.
- Lanier, J. (2010). *You are not a Gadget: A Manifesto*. New York: Random House.
- Lanier, J. (2013). *Who Owns the Future*. New York: Simon & Schuster.
- Lanier, J. (2017). *Dawn of the New Everthing. Encounters with reality and virtual reality*. New York: Henry Holt Company.
- Latorre, J. (2019). *Ética para máquinas*. Barcelona: Ariel.
- Liz, M. (ed.) (2013). *Puntos de vista. Una investigación filosófica*. Barcelona: Laertes.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83 (4), pp. 435-450. <https://doi.org/10.2307/2183914>
- Peirano, M. y Bueno, S. (2009). *El rival de Prometeo. Vidas de autómatas ilustres*. Madrid: Impedimenta.
- Savulescu, J. y Bostrom, N. (eds.) (2009). *Human Enhancement*. Oxford: Oxford University Press.
- Searle, J. (1980). Minds, Brains, and Programs. *The Behavioral and Brain Sciences*, 3 (3), pp. 417-424. <https://doi.org/10.1017/S0140525X00005756>

Searle, J. (1985). *Minds, Brains and Science. The 1984 Reith Lectures*. Cambridge: Harvard University Press.

Searle, J. (1990). ¿Es la mente un programa informático? *Investigación y Ciencia*, 162, pp. 9-16.

Turing, A. (1950). Computing Machinery and Intelligence. *Mind*, 59 (236), pp. 433-460. <https://doi.org/10.1093/mind/LIX.236.433>

Vázquez, M. y Liz, M. (eds.) (2015). *Temporal Points of View. Subjective and Objective Aspects*. Heidelberg: Springer. <https://doi.org/10.1007/978-3-319-19815-6>



a572

Antonio Manuel Liz Gutiérrez