
Inteligencia artificial y robótica

PID_00270335

Pablo Jiménez Schlegl
Carme Torras Genís

Tiempo mínimo de dedicación recomendado: 6 horas



**Pablo Jiménez Schlegl**

Pablo Jiménez Schlegl es doctor ingeniero industrial por la UPC y científico titular del CSIC en el Institut de Robòtica i Informàtica Industrial (CSIC-UPC), donde actualmente ocupa el cargo de jefe de departamento. Sus intereses científicos incluyen la planificación de tareas, el aprendizaje, y más recientemente la manipulación robótica de objetos deformables de uso cotidiano, como la ropa, con el objetivo de contribuir a la robótica asistencial. Últimamente también se ha interesado por temas relacionados con el impacto de la robótica en la sociedad humana desde un punto de vista ético.

**Carme Torras Genís**

Carme Torras es licenciada en matemáticas, doctora en informática y profesora de investigación en el Instituto de Robótica e Informàtica Industrial (CSIC-UPC) en Barcelona, donde dirige un grupo de investigación en robótica asistencial.

Es *Fellow* y EurAI, miembro numerario de la Real Academia de Ciencias y Artes de Barcelona, y de la Academia Europaea; y ha obtenido reconocimientos como la medalla Narcís Monturiol de la Generalitat de Catalunya y el premio Julio Peláez a las mujeres pioneras en ciencias físicas, químicas y matemáticas. Su novela de ciencia ficción, *La mutación sentimental*, fue publicada en inglés con el título *The Vestigial Heart*.

El encargo y la creación de este recurso de aprendizaje UOC han sido coordinados por la profesora: Mònica Vilasau Solana (2020)

Primera edición: febrero 2020
© Pablo Jiménez Schlegl, Carme Torras Genís
Todos los derechos reservados
© de esta edición, FUOC, 2020
Av. Tibidabo, 39-43, 08035 Barcelona
Realización editorial: FUOC

Ninguna parte de esta publicación, incluido el diseño general y la cubierta, puede ser copiada, reproducida, almacenada o transmitida de ninguna forma, ni por ningún medio, sea este eléctrico, químico, mecánico, óptico, grabación, fotocopia, o cualquier otro, sin la previa autorización escrita de los titulares de los derechos.

Índice

1. Robótica e IA para profanos.....	5
1.1. Robots e IA	5
1.2. ¿Qué es un robot?	7
1.3. Aplicaciones	12
1.4. Técnicas de IA en robótica	18
2. Consideraciones éticas.....	22
2.1. ¿Es posible construir máquinas éticas?	23
2.2. Relevancia ética de los robots en diferentes ámbitos de la experiencia humana	29
2.2.1. El ámbito laboral: impacto del robot sobre el mercado de trabajo	29
2.2.2. El ámbito del transporte: vehículos autónomos	32
2.2.3. El ámbito del conflicto: robots militares	35
2.2.4. El ámbito de la salud: prótesis, sistemas expertos y robots cirujanos	38
2.2.5. El ámbito doméstico: seguridad, privacidad, respeto a la intimidad	41
2.2.6. El ámbito de la intimidad: relaciones afectivas	43
3. Iniciativas legales.....	48
3.1. IA y el mundo de las leyes	48
3.2. Responsabilidad civil y penal de los actores implicados en el desarrollo del robot inteligente	50
3.2.1. Accidentes: los robots pueden causar daño	50
3.2.2. Un marco legal para robots	53
3.3. Iniciativas legales en el ámbito europeo	58
3.3.1. Precedentes	58
3.3.2. Iniciativa de legislación del Parlamento Europeo	59
3.4. Iniciativas normativas y legales en el ámbito internacional y extracomunitario	61
3.4.1. Normas ISO/TC 299 Robotics	61
3.4.2. Ámbito extracomunitario	63
3.5. Derechos del robot: visiones de futuro	65
Resumen.....	67
Bibliografía.....	69

1. Robótica e IA para profanos

No es necesario ser un experto tecnólogo, ni tan solo un aficionado a la ciencia-ficción, para que los términos de inteligencia artificial y de robot tengan muchas connotaciones familiares. También sabemos que empiezan a desempeñar un papel cada vez más relevante en nuestras vidas. Sabemos que en muchos aspectos nos harán la vida más fácil (tal es su objetivo), pero que también serán fuente y objeto de conflictos, disputas, accidentes, vulnerabilidades y otros efectos negativos. Lo cierto es que han venido para quedarse, y de los humanos depende que su desarrollo y expansión no solo sea inocuo, sino también beneficioso en conjunto. Por ello es decisivo que, en primer lugar, sepamos de qué estamos hablando realmente. Tal es el objetivo de esta primera parte: sentar las bases terminológicas y de concepto.

1.1. Robots e IA

La **inteligencia artificial** (IA) comprende todo el conjunto de teorías, métodos y algoritmos que pretenden emular procesos cognitivos. Aunque existen otros tipos de inteligencia en el planeta (que reviste formas variadas como la inteligencia mostrada por mamíferos o la inteligencia colectiva de enjambres y hormigueros), el principal referente de la IA es la inteligencia humana. En su libro divulgativo (López de Mántaras y Meseguer, 2017), los expertos en IA López de Mántaras y Meseguer plantean la doble pregunta de si es posible construir máquinas inteligentes, y si el cerebro es una máquina. Como afirman estos autores, «el objetivo último de la IA –lograr que una máquina tenga una inteligencia *general* similar a la humana– es uno de los más ambiciosos que se ha planteado la ciencia» (López de Mántaras y Meseguer, 2017, pág. 7). El énfasis, del original, da pie a diferenciar entre la IA *específica* y la *general*, siendo la primera la inherente a los sistemas expertos (SE) y los programas capaces de jugar al ajedrez (llegando a derrotar a campeones mundiales) o al póquer, mientras que la segunda ha de exhibir la versatilidad de la inteligencia humana. Otra distinción igualmente destacada es entre IA *débil* y *fuerte*, siendo la primera sinónimo de la específica, mientras que la segunda, además de ser general (el recíproco no es cierto, es decir, no son sinónimos) va más allá de reproducir o simular los procesos mentales: *es una mente*. Algunos autores, como el filósofo John Searle, cuestionan que se pueda construir una IA fuerte, que ilustra con el experimento mental de la habitación china: una persona equipada con un manual inglés-chino en una habitación puede proporcionar traducciones al inglés de los ideogramas que va recibiendo, sin que ello signifique que realmente sabe chino (es decir, se limita a establecer correspondencias). Al margen de esta discusión, lo cierto es que la IA débil ha proporcionado numerosos éxitos y aplicaciones de gran utilidad, como los ya mencionados

sistemas expertos, y en lugar de duplicar la actividad mental humana, lo que puede hacer de forma muy eficiente es complementarla gracias a sus capacidades de procesamiento y de gestión de la memoria.

Siguiendo con el texto de López de Mántaras y Meseguer, existen cuatro tipos de modelos de IA:

- **Simbólico.** Se basa en lo que se denomina la hipótesis de los sistemas de símbolos físicos (SSF), que afirma «todo sistema de símbolos físicos posee los medios necesarios y suficientes para llevar a cabo acciones inteligentes». Este ha sido el modelo dominante en IA y es «un modelo *top-down* que se basa en el razonamiento lógico y la búsqueda heurística como pilares para la resolución de problemas» (López de Mántaras y Meseguer, 2017, pág. 11). La mayoría de los sistemas expertos a los que aludíamos más arriba están basados en este modelo, los símbolos que manejan son palabras de la terminología característica del dominio especializado en cuestión (médico, legal, ingenieril, etc.).
- **Conexionista.** Es un modelo *bottom-up* que considera que la actividad inteligente emerge a partir de procesos más simples altamente interconectados. La versión más extendida es la de la red neuronal artificial, donde los procesos simples se modelizan como neuronas, las interconexiones corresponden a las sinapsis, y los procesos de entrenamiento de la red (modificación local de la intensidad de las conexiones) emulan el aprendizaje de las correlaciones entre entradas y salidas del sistema.
- **Evolutivo.** En este modelo, los comportamientos inteligentes emergen en un sistema que emula los mecanismos de la evolución biológica: la información contenida en cromosomas está sujeta a mutaciones aleatorias, a las leyes de la herencia genética y de la selección natural. Los individuos que logran sobrevivir y prosperar representan procesos mentales mejor adaptados a la resolución de los problemas planteados por un entorno específico (que representa la categoría de problemas a cuya resolución va dirigido el sistema).
- **Corpóreo.** En este modelo se pone énfasis en la *cognición situada*: para dotar a los mecanismos cognitivos de contenido semántico, es necesario que estos experimenten su entorno directamente mediante sensores y puedan también modificarlo a través de sus acciones. Como dicen López de Mántaras y Meseguer, muchos investigadores (entre los cuales ellos mismos) están convencidos de que la aproximación corpórea es «imprescindible para avanzar hacia inteligencias de tipo general», y que «el cuerpo da forma a la inteligencia» (López de Mántaras y Meseguer, 2017, pág. 14).

Allen Newell y Herbert Simon

López de Mántaras y Meseguer citan a Allen Newell y Herbert Simon en su ponencia al recibir el premio Turing de 1975.

Reflexión

Aunque el modelo conexionista exhibe evidentes paralelismos con el funcionamiento del cerebro, está todavía muy lejos de capturar su inmensa complejidad. Véase al respecto López de Mántaras y Meseguer (2017, pág. 13).

La inteligencia humana es obviamente fuente de inspiración de todos estos modelos, pero así como el primero es un intento de formulación abstracta de los procesos mentales, los otros tres pretenden emular los mecanismos físicos

que han llevado a la misma. De todos ellos es el último el que más nos interesa en el presente curso. Se ha de subrayar que no es incompatible con los anteriores: de hecho, puede recurrir a aquellos como modelos de representación interna. El cuerpo en cuestión, como la inteligencia a la que «encarna», es artificial. Por supuesto, estamos hablando del robot.

En definitiva, un robot da cuerpo a la inteligencia artificial, le permite percibir el mundo de primera mano, moverse por el entorno físico, actuar. Recíprocamente, gracias a las técnicas de la IA, el robot no es una máquina más: es una máquina capaz de recordar, aprender, planificar, razonar, tomar decisiones más o menos complejas. Por descontado, la IA incorporada tiene muchas aplicaciones e implicaciones éticas y legales por los daños que potencialmente puede ocasionar. A lo largo de este módulo se comentarán los casos más relevantes, pero el énfasis se pondrá en la IA que actúa en el mundo físico a través de la robótica. Para ello, en primer lugar debemos aclarar qué es un robot.

1.2. ¿Qué es un robot?

Todo el mundo sabe qué es un robot. La palabra, acuñada hace un siglo por el escritor checo Karel Čapek en su obra de 1920 *R.U.R. (Rossum's Universal Robots)*, evoca preferentemente a un humanoide artificial, metálico, que realiza movimientos bruscos, sincopados, y tiene voz enlatada. La ficción televisiva y cinematográfica ha ido ampliando el rango de acepciones de dicha imagen, por lo que el imaginario popular no tiene inconveniente en admitir bajo la misma categoría al símil de papelera o aspiradora rodante que es R2D2 de *Star Wars*, y los replicantes prácticamente indistinguibles de los humanos de *Blade Runner*.

Pero al mismo tiempo muy pocos saben qué es un robot realmente. Un robot ¿ha de tener forma humana, siquiera vagamente? ¿Un autómatas del siglo XVII es un robot? Si la forma humana no es relevante, ¿puede un vehículo autónomo considerarse un robot? ¿Y una lavadora inteligente? ¿Todas las máquinas que poseen capacidad de procesar información externa son robots? ¿Un PC es un robot? Uno de los padres de la robótica industrial (una categoría bien definida de robots), Joseph Engelberger, declaró en una ocasión que él no sabría definir «robot», pero que sabía reconocer uno en cuanto lo veía. No hay un consenso generalizado sobre la definición precisa de robot, y muchas discusiones sobre si determinada máquina puede calificarse como tal, pero sí hay cierto acuerdo en que el requisito fundamental es el que puede resumirse en el ciclo **percibe-piensa/planifica-actúa**:

Rossum's Universal Robots

Y estrenada el 25 de enero de 1921 en el Teatro Nacional de Praga. El término utilizado para designar a los humanoides fabricados para trabajar en lugar de los humanos le fue sugerido al dramaturgo por su hermano pintor, Josef.

- un robot ha de ser capaz de percibir su entorno mediante sensores (dependiendo de lo que se espera de la máquina dichos sensores serán más o menos sofisticados),
- ha de ser capaz de decidir, si existen diferentes alternativas, cuál es el curso de acción más conveniente para alcanzar los objetivos encomendados acorde con su programación (el término «pensar» incomoda a muchos científicos por designar una actividad humana sobre la que aún hay muchos interrogantes, y prefieren los términos más precisos de «toma de decisiones» para evaluar acciones de efectos inmediatos, o «planificación» si se trata de determinar una secuencia de acciones que llevarán a la meta),
- ha de ser capaz de llevar a cabo acciones en un entorno físico a través de sus actuadores, acciones que van desde meros desplazamientos hasta el traslado o manipulación de objetos, modificación de su propia estructura, interacción con personas, etc.

Según la Norma ISO 8373:2012, un robot es un

«mecanismo actuado programable en dos o más ejes con un grado de autonomía, que se mueve en su entorno, para realizar las tareas previstas».

Así pues, nadie calificaría un termostato de robot, pero tampoco una casa domótica donde un ordenador central, programado de acuerdo con las preferencias de sus habitantes, regula el conjunto de termostatos del hogar, además de las luces, persianas, limpieza, comunicaciones, o accesos: por sofisticada que sea, y a pesar de ser programable, falta la componente del *movimiento* en un entorno (este argumento no es incontrovertible, ya que la apertura y cierre de ventanas conlleva el movimiento de una parte del conjunto, como en un robot reconfigurable)¹. Más sutil es el caso de un automatismo de actuadores neumáticos (sistemas cilindro-émbolo actuados por aire a presión): puede ser programado si es controlado mediante un PLC, pero esta programación solo se refiere a la *secuencia*, el orden en que son activados los actuadores, pero no se puede variar por programa la *posición* de los mismos, que viene dada por los extremos del recorrido del émbolo. Por consiguiente, tampoco son robots, ya que estos pueden posicionarse en cualquier punto de su espacio de trabajo.

De acuerdo con su aspecto físico, muy ligado a su funcionalidad, los robots pueden clasificarse de acuerdo con la siguiente tipología:

- **Brazos manipuladores o brazos robóticos** (*manipulators, robotic arms*). Es el tipo más extendido en la industria (véase el subapartado 1.3), y también está muy presente en los laboratorios de investigación. Son mecanismos poliarticulados, accionados por motores, con un elemento terminal en un extremo que puede ser una garra, mano o pinza para agarrar y desplazar objetos o realizar algún tipo de manipulación, o una herramienta (de corte, soldadura, pintura, etc.). Habitualmente son fijos (la base está anclada

Behavioral robotics

También ha habido propuestas, como la de la robótica de comportamientos (*behavioral robotics*) de Rodney Brooks, que enlazaban directamente percibe-actúa en multitud de comportamientos elementales, de cuya acción conjunta emergía un comportamiento más sofisticado, arguyendo que el mundo físico es su mejor representación. Más adelante se decantaron por una arquitectura híbrida, en la que al menos a alto nivel se producía cierta actividad deliberativa.

Bibliografía

<https://www.iso.org/obp/ui/#iso:std:iso:8373:ed-2:v1:en>

⁽¹⁾En comparación, un brazo robot como el que se describe a continuación, aunque su base esté fija, se mueve dentro de su espacio de trabajo.

Un sistema cilindro-émbolo...

sería una jeringa, o los cilindros del motor de un coche. En un actuador neumático, en lugar de la fuerza ejercida por los dedos, es el aire a presión que se introduce por uno de los lados del cilindro lo que hace mover el émbolo.

PLC

Controlador lógico programable, por sus siglas en inglés: es un tipo de miniordenador en el que se puede programar una secuencia de activación de válvulas (en este caso) que permiten o interrumpen el paso del aire a los extremos del cilindro, moviendo el émbolo a un lado o al otro.

en el suelo o suspendida de una estructura), pero también pueden estar montados sobre una plataforma móvil. Un robot como Baxter, diseñado para colaborar con trabajadores humanos, consistente en dos brazos montados sobre una columna que alberga el ordenador de control y con una pantalla que muestra expresiones faciales estaría en la frontera entre los brazos manipuladores y los robots humanoides. Más abajo se describen los componentes y principios de funcionamiento de un brazo robot.

- **Robots móviles.** Su principal característica es que son capaces de desplazarse de un lugar a otro. Suelen hacerlo sobre ruedas, si el suelo es más o menos liso, y sobre patas u orugas si el terreno es abrupto (o ruedas montadas sobre mecanismos con cierta flexibilidad, como los *rovers* de exploración). También hay robots que se desplazan por el medio acuático (robots experimentales, de inspección y de exploración) y aéreo (vehículos aéreos no tripulados, drones). En esta categoría se incluyen desde los AGV (*automated/autonomous guided vehicle*) empleados en logística industrial hasta el vehículo autónomo, del que se hablará extensamente en el subapartado 2.2.2.
- **Robots humanoides.** Son robots de apariencia antropomórfica, aunque esta similitud no necesariamente ha de ser muy fidedigna, aceptándose como tales incluso robots que se desplazan sobre ruedas o torsos (con brazos y cabeza) fijos. Lo más relevante es que disponga de extremidades superiores (preferentemente dos) y algo que pueda interpretarse como un rostro para facilitar la comunicación. Algunos investigadores ponen el énfasis en la funcionalidad, desarrollando robots que se puedan desplazar sobre piernas y coger cosas con los brazos (desde el famoso ASIMO de Honda hasta el impresionante Atlas de Boston Dynamics), mientras que otros buscan un aspecto lo más humano posible, aunque de momento la motricidad de las extremidades inferiores es muy limitada (por ejemplo, las *actroid* de la Universidad de Osaka/Kokoro Ltd, Sophia de Hanson Robotics, o las muñecas y muñecos sexuales de RealDoll).
- **Robots experimentales.** En este cajón de sastre incluimos desde robots biomiméticos hasta robots autoreconfigurables, micro y nanorobots, robots accionados por cables, y otras pruebas de concepto tanto mecánicas como de control.

En un robot cabe distinguir el hardware y el software:

- **Hardware:** Es la parte física del robot, su cuerpo, a través del cual interactúa con su entorno. A su vez consta de las siguientes partes:
 - **Estructura.** En el caso de un brazo, la estructura consiste en una serie de elementos rígidos conectados entre sí mediante **articulaciones**, que pueden ser rotativas (giratorias, como el codo humano) o lineales (prismáticas, como el par cilindro-émbolo). Dichos elementos pueden estar conectados en serie (es decir, uno detrás de otro, como en el bra-

zo humano), o en paralelo (como por ejemplo la plataforma de un simulador de vuelo). En el caso de un robot móvil, la estructura sería la plataforma que contiene los motores, sensores y electrónica, y si es un robot caminante también podríamos incluir la estructura de cada pata. De forma más genérica, la estructura es lo que da consistencia al cuerpo del robot, contiene algunos elementos como actuadores y sensores, y permite moverse al robot. Puede establecerse una analogía anatómica con el esqueleto.

- **Actuadores o motores.** Para seguir con la analogía, corresponderían a la musculatura del cuerpo. Son transductores de energía (generalmente eléctrica) en movimiento. La mayoría son motores eléctricos, que activan tanto las articulaciones de un brazo o de una pierna, como las ruedas de un vehículo. También pueden actuar sobre el cierre y abertura de una mano o pinza, el movimiento de la cabeza o de las cámaras, o de rasgos faciales de robots humanoides. Habitualmente incorporan algún tipo de reducción mecánica, para reducir la velocidad de salida del motor y aumentar el par (la fuerza), y un **codificador** (*encoder*) que es un dispositivo que «lee» la posición de un eje (esto es, de una articulación) y la transforma en una señal digital. Estas señales sirven al controlador del robot para registrar en todo momento la configuración (la pose) del robot, es lo que se conoce como **propiocepción**. En muchos casos será necesario algún tipo de **transmisión mecánica**, a través de engranajes, correas dentadas o incluso sistemas de cables y poleas, para llevar el movimiento generado en el actuador hasta la articulación a la que mueve (por ejemplo, en un brazo los motores suelen situarse cerca de la base para reducir el peso y las fuerzas de inercia del robot).
- **Sensores.** Son dispositivos que informan al robot sobre su propio estado (por ejemplo, los ya mencionados codificadores, también podríamos incluir los medidores de carga de la batería) y sobre el estado del entorno. Centrándonos en estos últimos, pueden ser tan simples como un detector de presencia o proximidad (por ejemplo, una célula fotoeléctrica) que informa sobre la presencia de una pieza que puede ser agarrada por el robot, o tan complejos como las cámaras de un sistema de visión, sensores de fuerzas y momentos montados en la muñeca de un robot (informan sobre las fuerzas ejercidas en la garra del robot), o la piel artificial (matrices de detectores de contacto, de presión, quizá también otras variables como la temperatura, etc.). Un robot también puede incorporar micrófonos para registrar los comandos verbales que le comunica oralmente el usuario.
- **Electrónica.** Comprende tanto la electrónica de potencia (la que regula el suministro energético a los actuadores) como la de control y el ordenador (con los típicos componentes como procesadores, memorias, disco duro, etc.). Este último se encarga del control a más alto nivel, básicamente de interpretar los programas almacenados en función de

la información suministrada por los sensores, y en consecuencia enviar señales de activación a los diferentes actuadores. Es la presencia del ordenador la que confiere su versatilidad al robot, su adaptabilidad, sofisticación y facilidad de reprogramación. El robot puede contar con una unidad de programación para introducir los programas, o estos pueden ser escritos en un PC de propósito general y luego transferidos a la memoria del ordenador del robot. En los brazos manipuladores, la electrónica de potencia, el ordenador y la unidad de programación suelen estar separados del brazo, integrados en un armario de control. En un robot móvil, en cambio, tanto la electrónica de control como la batería que suministra energía al robot van a bordo.

- **Elemento terminal.** Es el dispositivo que permite al robot realizar acciones sobre su entorno. El más común es la **pinza** (de dos dedos), la **garra** o la **mano** (este término suele aplicarse cuando hay más de dos dedos) para agarrar y desplazar objetos, o realizar otro tipo de manipulaciones sobre los mismos. Algunos investigadores intentan emular la versatilidad y la destreza de la mano humana, pero es un reto que no se alcanzará en un futuro próximo. Otros elementos terminales, como se ha apuntado antes, pueden ser **herramientas**: la más utilizada, por la masiva presencia de robots en la industria automovilística, es la pinza de soldadura por puntos, también son frecuentes las herramientas de corte (convencionales, por láser, plasma o chorro de agua), las pistolas de aplicación de pinturas y lacas, herramientas de lijado y pulido, entre otras. Los robots de exploración pueden llevar dispositivos medidores y de recogida de muestras.
- **Otros dispositivos.** Dependiendo de sus características, el robot puede incorporar dispositivos que faciliten la interacción del robot con el usuario como una pantalla gráfica o un sintetizador de voz y altavoces para la comunicación oral, elementos estilísticos o decorativos, etc.
- **Software:** Es la parte intangible del robot, su **programación**. En su forma más simple, consiste en una secuencia de especificaciones de acciones expresadas en el lenguaje de programación del robot, que incluye un repertorio limitado de instrucciones de movimiento (que se interpretan como algo del tipo «ve a la posición 12 a velocidad 0.8»), de accionamiento del elemento terminal («abre/cierra pinza»), e incluso de control del flujo del programa (bucles tipo «mientras haya piezas en la cinta transportadora, colócalas en cajas» o saltos condicionales «si hay pieza en el alimentador J entonces ejecuta las instrucciones 14-18, si no ejecuta las instrucciones 56-61»). Todo ello en un lenguaje formal propio del fabricante, y donde parámetros tales como «posición 12» o «alimentador J» se han especificado con anterioridad, por ejemplo, guiando manualmente el robot a las posiciones en cuestión y asignándolas dichos nombres. La mayoría de robots industriales actuales se programan de este modo, y todo lo más requieren simples formas de percepción (por ejemplo, detectar la presencia o ausen-

cia de una pieza en un lugar especificado). Ello es posible gracias a que estos robots trabajan en un entorno controlado, es decir, donde cada cosa está en su sitio y no hay imprevistos. La situación cambia radicalmente en el caso de robots de servicios, sociales o de exploración. En estos casos, el robot debe desenvolverse en un mundo cambiante, lleno de contingencias inesperadas, de información incompleta o no precisa. Es evidente que, en estas circunstancias, el robot requiere un sistema sensorial que le proporcione una descripción actualizada del entorno. Si el robot es una aspiradora autónoma que circula por la casa, no precisará un sistema de sensores muy sofisticado, bastará que detecte obstáculos y localice el cargador de la batería. El software de control podrá equipararse al de un insecto en el mejor de los casos. En cambio, si ha de interactuar con su propietario, distinguirlo de otras personas, conocer sus preferencias y mantener una conversación estimulante, necesitará un sistema robusto de reconocimiento facial (robusto en el sentido de que sea capaz de reconocerlo bajo distintas condiciones de iluminación, puntos de vista u opciones de peinado), así como diversas técnicas de IA de las que se hablará en el subapartado 1.4. En cualquier caso, el hecho de que el comportamiento esté condicionado por software, sea codificado por un programador o aprendido por observación es uno de los rasgos distintivos del robot. Por consiguiente, un autómatas como los que fueron muy populares entre los siglos XVIII y XIX no es un robot a pesar de su apariencia humanoide, la sofisticación de sus movimientos y su naturaleza mecánica, puesto que para variar la tarea que está realizando será necesario introducir modificaciones físicas en el mecanismo, como cambiar el sistema de levas que rige los movimientos de las manos de la clavecinista de los hermanos Droz. En cambio, un pariente próximo son las máquinas de control numérico o las impresoras 3D, que crean piezas físicas a partir de las especificaciones de un programa.

1.3. Aplicaciones

Los ámbitos de aplicación de los robots son múltiples y variados. El sector dominante es el industrial, por motivos históricos y económicos. Pero, poco a poco, van penetrando en otros sectores y realizando tareas que parecían reservadas a los humanos. Vamos a repasarlas brevemente:

- **Robots industriales.** Es el sector pionero: la invención del primer «aparato programado para la transferencia de artículos» de George Devol, en 1954, quien conoció al empresario Joseph Engelberger en 1956, marcó el inicio cuando ambos desarrollaron al año siguiente el primer prototipo, el Unimate 001 (Robotic Industries Association, 2019). En 1961, Engelberger fundó Unimation Inc., la primera empresa dedicada a fabricar robots industriales. El principal cliente de las compañías de fabricación de robots industriales fue, desde sus inicios, la industria automovilística, y aún hoy en día lidera el sector, seguida muy de cerca por la industria de aparatos y componentes electrónicos. Ello se debe a que en su momento la industria automovilística tenía la suficiente envergadura (y beneficios) como para

Robot industrial

La misma norma da definiciones para algunos de los términos. Así, *reprogramable* significa «diseñado de modo que los movimientos o funciones auxiliares programados pueden ser modificados sin alteración física», *multipropósito* es «capaz de ser adaptado a una aplicación diferente con alteración física» (por ejemplo, un cambio de elemento terminal, que también puede ser automático), y *manipulador* es una «máquina donde el mecanismo usualmente consiste en una serie de segmentos, articulados o deslizando uno respecto a otro, para el propósito de agarrar y/o mover objetos (piezas o herramientas) generalmente en varios grados de libertad».

poder permitirse apostar por una tecnología nueva, cuyos resultados todavía se desconocían. La producción en masa de sus productos, y el tipo de tareas encomendadas a los robots (tareas peligrosas en ambientes poco saludables, penosas y monótonas)² propiciaron la progresiva sustitución de trabajadores humanos en las líneas de soldadura y de pintura. Según datos de la International Federation of Robotics (IFR), la venta de robots en 2018 fue de 384.000 unidades (World Robotics, 2019), situando el número total de robots industriales instalados en el mundo sobre los 2,5 millones de unidades (World Robotics, 2018). La Norma ISO 8373:2012 proporciona una definición precisa de robot industrial:

«manipulador controlado automáticamente, reprogramable, multipropósito, programable en tres o más ejes, que puede ser fijo en un lugar o móvil para uso en aplicaciones de automatización industrial».

Las tareas en que se emplean robots industriales incluyen soldadura por puntos y soldadura continua, pintura, corte, acabado (lijado, pulido, encerado, etc.), ensamblado/desensamblado, paletizado, alimentación de máquinas (esto es, transferencia de piezas a y de máquinas), transporte o logística. Las operaciones de desensamblado, en particular, son de gran interés para la incipiente industria del reciclaje y recuperación de componentes o metales valiosos. Otro reto interesante, con grandes dificultades pero también grandes beneficios potenciales, es la manipulación de objetos deformables como comida y ropa.

- **Robots en minería.** Así como la tecnología robótica está más que probada y operativa en la industria, en el sector primario los robots aún se encuentran en fase experimental. No obstante, las ventajas de una futura robotización son evidentes, sobre todo en el sector de la minería, donde puede salvar vidas. Se estima que unos 12.000 mineros mueren en accidente cada año en el mundo, sobre todo en minas de carbón (Chakravorty, 2019). Existen varios proyectos de investigación sobre la exploración de minas inundadas, combinación de redes densas de sensores y robots para la exploración y monitorización de minas, o colaboración en equipos formados por humanos y robots en minería en la Colorado School of Mines y en el Council for Scientific and Industrial Research de Sudáfrica. El uso de robots, además de posibilitar la explotación de minas en presencia de gases tóxicos o con riesgo de derrumbamiento, también permitirá reconsiderar la explotación de minas en búsqueda de metales raros cuya demanda, antes inexistente, está creciendo en los últimos años. También la minería de superficie puede beneficiarse de robots de taladrado (menos invasivos que el uso de explosivos) y el transporte de mineral en camiones autónomos (Gendron, 2019).

⁽²⁾Es lo que se conoce en inglés como la triple «D»: *dangerous, dirty and dull*.

Proyectos de investigación

Proyecto Europeo UNEXMIN, <https://www.unexmin.eu/>

Proyecto ARIDuA, Autonomous Robots and the Internet of Things in Underground Mining, <https://tu-freiberg.de/en/aridua/summary>

- **Robots en agricultura, ganadería y pesca.** En agricultura, las aplicaciones (algunas de ellas ya en fase de explotación) incluyen el uso de drones e imágenes aéreas para evaluar el estado de los cultivos, la aplicación localizada de herbicidas y pesticidas con drones o con robots terrestres, la eliminación mecánica de mala hierba con robots (prescindiendo de herbicidas), la navegación autónoma de tractores, la recolección robotizada de fruta, la cosecha en plantaciones de invernadero, o la manipulación de plantas de invernadero (Postscapes, 2019). En ganadería podemos mencionar un robot teleguiado para la conducción de ganado en recintos vallados (Cargill, 2018), los sistemas robotizados de ordeñado (Michigan State University, 2019), y, de forma más experimental, sistemas multirrobot para el pastoreo de rebaños de ovejas (Pierson, 2015). Y en pesca, a modo de curiosidad, se puede hacer referencia a los llamados cebos robóticos, que tienen el aspecto e imitan los movimientos aleatorios de pececillos, para actuar como cebos de peces mayores, aunque no son realmente robots ya que no perciben su entorno ni actúan en consecuencia.
- **Robots de exploración.** Son robots móviles equipados para explorar y tomar muestras o realizar alguna otra tarea en entornos hostiles para el ser humano, como las profundidades submarinas, los volcanes, las zonas afectadas por una catástrofe, o el espacio exterior (incluyendo otros astros). Generalmente llevan una o varias cámaras, que permiten al operador humano visualizar el entorno inmediato del robot. También pueden llevar instrumental especializado para la tarea específica que han de llevar a cabo. Por ejemplo, los robots de búsqueda y rescate, cuya misión es localizar personas con vida entre los escombros o en condiciones de mala visibilidad, están equipados con sensores infrarrojos pasivos, que detectan fuentes de calor. También han de estar equipados para resistir durante un tiempo razonable en las adversas condiciones ambientales en las que desarrollan su actividad, sean presiones y/o temperaturas extremas, radiación, microimpactos, etc. Los robots de exploración submarina (Smithsonian Ocean Team, 2009), volcánica (Muscato *et al.* 2012), o de zonas afectadas por radiaciones (González, 2015) suelen ser unidades teleguiadas, ya que la distancia es relativamente corta y el operador del robot puede dar respuesta inmediata a las situaciones captadas por las cámaras de a bordo. Las dificultades técnicas radican, sobre todo, en la protección de las partes electrónicas más sensibles a las condiciones extremas en que operan. Sin embargo, las auténticas estrellas de la exploración son los robots espaciales, y en particular los *rovers* de exploración planetaria enviados a Marte. Ya en 1970 y 1973 los soviéticos habían enviado *rovers* de exploración a la Luna, en el programa Lunokhod. Estas unidades operaban remotamente, lo cual no era factible para los *rovers* marcianos debido al dilatado retardo de señal. El primero fue el Mars Pathfinder (1997), al que siguieron el Spirit y el Opportunity (llegaron a Marte en 2004) y el Curiosity (2012) (Mars Exploration Program, 2019). Estos robots poseen cierta autonomía para determinar la dirección a seguir o evitar un obstáculo, ya que si dependieran enteramente de las instrucciones enviadas desde la Tierra, avanzarían

Más sobre los cebos robóticos:

<https://roboticlure.com/>

muy poco a poco. Todos ellos llevan cámaras para enviar vídeos y fotografías a la Tierra, así como equipo científico para tomar y analizar muestras (cuyos resultados también son enviados a la Tierra). Los robots desactivadores de minas también se podrían considerar robots de exploración, pero se han incluido en el apartado de robots militares por el contexto bélico o posbélico en que se sitúan (véase más abajo «Robots militares» y en el subapartado 2.2.3).

- **Robots de construcción y mantenimiento.** Los robots de construcción con un futuro inmediato más prometedor son las impresoras 3D de edificios, robots que colocan ladrillos, y robots de demolición (Robotics Online, 2018). Por lo que se refiere al mantenimiento, ya existen robots de limpieza de fachadas (generalmente grandes superficies de vidrio) o de inspección y limpieza de tuberías, entre otros.
- **Robots de transporte de personas o mercancías.** Esta categoría incluye los vehículos autónomos terrestres y los vehículos aéreos no tripulados, de los que se hablará extensamente en el subapartado 2.2.2. Muchos tienen la apariencia de vehículos convencionales (coches, camiones, autocares), pero van equipados con un complejo sistema sensorial.
- **Robots militares.** En el subapartado 2.2.3 se describe brevemente la tipología, de acuerdo con el tipo de misiones encomendadas, y de forma más extensa los vehículos aéreos no tripulados (UAV). Desde un punto de vista constructivo, además de los UAV hay también sistemas con armas que se reorientan autónomamente sobre una base fija a gran velocidad para repeler ataques, así como armas o manipuladores (para desactivar explosivos) montados sobre *rovers*, generalmente teleguiados.
- **Robots de servicios.** La Norma ISO 8373:2012 define el robot de servicio como un «robot que realiza tareas útiles para los humanos o el equipamiento, excluyendo aplicaciones de automatización industrial». Según esta definición, vendría a ser un cajón de sastre en el que tienen cabida todo tipo de robots excepto los específicamente industriales. No obstante, en la práctica robótica suelen considerarse los que operan en un exterior no urbano como «robots de campo» (*field robots*), mientras que a los robots de servicios se les atribuye una mayor componente social, de interacción con los humanos. La frontera respecto a las otras categorías es difusa, por ejemplo, robots de limpieza en edificios o en las calles, que no dejan de ser robots de mantenimiento, podrían considerarse perfectamente robots de servicios. Entrarían aquí también los robots empleados en hostelería (camareros, cocineros, servicio de habitaciones) y en centros educativos, hospitales, etc. desempeñando tareas similares. Robots informativos y guía para visitantes y turistas, en entornos urbanos, museos, congresos, ferias, parques, etc. también se consideran robots de servicios. Estos robots, dependiendo de las tareas encomendadas, han de estar muy bien equipados desde el punto de vista sensorial y perceptivo, y también muy desarrolla-

dos en aspectos de interacción con los seres humanos (reconocimiento facial y de emociones, interpretación del lenguaje natural o al menos de ciertas frases y gestos, síntesis de voz, *displays* gráficos informativos, entre otros). Para una revisión de retos y logros alcanzados recientemente para este tipo de robots, véase Torras (2016). Precisamente, los robots que ponen énfasis en la interacción con los seres humanos constituyen la categoría de **robots sociales**, en los que también entrarían los del siguiente apartado. Véase Torras (2018b) para una revisión de las implicaciones éticas asociadas a este tipo de robots y a las redes sociales, así como sobre el papel especulativo (y educativo) de la ciencia-ficción sobre las ventajas e inconvenientes de posibles escenarios relacionados con estas tecnologías.

- **Robots domésticos y personales.** De características muy similares a las expuestas en el apartado anterior, en esta categoría, sin embargo, se pone énfasis en la adaptación al usuario (o grupo reducido de usuarios, como una familia, por ejemplo). Estos robots, además de venir de fábrica con una serie de habilidades manuales y sociales, han de tener una gran capacidad de aprendizaje. Entrarían en esta categoría todos los robots que se mueven en el ámbito doméstico, lo que incluye desde sencillos robots de limpieza hasta robots más sofisticados que actuarían a guisa de mayordomo o de secretario personal, robots de entretenimiento, y hasta compañeros sexuales o sentimentales. Los riesgos que suponen estos robots para la privacidad del individuo se analizan en el subapartado 2.2.5, y para su equilibrio emocional en el subapartado 2.2.6.
- **Robots educativos y de entretenimiento, «niñeras» robóticas.** Pensados para educar de forma práctica en robótica o en otras materias, o para actuar como juguetes sofisticados, estos robots pueden incluir la posibilidad de ser ensamblados o reconfigurables por sus usuarios. Pueden utilizarse tanto en el ámbito doméstico como en centros escolares, donde actúan como elemento de apoyo al profesorado. Tienen la ventaja de poder dedicar una atención exclusiva e ininterrumpida a cada niño, una paciencia y una constancia ilimitada (cualidades muy apreciadas en este contexto), al mismo tiempo que una firmeza inmune al chantaje emocional. Resultan muy atractivos para los pequeños, lo cual facilita que estos se involucren más en el aprendizaje. También contribuye a que los niños puedan confiar en que no serán criticados por el robot si se equivocan. Pueden constituir una herramienta inclusiva para aquellos niños con dificultades de socialización, si se usan correctamente, aunque también pueden producir justamente el efecto contrario. Entre los efectos negativos potenciales están una disminución de la tolerancia a las discrepancias de los demás, una dependencia excesiva, la atribución de deseos, emociones, objetivos propios y afecto a la máquina (y la consiguiente decepción al darse cuenta que no es así), y la exposición del niño a la posible actividad de un *hacker*. Algunos robots de entretenimiento también pueden fomentar la pasividad y la falta de imaginación. El colectivo infantil es muy vulnerable y maleable, por

lo que el diseño y la programación de tales robots deberá poner especial énfasis en la seguridad y en la pedagogía.

- **Robots médicos, de rehabilitación y asistenciales.** Los robots empleados en medicina realizan operaciones de cirugía, asisten en labores de enfermería con tareas rutinarias como distribución de medicamentos o bien prestando asistencia física en el movimiento de pacientes, y pueden ser empleados también en terapias de rehabilitación. Además de prestar sus servicios en hospitales, también pueden cumplir un papel en la asistencia domiciliar de pacientes impedidos o de avanzada edad, constituyendo en tal caso una categoría especial de los robots personales y domésticos (de nuevo, véanse los subapartados 2.2.5 y 2.2.6 para algunas consideraciones éticas de seguridad y dependencia emocional). La mayoría de robots cirujanos son en realidad telerrobots, que reproducen a una escala mucho menor y evitando posibles temblores los movimientos realizados por el cirujano humano desde su interfase. El cirujano puede seguir el progreso de la operación gracias a las imágenes mostradas por un monitor y capturadas por microcámaras introducidas en el paciente u otros sistemas de captura de imágenes. Ello permite aumentar el campo visual y facilita una experiencia inmersiva del cirujano en el entorno anatómico de la operación, miniaturizando, como se ha dicho, sus movimientos, y alcanzando por tanto una mayor precisión. Además, el cirujano puede predefinir zonas prohibidas, tejidos como por ejemplo nervios o algunos vasos sanguíneos que no se deben tocar, y el sistema de control impide en tiempo real el acceso a las mismas. El uso de robots suele estar asociado a cirugía mínimamente invasiva, y el ejemplo más conocido es el robot Da Vinci (Da Vinci, 2019), que ya se usa en numerosos hospitales. Los robots que trabajan de forma totalmente autónoma, como el robot dentista que realizó dos implantes en 2017 (Yan, 2017), aún están en fase experimental, como también lo están los micro y nanorobots, que se introducirían en el cuerpo del paciente para actuar *in situ* (destruyendo un coágulo o un tumor, por ejemplo). Los robots auxiliares de enfermería son versiones hospitalarias de robots de servicios más comunes, excepto aquellos que han de entrar en contacto físico con el paciente para moverlo. Aun actuando bajo supervisión del profesional sanitario, estos robots han de cumplir estrictos requisitos de seguridad, presentar una elevada estabilidad, un diseño ergonómico para una interacción física más adecuada y precisas capacidades sensoriales. Su introducción en los hospitales mejorará en gran medida las condiciones laborales del personal de enfermería, expuesto al riesgo de lesiones al desplazar pacientes voluminosos. En cuanto a los robots de asistencia terapéutica y rehabilitación, se trata de sistemas capaces de ajustar las terapias de acuerdo con las características y el progreso del paciente. Los ejercicios, en sus variantes de refuerzo del aparato locomotor, coordinación, destreza, memoria, etc. son propuestos y supervisados por el terapeuta humano, pero es el robot el que, en función de la respuesta del paciente, puede aumentar o disminuir el grado de dificultad de los mismos. Además de descargar al terapeuta del trabajo más rutinario, el paciente

puede sentirse más confiado al no sentirse sometido a eventuales críticas por el sistema.

- **Exoesqueletos y prótesis robóticas.** De las prótesis robotizadas y de sus implicaciones éticas se hablará extensamente en el subapartado 2.2.4. En cuanto a los exoesqueletos, se trata de estructuras que se ajustan al cuerpo del usuario, dotadas de sensores que pueden captar cierta actividad muscular o neuronal del paciente, y a través de sus actuadores dar una respuesta en forma de movimiento. Pueden amplificar la fuerza física de un usuario sano (de hecho, se está investigando su posible uso en aplicaciones militares), pero sobre todo pueden proporcionar movilidad a un usuario impedido.

1.4. Técnicas de IA en robótica

La mayoría de robots existentes hoy en día actúan de acuerdo con una secuencia de instrucciones codificada por un programador humano. Como ya se ha dicho, esto es suficiente para robots que desempeñan su actividad en un entorno controlado, como es el caso de los robots industriales, o que realizan tareas simples como un robot aspirador. Cuando el robot se enfrenta a un mundo complejo, dinámico, parcialmente observable y lleno de incertidumbres, y ha de ser capaz de perseguir los objetivos marcados a pesar de obstáculos y de contingencias, y ha de afrontar las ambigüedades y matices de las relaciones humanas, no cabe duda de que solo puede hacerlo con éxito si está dotado de las herramientas cognitivas de la IA. En lo que sigue, vamos a explicar sucintamente en qué consisten dichas herramientas; para más detalles, véase Jiménez (2018):

- **Percepción.** Es la adquisición, procesamiento e interpretación, en términos inteligibles para el sistema de representación del conocimiento y control del robot, de las señales visuales, de fuerzas y momentos, acústicas, táctiles, etc. que proporcionan los sensores. Es la premisa para los procesos de aprendizaje y de control de un robot, lo que le permite estar situado en el mundo y dar una respuesta coherente con lo que interpreta que pasa alrededor. El canal sensorial más informativo aunque también más complejo es la visión por computador. Las imágenes adquiridas por la cámara son segmentadas en diferentes regiones significativas, y sobre esta partición tienen lugar los diferentes procesos de detección (por ejemplo, de ojos y boca en una cara), de identificación (por ejemplo, asociando nombres a diferentes caras), o de reconocimiento (interpretando una escena como «mujer montada en bicicleta con un perro corriendo a su lado»). La integración de la información sensorial de diferentes canales (visión y tacto, por ejemplo) permite resolver ambigüedades de los canales por separado, y se denomina **fusión sensorial**.

- **Aprendizaje.** Consiste en la adquisición o actualización de un modelo del mundo de primera mano, esto es, procesando la información que suministra el sistema sensorial. Las técnicas de base son métodos matemáticos de correlación, probabilísticos, estadísticos o de reconocimiento de patrones. Las técnicas de aprendizaje se pueden aplicar a la interpretación de imágenes y a la interpretación del lenguaje natural, asignando estadísticamente un significado a cada elemento de la imagen o del texto. Un método con gran éxito son las ya mencionadas redes neuronales artificiales (subapartado 1.1) en una versión actualizada que incluye muchas capas de neuronas (las llamadas *deep neural networks* que posibilitan el famoso *deep learning*), aunque tienen el gran inconveniente de la falta de trazabilidad (es un proceso de tipo «caja negra», en que no se sabe lo que ocurre en su interior y por qué aprende lo que aprende). El aprendizaje puede ser **supervisado** (por un humano, por ejemplo), mediante técnicas de clasificación (el supervisor indica la clase a la que pertenece cada muestra de aprendizaje o entrada) o de regresión (ajuste automático de una función matemática a los pares entrada-salida, dada esta por el supervisor), o **no supervisado**, aplicando técnicas de *clustering* (agrupación por similitud) o aprendizaje de reglas asociativas. Los dos grandes paradigmas de aprendizaje en robótica son el **aprendizaje por refuerzo** (*reinforcement learning*) donde el robot aprende qué acciones son más convenientes en cada estado del entorno mediante observación y una función de recompensa (o penalización), y el **aprendizaje por demostración**, donde un «profesor» humano enseña al robot la mejor manera de realizar una acción o una tarea. El aprendizaje por refuerzo tiene la ventaja de que permite al robot aprender autónomamente, aunque puede alargarse mucho ya que ha de ir probando las diferentes acciones en entornos cambiantes. Esto no ocurre con el aprendizaje por demostración, que sin embargo tiene el inconveniente de requerir repetidas demostraciones del profesor (para adquirir cierta generalidad ante ligeras variaciones en el entorno y aprender solo lo realmente significativo), lo cual puede ser engorroso o tedioso para este. Ambas técnicas se pueden combinar: con pocas demostraciones el profesor puede dejar al robot aprendiendo por refuerzo por su cuenta, mejorando las acciones recién aprendidas.
- **Planificación.** Es determinar la secuencia de movimientos o acciones necesaria para alcanzar un objetivo. Cabe distinguir la planificación de movimientos de la de tareas. La **planificación de movimientos** es puramente geométrica (aunque también puede incluir velocidades y aceleraciones), y consiste en determinar la secuencia de poses (posición y orientación) de un robot para llegar de un punto de inicio a una meta evitando los obstáculos (fijos o móviles) que pueda haber en su entorno, a partir de una representación geométrica explícita o implícita del mismo. La **planificación de tareas** opera a un nivel simbólico, con un repertorio de acciones básicas que puede ejecutar el robot. Dichas acciones solo pueden ejecutarse si se dan ciertas precondiciones y, una vez ejecutadas, producen ciertos efectos en el entorno, modificándolo de modo que ahora se dan

Planificación de tareas

Propiamente con representaciones simbólicas de tales acciones, por ejemplo, en forma de reglas «si se dan estas *precondiciones*, **entonces** la acción puede ejecutarse y produce estos *efectos*».

las precondiciones de otras acciones diferentes. Ello permite planificar, esto es, encadenar las acciones básicas de manera que se vaya modificando el entorno hasta alcanzar el estado deseado. En un inicio los métodos de planificación eran deterministas, esto es, partían de la base de que toda la información relevante estaba disponible, y que los efectos de las acciones sobre el entorno eran fijos y determinados. Hoy en día, existen planificadores capaces de proporcionar un buen plan en presencia de modelos incompletos del mundo e incertidumbre en los efectos de las acciones.

- **Representación del conocimiento y razonamiento.** Junto con representaciones elementales como conjuntos de datos, descripciones geométricas, grafos (y las correspondientes técnicas de búsqueda en grafos), etc., el conocimiento puede ser representado a través de la **lógica**, donde las afirmaciones tienen un valor binario (verdadero o falso) y que permite razonar por inferencia, sea inductiva, deductiva o abductiva. Una forma muy efectiva de utilizar la lógica es mediante la programación de satisfacción de restricciones (*constraint logic programming*). Por otro lado, la lógica de descripciones (*description logics*) parte de una estructura terminológica jerarquizada y utiliza la consistencia, subsunción (en su acepción más general, no jurídica), o disjunción como mecanismos de razonamiento (entre otros). Está en la base del OWL (*ontology web language*) para la representación del conocimiento en internet (para ser usada por aplicaciones más que por humanos), y también de RoboEarth, un repositorio en red para robots (véase el subapartado 2.2.5). Otra forma de representación es la **probabilística**, más acorde con las incertidumbres del mundo real, donde las afirmaciones son ciertas o falsas con cierta probabilidad, y el razonamiento se funda en el teorema de Bayes para determinar la probabilidad de una hipótesis dada la evidencia observada (se utilizan formalismos como las redes bayesianas). Por último, en la **lógica difusa** (*fuzzy logic*) las afirmaciones tienen un determinado grado de certeza: diferentes afirmaciones pueden ser ciertas simultáneamente, pero en cada situación unas lo son más que otras.

Un reto que todavía se está lejos de alcanzar es la recreación artificial de conciencia y emociones. La primera tiene que ver con la conciencia de uno mismo (*selfawareness*) que muestran muchos animales, sobre todo mamíferos superiores, con las experiencias subjetivas (*qualia*), que algunos filósofos y científicos consideran el principal obstáculo para conseguir una conciencia genuina, con la voluntad y los objetivos propios, etc. Cabe preguntarse si, al margen de proporcionar nueva luz sobre qué es la conciencia humana, el objetivo de alcanzar una conciencia artificial es deseable en sí mismo, y si es conveniente o no dotar a las máquinas de conciencia. En una línea similar están las emociones. Por descontento que puede ser muy útil que a través del aprendizaje el robot pueda llegar a identificar emociones y el estado de ánimo de las personas con las que ha de interactuar, de modo que pueda modular esta interacción de la forma más conveniente (por ejemplo, no haciéndose pesado). Incluso puede facilitar la interacción que el robot pueda simular ciertas emociones

(por ejemplo, mostrarse cansado si está bajo de baterías). Las posibilidades y la utilidad, empero, de que algún día una máquina realmente sienta emociones son más que cuestionables.

2. Consideraciones éticas

Los avances en IA y robótica, y los productos que de este desarrollo se derivan, tienen un impacto potencial de gran calado en la sociedad humana. Otras tecnologías, como por ejemplo internet o la telefonía móvil, ya han dado pruebas de su poder transformador, en la forma que tenemos de comunicarnos, comprar, viajar, consumir productos culturales o relacionarnos. Incluso si se cumplen únicamente las expectativas de los más escépticos (o los menos agoreros), no cabe duda de que los robots afectarán de diversas maneras a nuestra forma de vida conforme vayan ganando en prestaciones, destreza, autonomía y habilidades cognitivas. La propia comunidad científica ya dio los primeros pasos en cuanto a considerar las implicaciones éticas de esta tecnología. El comienzo oficial de la **roboética** o ética del desarrollo de los robots para el progreso humano y social, término acuñado por el científico y experto en robótica Gianmarco Veruggio en 2002, se sitúa en el año 2004, con el Primer Congreso de Roboética en San Remo (Italia). Ese mismo año se creó el Comité Técnico en Roboética del IEEE-RAS, y tuvo lugar la Declaración Mundial del Robot en Fukuoka (Japón), sobre las expectativas de los robots de próxima generación (Fukuoka, 2004).

En esta parte repasaremos algunos de los ámbitos de la actividad humana en los que la IA y la robótica pueden tener una incidencia más acusada, así como las cuestiones de índole ética y moral que cabe plantearse acerca de los posibles riesgos que se derivan de su implantación. A la corta, estos riesgos no difieren sustancialmente de los derivados del mal funcionamiento de cualquier otra máquina, incluyendo fallos mecánicos y errores humanos. A la larga, y esto es pura especulación, los riesgos serían los asociados a la existencia de una nueva especie inteligente en la Tierra, con la salvaguarda de una eventual fusión entre la humanidad biológica y sus criaturas cibernéticas. Y entremedio, los riesgos provienen de la presencia ubicua de máquinas que permearán cada vez más sectores y harán cada vez más cosas; muchas de ellas las realizarán mejor que sus creadores humanos, y por tanto irán ganando más atribuciones y responsabilidades. En lo que sigue utilizaremos frecuentemente ética y moral como términos intercambiables, aunque el lector debe ser consciente de las diferencias entre ambos: mientras que la moral se refiere al conjunto de costumbres impuestas por el entorno, a los usos de una sociedad determinada en el tiempo y en el lugar, la ética es fruto de una reflexión individual, racional, e influye en la conducta personal de forma voluntaria y consciente. La moral tiene un carácter más práctico dando respuesta a situaciones concretas, vividas en el día a día, mientras que la ética, desde un punto de vista teórico, intenta dar respuesta a por qué deberíamos (o no) actuar del modo que propone la moral (Román, 2016).

Primer Congreso de Roboética

http://www.roboethics.org/sanremo2004/ROBOETHICS_Program.html

La IEEE...

es la asociación internacional de ingenieros eléctricos y electrónicos más importante del mundo, de la cual forma parte RAS (Robotics and Automation Society). <http://www.ieee-ras.org/robot-ethics>

2.1. ¿Es posible construir máquinas éticas?

La pregunta que encabeza este subapartado (¿es posible construir máquinas éticas?) es lo suficientemente inconcreta como para admitir varias interpretaciones y, por tanto, varias respuestas. En primer lugar, a raíz de lo visto hasta el momento, sabemos que los robots, en tanto que pueden ocasionar a través de sus acciones diferentes efectos que afectan tanto a particulares como a sectores y al conjunto de la sociedad, en forma de daños físicos, económicos, psicológicos o morales, son máquinas éticamente relevantes. En segundo lugar, dado el estado actual de la IA, los robots no están en condiciones de responder por sus acciones. El robot que causó el accidente mortal de la planta de Akashi y fue retirado no *sufrió* realmente un castigo y podría haber seguido funcionando igual que las otras unidades: la decisión de retirarlo respondió a la compulsión humana del resarcimiento, del mismo modo que un niño «castiga» propinándole un golpe al mueble contra el que acaba de darse un coscorrón. Por tanto, a día de hoy lo que tenemos son agentes (máquinas) *implícitamente* éticos, en terminología de James Moor (2006), o en palabras de Wallach y Allen (2009), máquinas dotadas de **moralidad operacional**: máquinas cuya

«significación moral está enteramente en manos de diseñadores y usuarios» (Allen y Wallach, 2014).

Por ejemplo, a día de hoy, los vehículos autónomos no van más allá de esta modalidad, según Loh y Loh (2017). A pesar de constituir (con el conductor) lo que estos autores denominan un sistema híbrido por lo que se refiere a la toma de decisiones durante la conducción, la responsabilidad en situaciones de dilema moral recae en el conductor. En contraposición, habría aquellas máquinas que son *agentes éticos explícitos* cuyas acciones son fruto de un razonamiento explícito y deliberado sobre los efectos éticos de las mismas. Para Wallach y Allen, la máxima expresión del agente ético explícito se situaría en lo que ellos denominan **agencia moral plena** (*full moral agency*), característica hoy por hoy exclusiva del ser humano. No es descartable que algún día las máquinas puedan adquirir algo semejante a la conciencia, a la voluntad y a los propios objetivos (aunque cabe preguntarse si es necesario o deseable...); posiblemente este desarrollo venga de la mano de la IA general o fuerte.

Mientras tanto, Wallach y Allen postulan la posibilidad de que los robots puedan tener lo que ellos llaman **moralidad funcional**, a través de la cual y fruto de una actividad deliberativa (explícita), «las propias máquinas tengan la capacidad para evaluar y responder a cuestiones morales». Es decir, en el momento de decidir la siguiente acción a ejecutar (o de planificar una secuencia de acciones), una máquina moralmente funcional ha de ser capaz de evaluar el valor ético de los efectos de esta acción y determinar si, a la luz de dicha evaluación, estos efectos (y por ende la propia acción) son aceptables o no. La moralidad funcional se situaría entre los extremos de la moralidad operacional y la agencia moral plena. Para los autores (y de hecho hay un amplio consenso en torno a esta cuestión), esta capacidad no es solo deseable sino que es nece-

saría conforme los robots ganen en competencias y atribuciones. La pregunta siguiente es: ¿cómo se implementa esta moralidad funcional? O en términos de Wallach y Allen: ¿cómo se construye un *agente moral artificial* (AMA)?

Los filósofos llevan siglos debatiendo sobre ética. Han desarrollado abstracciones, teorías y sistemas con pretensiones de universalidad, que trascienden hechos y situaciones concretas, pero que por este mismo motivo deberían poder aplicarse en cualquier circunstancia y contexto particular. Los ingenieros e investigadores en IA y robótica, por otro lado, están interesados precisamente en el aspecto práctico del diseño e implementación de un sistema que funcione en circunstancias y contextos particulares. ¿Pueden aprovechar los logros de la filosofía y evitar tener que reinventar la rueda? Para responder a esta pregunta hemos de repasar las principales teorías de la ética y examinar sus posibilidades de implementación en una máquina:

1) La **deontología** se refiere al estudio de los fundamentos del deber y las normas morales. El postulado básico de la deontología es que las acciones tienen un valor en sí mismas, y que hay deberes que se deben cumplir independientemente de sus consecuencias. En la práctica, se traduce en el establecimiento de conjuntos ordenados de reglas que rigen el comportamiento moral de un colectivo, por ejemplo, de quienes ejercen una determinada profesión (deontología profesional). Probablemente, el representante más destacado de esta teoría ética sea Immanuel Kant (1724-1804) y su imperativo categórico. La ambigüedad inherente al lenguaje humano posibilita cierto margen de interpretación y el debate sobre si esta debe ceñirse al texto o al espíritu de las reglas. La deontología evoca inmediatamente las famosas tres leyes de la robótica del escritor de ciencia-ficción Isaac Asimov (1920-1992), de las que hablaremos más adelante.

2) El **consecuencialismo** adopta la perspectiva opuesta, en tanto que no valora las acciones en sí mismas ni en la intención que las anima, sino por sus consecuencias. Una acción será buena, por tanto, si lo son las consecuencias de la misma. Una de las variantes más destacadas del consecuencialismo, el **utilitarismo**, formalmente desarrollado por Jeremy Bentham (1748-1832) –aunque el término fue acuñado por su discípulo John Stuart Mill (1806-1873)–, suele resumirse en el principio de utilidad como conseguir «el máximo beneficio para el máximo número de personas». Suele contraponerse este *utilitarismo del acto* al *utilitarismo de las normas*, definiendo este último la mejor acción como la que sigue una norma que aporta la máxima utilidad. En cierto modo, el utilitarismo de las normas admite una conciliación del consecuencialismo con la deontología. Otras variantes del consecuencialismo incluyen el egoísmo ético, el altruismo ético o el consecuencialismo (o utilitarismo) negativo (que propugna concentrarse en reducir el sufrimiento más que en aumentar el placer).

3) La **ética de las virtudes** está relacionada con el carácter de una persona: una acción es éticamente correcta si es la que realizaría una persona virtuosa en aquella situación. Tiene su origen en Platón y Aristóteles, y los conceptos que

aún hoy en día manejan las versiones actualizadas de esta corriente provienen de la filosofía griega, e incluyen la *areté* (excelencia o virtud), la *phronesis* (sabiduría práctica o moral, que permite al individuo saber hacer lo correcto en cada situación) y la *eudaimonia* (felicidad, un estado que se alcanza precisamente por la práctica de las virtudes). Véase Wikipedia, «Virtue ethics» (2019), para una lista de autores modernos que suscriben la ética de las virtudes, así como de las virtudes aristotélicas y los vicios que resultan de su exceso o de su defecto.

¿Es posible implementar alguna de estas teorías en un robot de manera que este sea capaz de pasar sus acciones por el tamiz de la ética antes de ejecutarlas? La ciencia-ficción ya propuso una manera de hacerlo en los años cuarenta del pasado siglo, a través de las famosas tres leyes de la robótica, que aparecen por primera vez en el relato «Círculo vicioso» («Runaround», en el número de marzo de 1942 de la revista *Astounding Science Fiction*), del escritor Isaac Asimov, y posteriormente, entre otros, en la colección de relatos *Yo, robot* (1950):

- «1. Un robot no hará daño a un ser humano o, por inacción, permitirá que un ser humano sufra daño.
2. Un robot debe cumplir las órdenes dadas por los seres humanos, a excepción de aquellas que entren en conflicto con la primera ley.
3. Un robot debe proteger su propia existencia en la medida en que esta protección no entre en conflicto con la primera o con la segunda ley».

Posteriormente, se añadiría una «ley cero» a las que las anteriores van supeditadas, según la cual:

«Un robot no hará daño a la Humanidad o, por inacción, permitir que la Humanidad sufra daño».

En aquella época eran muy populares las novelas de género en las que, invariablemente, los robots se rebelaban y constituían una amenaza a la altura de la heroicidad del protagonista. Asimov consideraba absurdo que el hipotético desarrollo de robots inteligentes no conllevara una salvaguarda para sus creadores humanos. Las tres (o cuatro) leyes formaban parte de la propia estructura de los cerebros *positrónicos* de los robots, de modo que no podían ser *hackeadas*. Evidentemente, Asimov no proporciona detalles sobre la construcción de este tipo de cerebros artificiales, ni pretendía que estas leyes se implementaran de algún modo en máquinas reales: no eran más que un recurso literario que le permitía desarrollar las tramas asociadas a situaciones en que entraban en conflicto. No obstante, se les puede reconocer el mérito de constituir un primer código deontológico para los robots, aunque sea en la ficción. Diversas asociaciones profesionales, universidades y otras instituciones han desarrollado sus propias listas de principios éticos de la robótica y la IA, un compendio de las cuales puede encontrarse en Winfield (2018).

Las teorías normativas de la ética, sean deontológicas o consecuencialistas, proceden de forma descendente (*top-down*): un principio genérico, sea la adhesión a un deber expresado en una regla o la maximización de un beneficio, ha de ser aplicado a una situación concreta. En robótica y en IA es frecuente programar instrucciones en forma de reglas que han de ser instanciadas para poder ejecutarse (esto es, la información que procede de los sensores indica si se dan las condiciones en el entorno para ejecutar la acción expresada en la regla) o funciones cuyas variables también corresponden a valores determinados por los sensores. No obstante, aplicar estos procedimientos a la evaluación ética de una acción en un entorno poblado de humanos (más allá de un entorno restringido y controlado como el industrial) no es tan inmediato. En el caso de las formulaciones deontológicas, el primer problema con el que nos encontramos es el de la correcta interpretación de la regla en cuestión. Una interpretación literal de la primera ley de Asimov impulsaría al robot a detener la mano del cirujano que está a punto de practicar una incisión. La casuística de excepciones a una regla genérica puede ser abrumadora. También lo es la de posibles realizaciones o concreciones de una regla general: siguiendo con el ejemplo de la primera ley, el «daño» infligido a un ser humano puede ir desde crearle una leve contrariedad hasta causarle la muerte. Otra fuente de problemas es la de reglas que entran en conflicto. Establecer prioridades entre las mismas no acaba de ser una solución: una regla de prioridad más elevada puede tener una probabilidad mínima de aplicabilidad y, por tanto, no ser relevante en una situación concreta.

La ética basada en el consecuencialismo topa con similares dificultades. Wallach y Allen citan las cuatro habilidades computacionales requeridas por un robot consecuencialista, según estableció el informático James Gips en 1995:

- 1) Una forma de describir el estado del mundo,
- 2) una forma de generar posibles acciones,
- 3) una forma de predecir la situación o estado resultante del mundo si se aplica determinada acción en el estado actual, y
- 4) un método para evaluar una situación en términos de bondad o deseabilidad.

Aunque estas habilidades ciertamente no definen un algoritmo ni un método a implementar, sí pueden contemplarse como un conjunto esencial de habilidades que requeriría un robot consecuencialista. Ahora bien, ya la primera tarea plantea la espinosa cuestión de determinar los elementos relevantes que definen una situación, el estado del mundo en aquel momento. Una acción puede afectar a múltiples sujetos, sean humanos, otros seres vivos, incluso el ecosistema u objetos inanimados. ¿Dónde detener la evaluación del estado de cada uno de los potenciales afectados? El rango de posibles acciones generadas (segunda tarea) también depende de la identidad y el número de posibles

Bibliografía

(Wallach y Allen, 2009, pág. 87)

afectados por las mismas. Por lo que se refiere a la tercera tarea, si hemos sido capaces de solventar la primera y tenemos alguna manera de manejar la incertidumbre en los efectos asociada a la ejecución de cualquier acción (a causa de las contingencias que se puedan producir), ciertamente puede esperarse que seamos capaces de predecir el resultado *inmediato* de la aplicación de determinada acción, pero una acción puede tener también consecuencias a medio y largo plazo, algunas improbables pero no imposibles. ¿Dónde detener el cálculo de las posibles ramificaciones, hasta qué punto en el futuro nos debemos adentrar? Y por último, ¿cómo calcular objetivamente la bondad, utilidad o deseabilidad de una situación resultante de una acción? ¿Pueden priorizarse los placeres resultantes? ¿Es mejor el regocijo banal de miles de telespectadores de un programa basura que el placer estético de unas docenas de entendidos en escuchar en directo una interpretación musical de calidad? ¿Se puede cuantificar el placer que experimenta cada uno?

En el fondo, las teorías que proceden de forma descendente desde principios de pretendida universalidad, sean estas formulaciones deontológicas del deber o funciones de beneficio o utilidad, se enfrentan al mismo problema fundamental cuando se plantea su eventual implementación en una máquina: delimitar, discernir los datos, hechos y circunstancias relevantes en el universo de las infinitas posibilidades del día a día, es decir, no en un área especializada y restringida, para tomar decisiones moralmente aceptables. No es más que otra versión del *frame problem* del sentido común y de la IA general. Es lo que otros autores denominan «isotropía» (Guarini y Bello, 2014): la relevancia de cualquier cosa en relación con cualquier otra cosa. No obstante, a pesar de ser computacionalmente intratables, los principios éticos formulados a alto nivel (esto es, por ejemplo como reglas), sí pueden constituir heurísticas que guíen el proceso de selección de la mejor acción a emprender, así como ayudar en el proceso de evaluación *a posteriori* de las acciones realizadas por el robot (Allen y Wallach, 2014, pág. 59).

Una alternativa a la programación explícita de instrucciones es el aprendizaje. Sea por demostración, o sea de forma autónoma por refuerzo, estos procedimientos permiten adquirir conocimientos y habilidades –en nuestro caso, de discernimiento moral– de forma ascendente (*bottom-up*), a través de ejemplos o de ensayos de prueba y error. Otros métodos que proceden de forma ascendente son aquellos que simulan procesos evolucionistas, como los algoritmos genéticos. Estos procedimientos, aplicados a los AMA, parten de la base de que la capacidad de discernimiento ético emergerá como uno de los aspectos más generales de la inteligencia. Al fin y al cabo, es por aprendizaje que los humanos adquirimos nuestras convicciones éticas (y es posible que la humanidad y las culturas que la integran hayan desarrollado sus códigos morales de una forma análogamente evolutiva). El problema principal con los procedimientos *bottom-up* es determinar el conjunto de ejemplos (o de funciones de recompensa/penalización, o los criterios de selección) que garantizan el aprendizaje o la evolución de comportamientos éticamente apropiados. Debe tenerse presente que los algoritmos de aprendizaje se limitan a procesar datos, y que

por consiguiente los resultados de este aprendizaje pueden estar afectados por los sesgos que involuntaria y, a menudo, inconscientemente, se introducen en la fase de recopilación de datos. Ejemplo de ello son los algoritmos expertos de concesión de créditos o los procedimientos estadísticos que atribuyen una mayor tasa de reincidencia delictiva a sectores desfavorecidos socialmente o pertenecientes a determinadas etnias: una interpretación acrítica de los resultados perpetúa los prejuicios del investigador (cuando no es consciente de que «quién es más probable que cometa un crimen» y «quién es más probable que resulte condenado» son dos cuestiones enteramente diferentes y que en los Estados Unidos están fuertemente condicionadas por la etnia del sujeto). En palabras de Denyse O’Leary (2019):

«¿Puede un algoritmo ser racista? El *output* no tiene opinión. Como siempre, es el *input* el que nos ha de preocupar».

En el caso de AMA que están desarrollando un criterio ético de evaluación por aprendizaje, los sesgos inadvertidos de quienes seleccionan y anotan los datos con que se alimenta el sistema de aprendizaje son los que conformarán el tipo de ética del AMA. Quizá sea preferible que los supervisores del aprendizaje sí introduzcan algún tipo de sesgo, de forma consciente, a fin de favorecer la emergencia de comportamientos éticos deseables.

«Todos los sistemas de aprendizaje requieren sesgos y asunciones para funcionar bien, así que no perdemos nada incluyendo sesgos que son éticamente justos» (Montañez, 2019).

La ética de las virtudes puede contemplarse como una conciliación de los procedimientos *top-down* y *bottom-up*: las virtudes pueden ser explícitamente descritas (y los resultados de una acción contrastados con los que resultarían de un comportamiento éticamente virtuoso), mientras que la adquisición de los rasgos morales correspondientes a un carácter virtuoso se harían por aprendizaje. El sistema cognitivo propuesto en Wallach y Allen (2009) para la implementación de criterios morales en agentes artificiales, el modelo LIDA de Stan Franklin, proporciona mecanismos de selección de acciones que proceden de forma ascendente a partir de colecciones de datos del sistema sensorial, y procesos descendentes que le dan sentido a la situación actual. Es discutible que este o futuros sistemas que combinen procedimientos normativos con aprendizaje basado en la experiencia permitan llegar a obtener agencia moral plena, pero es muy posible que permitan alcanzar elevadas cotas de moralidad funcional.

Véase también

Véase también O’Neil (2016).

2.2. Relevancia ética de los robots en diferentes ámbitos de la experiencia humana

Los seres humanos interpretamos diferentes papeles a lo largo de la vida, tanto por evolución vital como dependiendo del entorno en el que nos encontramos. A continuación revisaremos cómo puede incidir la presencia de un robot inteligente en estos diferentes ámbitos, algunos cotidianos y otros más excepcionales, y las cuestiones éticas que se plantean debido a esta presencia.

2.2.1. El ámbito laboral: impacto del robot sobre el mercado de trabajo

La irrupción de los robots en el mercado laboral es percibida como una amenaza por amplios sectores de la sociedad. A ello contribuyen estudios de proyección cuyas conclusiones son reinterpretadas y amplificadas de forma catastrófica por algunos medios de comunicación. Hablaremos de estos estudios más adelante; de momento, analicemos algunos precedentes en transformaciones significativas en el trabajo humano, en sus características y posibles analogías con el caso que nos ocupa. Posiblemente, la primera transformación radical del *modus vivendi* humano haya sido lo que se denomina la revolución neolítica, que empezó hacia el año 10000 a. C. en Levante (este del Mediterráneo), y 8000-7000 a. C. en las riberas del Nilo, del Creciente Fértil (Tigris y Éufrates), del Indo y del río Amarillo, aunque tuvo lugar durante un largo período de tiempo (milenios) en su expansión hacia otros lugares habitados por humanos. Las sociedades de cazadores-recolectores nómadas se transformaron en sociedades agrícolas y ganaderas sedentarias.

Desde el punto de vista del trabajo, lo más destacado fue la **especialización**, y en consecuencia, el comercio. La figura del artesano, una persona que ha adquirido una serie de habilidades, tras años de formación, para la confección de un tipo de producto determinado, se mantendrá prácticamente hasta la Revolución Industrial, y aún existe en nuestros días (aunque de forma residual). Los gremios de artesanos mantendrán durante siglos el poder suficiente para oponerse a la mecanización de trabajos que les son propios, como ilustra el ejemplo de una máquina de hacer punto que fue inventada en 1589 por William Lee, a la que fue denegada la patente por la Corona inglesa (Frey y Osborne, 2013, pág. 6). La emergencia de una clase social capitalista, con creciente influencia en el Parlamento británico, que a su vez ganó poder frente a la Corona, creó el terreno abonado para la Revolución Industrial que se produjo en la segunda mitad del siglo XVIII gracias a la mejora de la máquina de vapor realizada por James Watt, incluyendo el regulador de presión (esto es, introduciendo el control automático). La Revolución Industrial significó la deriva de la demanda de mano de obra hacia el trabajador poco cualificado (en comparación con el artesano), al dividir la manufactura del producto en múltiples labores simples y muy especializadas. Con el advenimiento de las máquinas automatizadas (gracias a los avances en la electrónica de control en el siglo XIX), son precisamente estos trabajos de baja cualificación y altamente

Reflexión

Hay evidencia de semillas plantadas en fechas incluso más antiguas, pero sin conllevar los cambios de hábitos asociados al sedentarismo.

repetitivos los que se vieron amenazados (y gradualmente eliminados). Existe una nueva demanda de trabajadores con cierta cualificación técnica, especialistas en electromecánica y disciplinas afines, capaces de manejar y de mantener maquinaria cada vez más compleja. Así pues, de nuevo asistimos a lo que no es más que un desplazamiento en la demanda de mano de obra, y una vez más circunscrito al sector secundario.

Sin embargo, con la flexibilidad y creciente sofisticación que introduce la informatización de las máquinas, y en particular de los robots, el espectro de trabajos en los que los humanos pueden verse desplazados por máquinas se va ampliando. Las revoluciones tecnológico-industriales anteriores han causado la eliminación de ciertos tipos de trabajo pero han creado otros tantos, y han ocurrido a un ritmo en que la transición, la adaptación de los trabajadores a los nuevos requisitos se ha podido realizar de forma gradual (incluso en lo que respecta al relevo intergeneracional). Que suceda otro tanto en el presente caso es cuestionable, tanto por la rapidez como por el alcance con que pueden tener lugar los cambios. Los ordenadores y los robots tienen el potencial de sustituir la fuerza de trabajo humana en los tres sectores, de una forma sin precedentes en la historia de la humanidad.

Analicemos en primer lugar el sector industrial, donde ya hace décadas que se han ido introduciendo robots. Los robots han ido sustituyendo a trabajadores humanos en tareas peligrosas (como alimentar prensas o cizallas), en ambientes nocivos (para los que se necesitan protecciones especiales, como por ejemplo en pintura, con la presencia de disolventes), en labores monótonas (tareas repetitivas como en una cadena de montaje), o en combinaciones de las anteriores. A pesar de sustituir a las personas que realizaban dichas labores hasta el momento, hay cierto consenso en que los robots han supuesto una mejora en general de las condiciones laborales. La introducción de robots ha supuesto la creación de nuevos puestos de trabajo, sea directamente en las mismas plantas, como personal para la programación y para el mantenimiento de los robots, sea indirectamente, como las personas empleadas en las fábricas de robots, personal de formación y entrenamiento, y personal en investigación y desarrollo. Según un estudio realizado en 2015,

«entre 2009 y 2014, las 62 compañías con mayor base instalada de robots crearon netamente 1,25 M nuevos puestos de trabajo, un 20 % más» (Robotonomics, 2015).

¿Se mantendrá esta tendencia, o llegará el momento en que otras máquinas sean capaces de programar robots para tareas concretas, aparezcan robots de mantenimiento, o la demanda de robots industriales y por tanto la industria de fabricación de robots llegue a cierta saturación? Lo cierto es que la industria apuesta por lo que puede reportar el máximo beneficio en el plazo menor, y suele ser bastante conservadora en lo que se refiere a la introducción de tecnologías que no han sido suficientemente probadas. Por este motivo, aunque la sustitución de trabajadores humanos por máquinas resulte atractiva a la empresa por la reducción de costes asociada, lo es aún más la introducción de

equipos mixtos humanos-robots capaces de sacar partido de las mejores cualidades de cada uno. Esto es posible gracias al desarrollo de los **cobots** (*cooperative robots*), robots que pueden trabajar en colaboración y contacto con el ser humano sin suponer un riesgo para su integridad física. De confirmarse esta tendencia, la destrucción de puestos de trabajo en la industria sería muy inferior a lo augurado por las predicciones más pesimistas. Aun así, un informe reciente de Oxford Economics (2019) cifra en veinte millones los puestos de trabajo que se perderán en la industria manufacturera de aquí a 2030. A guisa de comparación, Naciones Unidas cifra la pérdida de lugares de trabajo por el estrés térmico debido al cambio climático en ochenta millones de personas en 2030 (Noticias ONU, 2019), mayoritariamente en el sector agrícola y en países de Asia meridional y África occidental (aunque, en este caso, se trata más bien del equivalente a la suma de horas de trabajo perdidas por exceso de calor que propiamente puestos de trabajo que desaparecen).

Por lo que se refiere a otros sectores, la mecanización también ha supuesto un desplazamiento de un gran número de trabajadores en el pasado. Las cosechadoras automáticas, por ejemplo, supusieron una drástica reducción en la mano de obra requerida en tareas asociadas a la siega. Los cajeros automáticos tuvieron un impacto igualmente notable en la banca, y más recientemente las compras y reservas en línea han afectado al minorista y a las agencias de viajes, solo por nombrar algunos ejemplos. ¿Cuál será la incidencia de la computerización y robotización del sector primario y el de servicios? En un estudio de 2013 de la Oxford Martin School y del Departamento de Ciencias de la Ingeniería de la Universidad de Oxford, los cuellos de botella en la susceptibilidad de un trabajo a ser computerizado son las habilidades perceptivas y destreza manual, por un lado, la creatividad, por otro, y la inteligencia social como tercero (Frey y Osborne, 2013, pág. 24 y sig.).

Los trabajos que exigen un mayor grado en cualquiera de estos factores son los menos susceptibles a ser computerizados. Por ejemplo, aunque la labor de venta a distancia requiera cierta interacción con el cliente, no son necesarias las habilidades sociales de un consejero matrimonial, que ha de ser capaz de mostrar un alto grado de empatía. El informe aventura que se producirán dos oleadas en la computerización de trabajos, la primera debida a tecnología ya existente de aprendizaje automático y de sensores, que afectará a labores en el sector de transportes y logística, en labores de soporte administrativo, y en labores de producción, y una segunda, que requerirá superar los cuellos de botella tecnológicos mencionados, situando los trabajos que exigen creatividad y un elevado grado de inteligencia social en la escala de los de menor riesgo de ser computerizados. Es la misma conclusión a la que llega el estudio más reciente antes mencionado:

«Será difícil que las máquinas reemplacen a los humanos en ocupaciones del sector de servicios que requieran compasión, creatividad e inteligencia social» (Oxford Economics, 2019, pág. 7).

Habilidades del ser computerizado

Cada uno de ellos es desglosado a su vez en otros factores, por ejemplo, en la inteligencia social se considera la perceptibilidad social, la capacidad de negociación, la persuasión, o la asistencia y cuidado de otros.

Este segundo estudio constata el desplazamiento de trabajadores del sector productivo por la automatización/robotización hacia los sectores del transporte, construcción y mantenimiento, y trabajo administrativo y de oficina, que son justamente los sectores más vulnerables a la automatización durante las próximas décadas. Los autores desaconsejan –a pesar de ello– llevar a cabo políticas para desacelerar la adopción de tecnología robótica, ya que esta repercute globalmente de forma muy positiva en la economía, y sugieren en cambio que el *dividendo robótico* se destine en parte a mitigar la situación de los trabajadores más vulnerables, así como favorecer su adaptación e integración en el nuevo panorama laboral y social.

Reflexión

Es importante destacar que el informe insiste en que únicamente considera en el estudio trabajos actualmente existentes, y no especula sobre posibles ocupaciones que emerjan debido precisamente a la computerización y robotización.

2.2.2. El ámbito del transporte: vehículos autónomos

Un vehículo autónomo es un medio de transporte automotriz y dotado de sensores que le permiten percibir su entorno y llegar a su destino sin la intervención de un conductor humano. Un avión operando en modo piloto automático puede considerarse un vehículo autónomo, pero el término se refiere en su acepción más difundida a vehículos terrestres, en concreto a coches y camiones. Los sensores y sistemas requeridos para la conducción autónoma incluyen detectores de proximidad ultrasónicos, radar y lidar, cámaras de visión y GPS. También se requiere un ordenador a bordo capaz de procesar toda la información suministrada por los sensores e interpretarla en términos de vías de circulación, señales de tránsito, otros vehículos, peatones, ciclistas y, en general, obstáculos fijos y móviles en la calzada. La posibilidad de comunicación con otros vehículos autónomos y/o con sistemas centralizados de gestión de tráfico abre las vías para una circulación más segura y eficiente, con menos accidentes y mayor fluidez. La incidencia del vehículo autónomo se produce tanto en el ámbito laboral (transporte de mercancías y de personas) como en el privado.

El Estándar SAE J3016 define los siguientes seis niveles de autonomía (SAE International, 2019):

- Nivel 0. Sin autonomía (puede haber avisos y control puntal).
- Nivel 1 (*hands on*). Control compartido; por ejemplo, Adaptive Cruise Control: el humano atiende a la dirección del vehículo mientras que el sistema controla la velocidad. Parking Assistance: viceversa. El humano ha de poder acceder al control total en cualquier momento.
- Nivel 2 (*hands off*). El sistema controla la dirección, aceleración y frenada. El humano supervisa y está dispuesto a intervenir en cualquier momento.
- Nivel 3 (*eyes off*). El humano puede distraerse, leer, mirar una película..., pero ha de poder intervenir en un tiempo determinado, avisado por el sistema.

- Nivel 4 (*mind off*). El humano puede dormir o abandonar el asiento del conductor, en áreas o situaciones determinadas. Fuera de estas, si el humano no retoma el control, el sistema aborta el viaje (aparca).
- Nivel 5 (*steering wheel optional*). Ninguna intervención humana (por ejemplo, taxi robot).

Cada nivel define un reparto diferente de responsabilidades entre el conductor y los fabricantes/programadores del sistema en caso de accidente. Entre las ventajas del vehículo autónomo se cuentan:

- Disminución del riesgo de accidentes, al eliminarse posibles imprudencias (consumo de estupefacientes, saltarse las reglas o señales de tráfico, distracciones, etc.) y evitar otros vehículos gracias a una red de interconexión.
- Mayor comodidad para usuarios, mejora de la movilidad para no conductores.
- Reducción de embotellamientos de tránsito, pacificación del tránsito urbano e interurbano.
- Favorece el vehículo compartido, menor ocupación de la vía.
- Reducción de los costes de los seguros.
- Mayor eficiencia energética.

En contra, pueden esgrimirse los siguientes argumentos:

- Reluctancia de los conductores a abandonar sus hábitos.
- Aumento del desempleo en el sector del transporte y de servicios asociado (restauración, hostelería...).
- Privacidad comprometida (seguimiento de los desplazamientos).
- Desplazamiento de la población a áreas suburbanas, mayor invasión humana del territorio.
- Posibilidades que ofrece al pirateo, control remoto, terrorismo.

Desde el punto de vista ético, una de las cuestiones más relevantes y polémicas es el de las reacciones programadas a situaciones límite. Son variantes de lo que en ética se conoce como el **dilema del tranvía**:

«Un tranvía corre fuera de control por una vía. En su camino se hallan cinco personas atadas a la vía por un filósofo malvado. Afortunadamente, es posible accionar un botón que encaminará al tranvía por una vía diferente, por desgracia, hay otra persona atada a esta. ¿Debería pulsarse el botón?» (Foot, 1978)

Una versión de este dilema, ya en el contexto de la conducción de un vehículo, es el **dilema del túnel**:

«Estás viajando solo en un vehículo autónomo por una carretera de montaña de un solo sentido y te acercas rápidamente a un túnel estrecho. Justo antes de entrar al túnel un niño intenta cruzar corriendo, pero tropieza y cae bloqueando la entrada al túnel. El vehículo [no puede frenar lo suficientemente rápido para evitar golpear al niño, por lo que] tiene dos opciones: golpear y matar al niño, o girar el volante y chocar contra la pared del túnel, matándote a ti. ¿Cómo debería reaccionar el coche?» (Millar, 2014)

Una cuestión relacionada es: ¿quién debería tomar esta decisión? El autor del dilema y del artículo sostiene que no debería ser una decisión impuesta por diseño, ya que de este modo los constructores/programadores del vehículo estarían incurriendo en *paternalismo* en lo que se refiere a decisiones que competen al conductor directamente afectado. En una encuesta llevada a cabo en el mismo portal web (Open Roboethics Initiative, 2014), un 64 % de los encuestados declararon que, de ir ellos como pasajeros del vehículo, este debería seguir recto y atropellar al niño. Por lo que se refiere a la pregunta de quién debería determinar la respuesta del vehículo, un 44 % apostaba por el pasajero, un 33 % por los legisladores, un 12 % por el fabricante/diseñador y un 11 % por otros.

En una plataforma de internet montada por investigadores del MIT es posible juzgar diferentes variantes del problema del coche autónomo que ha de decidir entre atropellar peatones que cruzan una calle o estrellarse contra una barrera sólida (o desviarse y atropellar otro grupo de peatones). Tanto los ocupantes del vehículo como los peatones varían en número y en características (edad, género, estado de salud, reconocimiento social, etc., incluyendo hasta animales de compañía), y también hay variantes tales como si los peatones están cruzando en rojo o en verde, o diferentes grados de afectación por el accidente (muerte, heridas de diferente gravedad, etc.). Los resultados obtenidos hasta mediados de 2018 de millones de encuestados de 233 países y territorios están recogidos y analizados en un artículo de la revista *Nature* (Awad *et al.*, 2018). Las diferencias culturales quedan en evidencia en el estudio: por ejemplo, países prósperos con instituciones fuertes son menos propensos a salvar a peatones que cruzaron en rojo. Esto ilustra las dificultades de elaborar un código ético universalmente válido, aunque tampoco era esta la finalidad del estudio, sino la de resaltar dichas diferencias culturales. Lo más probable es que los coches se limiten a seguir las normas de circulación de cada país (que de hecho ya contemplan algunas contingencias más probables, como que cerca de un parque infantil o un colegio pueda salir inesperadamente un niño entre los coches, por lo que hay que circular con especial precaución en estas zonas), y tendrán un seguro para cubrir daños.

Otra cuestión relacionada es que una reacción preprogramada puede interpretarse como un algoritmo que *apunta* a determinados colectivos o individuos. Citando el ejemplo de Millar (2017), un sistema que busque minimizar el daño causado «preferirá» atropellar a un ciclista que (cumpliendo con las normas de tráfico) lleve casco a uno que no. En este caso, estaría penalizando a los ciclistas responsables.

MIT<http://moralmachine.mit.edu/>

2.2.3. El ámbito del conflicto: robots militares

En un conflicto armado, la tipología de robots comprende las siguientes categorías, según la misión que lleven a cabo:

- Misiones humanitarias y tácticas no cruentas: desminado de un área, desactivación de explosivos, rescate de heridos, reconocimiento, logística y transporte.
- Misiones tácticas defensivas: dispositivos montados sobre tanques, buques u otros vehículos, o bien situados en una frontera o perímetro a defender (sistemas antimisiles), que responden de forma inmediata y automática a cualquier ataque detectado sobre los mismos.
- Misiones tácticas ofensivas: armas móviles (mayoritariamente por aire, aunque también hay que se desplazan por tierra) empleadas en la neutralización del enemigo.

Desde un punto de vista ético, las LAW (*lethal autonomous weapons*, armas letales autónomas) cobran una especial significación. Existen diferentes interpretaciones por lo que se refiere al grado de autonomía: algunos autores la formulan a muy alto nivel, entendiéndolo como tal que, describiendo un objetivo global, la máquina fija los objetivos particulares, los busca y detecta, y los elimina. Para otros autores, el calificativo de autónomo ya es aplicable si el robot es capaz de detectar y seguir un objetivo concreto, aunque la decisión de hacer fuego recaiga en un ser humano. El formato más corriente en la actualidad es la de un vehículo aéreo no tripulado (UAV, por *unmanned aerial vehicle*) equipado con misiles que pueden ser disparados sobre el objetivo. Estos UAV disponen de cámaras que pueden proporcionar una señal de vídeo al operador humano, de modo que este estará al tanto de la situación en todo momento. El operador humano puede estar «en el bucle» (*human in the loop*), teleoperando el vehículo la mayor parte del tiempo, excepto durante la fase más rutinaria de vuelo en piloto automático, o bien «sobre el bucle» (*human on the loop*), donde se limita a monitorizar el funcionamiento del vehículo, encargándose este de forma autónoma de las tareas de despegue y aterrizaje, así como de la determinación y búsqueda del objetivo. Algunas operaciones son llevadas a cabo por un grupo (*swarm*, enjambre) de UAV de forma coordinada: ante la imposibilidad de controlar cada vehículo de forma individual, se impone el modo de operación *on-the-loop*. En cualquiera de los casos, hay un consenso generalizado de que ha de ser un humano quien tome la decisión final sobre el establecimiento o la eliminación de un objetivo específico. De este modo, la atribución de responsabilidades por las acciones o incluso por el mal funcionamiento del robot se situaría claramente en los elementos humanos de la cadena de mando. Esto es evidente en el caso de los robots sin capacidad deliberativa sobre sus acciones. Volveremos sobre el tema de la responsabilidad y los robots autónomos en la tercera parte.

LAW

También conocidas como LAW (lethal autonomous weapons), LAR (lethal autonomous robots), robotic weapons (armas robóticas) o, más popularmente, killer robots (robots asesinos).

UAV

También conocidos por el anglicismo *drones*. El término análogo UAS (*unmanned aerial systems*) se refiere al conjunto formado por el vehículo aéreo y la estación de control en tierra, el sistema de comunicaciones, etc.

Human out of the loop

También podría estar «fuera del bucle» (*human out of the loop*), con el sistema operando de forma autónoma, sin supervisión.

Podría decirse que estos sistemas constituyen la culminación de un proceso observado a lo largo de la historia de la humanidad, y más específicamente de los conflictos bélicos, consistente en ir aumentando la distancia entre los contendientes, para contrarrestar los dos miedos a los que se enfrenta el soldado: el miedo a morir y el miedo a matar (Sharkey, 2014). En un UAS, el enemigo puede estar a miles de kilómetros del operador. A la distancia física se suma el efecto de seguir la operación a través de una pantalla, que se asemeja a la dinámica de los videojuegos, lo que incrementa el factor de desapego con la realidad del escenario bélico.

El Derecho Internacional Humanitario (DIH) establece los siguientes principios relativos a las acciones militares durante un conflicto armado (*ius in bello*):

- **Necesidad militar:** Un ataque o acción ha de servir para contribuir a derrotar al enemigo. Ha de ser un ataque sobre un objetivo militar legítimo.
- **Proporcionalidad:** Los daños causados a civiles o propiedades civiles han de ser proporcionales y no excesivos en relación con los beneficios militares previstos.
- **Distinción:** Los combatientes han de ser capaces de distinguir entre combatientes y civiles. Las leyes de protección se extienden incluso a combatientes heridos o que se han rendido.

También hay un consenso generalizado en no utilizar armas o métodos que generen un sufrimiento superfluo, gratuito e indiscriminado, como queda reflejado en disposiciones adicionales del DIH que se mencionan más adelante.

Un robot combatiente totalmente autónomo ¿cumplirá mejor o peor los principios del DIH? Algunos autores sostienen que la ausencia de emociones tales como miedo o sed de venganza facilita actuar con objetividad y ajustar el propio comportamiento a los preceptos del DIH (Arkin, 2009). Lokhorst y Van den Hoven (2014) consideran que un robot bélico puede ser programado para la incapacitación temporal del enemigo antes que matarlo, y proceder de este modo incluso en situaciones donde un soldado humano posiblemente no tuviera otra opción que matar. Ahora bien, existen fundadas reservas sobre que los sistemas de percepción actuales (y de un futuro próximo) con que están equipados los robots permitan interpretar una escena de modo que se garanticen los principios de proporcionalidad y de distinción. No es solo una cuestión sensorial, sino de interpretación: en lo que se refiere a la proporcionalidad, «no hay ninguna métrica conocida para medir objetivamente el sufrimiento innecesario, superfluo, o desproporcionado» (Sharkey, 2014, pág. 123). Y en cuanto al principio de distinción o de discriminación, requiere a menudo entender el contexto y las intenciones del otro³.

Derecho Internacional Humanitario

El DIH se aplica únicamente en caso de conflicto armado. Está contenido en los cuatro Convenios de Ginebra de 1949, de los que son firmantes casi todos los Estados. Estos Convenios se completaron con los Protocolos adicionales de 1977 relativos a la protección de las víctimas de los conflictos armados. Extraído del documento del Comité Internacional de la Cruz Roja *¿Qué es el Derecho Internacional Humanitario?* <https://www.icrc.org/es/doc/assets/files/other/dih.es.pdf>

⁽³⁾Lo que en filosofía y psicología se entiende como la teoría de la mente, la capacidad de atribuir a los demás estados de la mente, esto es, creencias, intenciones, deseos, emociones, etc.

En la guerra moderna, no hay una distinción manifiesta basada en el uniforme; muchas veces se trata de los denominados «insurgentes», que no visten de forma especial. Abundan las situaciones que requieren una capacidad de discriminación fina y matizada: niños obligados a llevar fusiles descargados, insurgentes enterrando a sus muertos, o el ejemplo citado en Guarini y Bello (2014, pág. 130) de un par de niños *sikhs* jugando a perseguirse con su *kirpan* (puñal ritual) en el momento en que las fuerzas de contrainsurgencia entran en la casa donde se encuentran. Estos autores insisten en el papel fundamental que las emociones juegan en resolver de forma eficiente la problemática asociada a la isotropía y la teoría de la mente, donde un proceso deliberativo necesitaría un tiempo del que no se dispone y/o unos recursos computacionales extraordinarios.

Un Estado puede apostar por la tecnología robótica para reducir los efectivos humanos en el teatro de operaciones y, por consiguiente, las posibles bajas humanas (entre sus filas). Este mismo factor, por otro lado, puede facilitar que dicho Estado tenga menos reticencias a la hora de emprender acciones beligerantes, en detrimento de la diplomacia, ya que no deberá enfrentarse al impopular efecto del «retorno de ataúdes». Aún más, el factor económico también será muy relevante. Como se apunta en Kahn (2017), no solo repercutirá el coste asociado a la pérdida de vidas humanas (que también se traduce en pensiones, indemnizaciones, etc.), sino que un UAV capaz de cargar una bomba de 250 kg como el MQ-9 Reaper cuesta, aproximadamente, una cuarta parte de lo que cuesta un avión a reacción tripulado con la misma capacidad, como el F/A-18F, y el coste asociado a entrenar un operador de UAV está en una relación de 1 a 360.000 con el de entrenar a un piloto. El efecto neto puede ser el aumento de la conflictividad, que a pesar de la presencia de robots en las contiendas también producirá más víctimas humanas. A ello cabe sumar que los Estados en desventaja tecnológica opten por otras vías como el terrorismo.

En 2013 se inició la campaña internacional «Stop Killer Robots»: 272 científicos, ingenieros, expertos en IA y robótica de treinta y siete países pidieron prohibir el desarrollo e instalación de armas plenamente autónomas. La cuestión es si prohibir las LAW garantizará que no se utilicen. Disposiciones adicionales del DIH hacen referencia a la prohibición de armas específicas, como la Convención sobre Armas Biológicas (1972), la Convención sobre Ciertas Armas Convencionales y sus cinco Protocolos (1980, son restricciones sobre armas de fragmentación, armas trampa, armas incendiarias, o armas láser cegadoras), la Convención sobre Armas Químicas (1993) y el Tratado de Ottawa de 1997 sobre las Minas Antipersona. Además, en 1970, se firmó el Tratado de No Proliferación Nuclear⁴, a revisar en 2020, y en 2017 el Tratado sobre la Prohibición de las Armas Nucleares (que solo ha sido ratificado por once países (a julio de 2018)⁵.

Reflexión

Que un sistema artificial pueda simular algo que cumple el cometido computacional de las emociones no supone que este sistema realmente sienta, como también indican los autores.

⁽⁴⁾Con la excepción de India, Pakistán, Israel, Sudán del Sur y Corea del Norte.

⁽⁵⁾Se requiere la ratificación de al menos cincuenta países para que entre en vigor.

Stop Killer Robots

A día de hoy (octubre de 2019), la coalición abarca 118 ONG de 59 países.

Aun así, Estados Unidos y Rusia poseen cerca de quince mil armas nucleares (en total). En muchos casos, la prohibición de un arma llega años después de haber sido utilizada, y en muchas ocasiones no todos los países firman o ratifican los acuerdos. En países como Estados Unidos toda nueva arma requiere una autorización donde debe demostrarse que puede usarse cumpliendo el principio de proporcionalidad, que puede ser controlada de modo que solo vaya dirigida a un objetivo específico legalmente aceptable, y que no hay ningún tratado o ley doméstica que la prohíba. No obstante, la combinación de dispositivos y armas ya aprobados no tiene por qué pasar el examen, como determinó la oficina del cuerpo jurídico de las fuerzas armadas americanas (Judge Advocate General) en el caso de un UAV, el Predator UCAV, destinado inicialmente a vigilancia, cuando fue armado con misiles Hellfire (Sharkey, 2014, pág. 119).

¿Y qué pasa cuando no son las fuerzas armadas regulares de un país democrático las que utilizan este tipo de armamento? Un militar de carrera generalmente concederá gran importancia al concepto de honor, que en su caso va asociado al desempeño de su actividad respetando las «reglas del juego»; es lo que se denomina la «ética del soldado» (O'Meara, 2014). Es consciente, además, de que sus acciones serán sometidas a escrutinio, dentro o fuera del ámbito militar. No sucede lo mismo con otros colectivos. Las agencias de inteligencia, por ejemplo, proceden de forma opaca: la CIA ha realizado numerosas acciones usando UAV para «decapitar» células insurgentes, amparándose en que se trata de operaciones encubiertas para no dar cuenta de estas ejecuciones extrajudiciales (Sharkey, 2014, pág. 114). A estas personas, además de la vida, se les ha despojado del derecho al recurso o a la rendición. Paramilitares, mercenarios que actúan como fuerzas de seguridad de grandes empresas y terroristas son otros colectivos que actúan al margen de los acuerdos internacionales de la *ius in bello*.

2.2.4. El ámbito de la salud: prótesis, sistemas expertos y robots cirujanos

Sea por accidente o agresión (esto es, por trauma), por enfermedad adquirida o por malformación genética, los seres humanos estamos sometidos a la eventualidad de una pérdida funcional sensorial o motriz asociada al deterioro o eliminación de un órgano, extremidad o tejido neuronal. A lo largo de la historia hemos intentado suplir alguna de estas pérdidas con adminículos o artilugios, siendo paradigmática la imagen del pirata con la pierna de palo, el garfio y el ojo de vidrio o el parche (que en este caso únicamente cumple una función estética, aunque tampoco debe menoscabarse su importancia). Pero no cabe duda de que es en tiempos actuales y años venideros cuando los avances tecnológicos permitirán disponer de prótesis cada vez más perfeccionadas y funcionales. Parte de los avances provienen de la ciencia de los materiales, que permiten obtener prótesis cada vez más ligeras y resistentes, así como más compatibles con los tejidos humanos. Pero es sin duda a través de la robótica que se han conseguido los mayores logros en cuanto a recuperar la movilidad

perdida. El paradigma *percibe-piensa-actúa* tiene pleno sentido en este caso: los biosensores captan la actividad muscular o neuronal de los tejidos funcionales del paciente, el controlador se encarga de procesar estas señales, así como las que provienen de la misma prótesis (por ejemplo, posición de los dedos o fuerza de impacto en el pie durante la pisada), y los actuadores son los motores o dispositivos que generan el movimiento o la adaptación activa de la prótesis. La función cognitiva queda repartida, de este modo, entre el controlador y el propio cerebro del paciente.

En principio, no debería existir ninguna objeción ética a las prótesis que suplen una funcionalidad perdida: se trata de restaurar las capacidades que un ser humano ya tenía, o que son comunes en un ser humano. No obstante, el debate ético está justificado en al menos tres cuestiones relacionadas con esta tecnología:

- **Seguridad.** Esta cuestión incluye dos aspectos: fiabilidad e intrusión. La fiabilidad se refiere al correcto funcionamiento del producto, que cumpla no ya las expectativas del usuario, sino las propias especificaciones del fabricante. En el caso de un producto con las características de una prótesis este aspecto es especialmente sensible, ya que no es un mero aparato que deja de funcionar (o funciona mal) y el usuario puede seguir su vida mientras es reparado, sino que es la propia integridad y movilidad/funcionalidad del usuario la que se pone en jaque. Un problema que en otras circunstancias no pasa de ser una molestia, como es el sobrecalentamiento de las baterías, en el caso de una prótesis puede causar quemaduras en el usuario (además de estropear la prótesis). Podría darse el caso de que una prótesis causa daño a un tercero (por ejemplo, una mano robotizada que aprieta tanto la mano de una persona a la que se está saludando que le causa lesiones). Tanto en un caso como en el otro habrá que determinar si el origen del fallo está en un mal diseño, una fabricación defectuosa o un mal uso de la prótesis. El segundo aspecto, el de la intrusión por parte de un *hacker* que toma el control sobre la prótesis, es especialmente tenebroso, ya que abre la puerta a un nuevo tipo de criminalidad. Ya se han dado casos de *hackers* que habían podido acceder al control de bombas de insulina de pacientes diabéticos, a los que potencialmente habrían podido asesinar con una simple sobredosis de insulina. En este caso, la responsabilidad recae evidentemente sobre el ciberintruso, aunque también puede que el fabricante no haya incluido suficientes cortafuegos para evitar su acceso, o que el usuario haya sido negligente a la hora de poner los correspondientes antivirus al día.
- **Desigualdad.** Las prótesis robóticas tan avanzadas al principio solo estarán al alcance de una minoría con suficiente poder económico. Incluso entre individuos con poder adquisitivo insuficiente puede darse algún tipo de discriminación, como en el caso de un Estado que financia la prótesis de un soldado que ha ido voluntariamente a luchar en una guerra lejana y no en cambio la de alguien que ha perdido un miembro en un accidente. Una

prótesis no será nunca un producto de consumo de masas, no solo porque afortunadamente los amputados representan un porcentaje reducido de la población total, sino sobre todo porque cada prótesis ha de ser adaptada a un usuario único, y al coste de la prótesis en sí hay que añadir el de la rehabilitación y entrenamiento del paciente. Con el avance y la mejora de las prótesis, estas pueden convertirse en otro factor de desigualdad social más acusado, como se comenta a continuación.

- **Capacidades sobrehumanas.** Ya en la actualidad una prótesis, junto con sus limitaciones y defectos, también aporta una serie de ventajas a su usuario: la ausencia de dolor experimentada en la misma⁶ y la dureza de los materiales constitutivos pueden otorgar cierta ventaja si el usuario decide emplearla agresivamente. Más allá de la restauración de capacidades, en un futuro una prótesis puede mejorar las mismas por encima de las posibilidades biológicas: puede otorgar mayor fuerza, velocidad, resistencia, destreza... Prótesis cerebrales podrían mejorar capacidades cognitivas como el aprendizaje o la memoria, y prótesis sensoriales abrir todo un nuevo mundo de percepciones. Estas perspectivas pueden inducir (a quien pueda permitírselo) a prescindir de miembros y órganos sanos para sustituirlos por las versiones artificiales mejoradas. Bien es cierto que los humanos pretendemos evolucionar y mejorar como especie, pero sin duda esta tendencia crearía a corto plazo una élite que añadiría a su fortaleza financiera una ventaja desleal en términos físicos y psicológicos. La complejidad de las causas legales que se derivarían de los conflictos inevitables entre humanos biológicos y humanos mejorados sería ciertamente elevada. De entrada, las recomendaciones presentadas en el Parlamento Europeo en 2017 promueven la creación de comités de ética en los hospitales que incluyan a expertos en roboética para hacer frente a las decisiones sobre la conveniencia de implantar prótesis para mejoras específicas.

⁽⁶⁾No estamos hablando de posibles llagas que por rozadura pueden producirse en la conexión con el miembro remanente, que pueden ser muy dolorosas.

Sin abandonar el campo de la salud humana, otros elementos de gran significación ética son los sistemas expertos utilizados para la diagnosis médica y los robots cirujanos. Los primeros son programas de IA que, a partir de la sintomatología (en forma de preguntas a las que los análisis y pruebas médicas van dando respuesta), son capaces de proporcionar posibles diagnósticos. Estos sistemas expertos pueden en gran medida contribuir a reducir el error médico. Como siempre, el dilema ético emerge cuando las cosas van mal: ¿quién es responsable de un diagnóstico erróneo, que posiblemente haya llevado a un tratamiento inadecuado, con nefastas consecuencias para el paciente? ¿Los expertos médicos que participaron en el desarrollo del mismo, los programadores, el médico que lo ha utilizado, el personal que ha hecho las pruebas o introducido los datos? En principio, una investigación debería poder dilucidar dónde radica la fuente del error. También podría darse el caso de un médico que cree que el diagnóstico que proporciona el sistema experto es erróneo, pero por no cargar con el riesgo de equivocarse decide darlo por bueno.

Por lo que se refiere a los robots cirujanos, a día de hoy actúan generalmente como brazos telemanipuladores, en el sentido de que no operan de forma autónoma sino reproduciendo los movimientos del cirujano, aunque con un pulso mucho más estable y a escala microscópica. Por regla general, su introducción ha supuesto una mejora en la calidad de las intervenciones, que resultan mucho menos invasivas y, por tanto, reducen el posoperatorio y favorecen la recuperación del paciente. Posiblemente, hayan evitado un considerable número de fallecimientos en la mesa de operaciones, pero no son del todo infalibles y también se ha dado el caso de un fallecimiento durante una operación por un fallo en el sistema de control. Un fallo de este tipo puede ser de diseño, de fabricación o de programación, más difícilmente de uso ya que precisamente estos sistemas incorporan salvaguardas para evitar reproducir movimientos bruscos o desmesurados del cirujano. No obstante, también podría ocurrir que un cirujano haya observado alguna anomalía en el sistema y, a pesar de ello, haya decidido seguir adelante con la operación. De poder demostrarse, sin duda parte de la responsabilidad recaería sobre el propio cirujano, aunque este podría ser víctima de un exceso de confianza (*overtrust* en inglés) en el sistema robótico, por el mero hecho de ser un sistema automático. Se trata de un fenómeno transversal: el médico al que aludíamos al final del párrafo anterior podría ser otro caso de confianza excesiva, o unos padres (o un fisioterapeuta) que confían excesivamente en la pauta impuesta por un robot rehabilitador a pesar de que tras las primeras repeticiones el paciente infantil muestra evidentes signos de malestar (Borenstein *et al.*, 2017). Esta confianza excesiva o incluso ciega en los sistemas automáticos también se da en otros ámbitos, como por ejemplo el bélico, con muertes causadas por fuego amigo en un sistema automático aún con un humano en el bucle (Kirkpatrick *et al.*, 2017), o el del transporte, con un accidente de aviación causado por confiar en lecturas erróneas del piloto automático, por no mencionar conductores que, confiando en su GPS, se metieron con el coche en un lago, en el mar, por un barranco, o en un desvío de más de 2.500 km (Borenstein *et al.*, 2017). Probablemente, los propios sistemas deban ser diseñados (o los usuarios entrenados) para no generar este tipo de exceso de confianza.

2.2.5. El ámbito doméstico: seguridad, privacidad, respeto a la intimidad

La frase del siglo XVI «la casa de un hombre es su castillo» pone de relieve la distinción, espacial en este caso, entre la esfera pública y privada. Exterior es sinónimo de estar expuesto, a los ojos de la sociedad, del Estado, de los vecinos y transeúntes, y también a innumerables peligros, que se asumen al traspasar el umbral. Puertas adentro es el ámbito de lo doméstico, de lo privado, de lo íntimo, también de la seguridad. De forma metafórica, los humanos también delimitamos y separamos ambas esferas, y vigilamos celosamente que no se traspase la frontera entre ambas. Aun así, hoy en día la privacidad está más entredicho que nunca. En gran medida, esto sucede de forma inadvertida: mediante el uso que hacemos de ordenadores personales y telefonía móvil, programas espía, no necesariamente de delincuentes cibernéticos ocultos sino

de empresas o instituciones, pueden seguir nuestros movimientos y nuestras transacciones, y mediante el análisis del *big data*, la ingente cantidad de datos que cada día vertemos en internet, también pueden elaborar un perfil de nuestras preferencias y afinidades. No siempre ocurre de forma inadvertida: frecuentemente, somos nosotros mismos quienes a través de las redes sociales exponemos nuestros datos, predilecciones y debilidades, en un ejercicio de exhibicionismo digital propio de una sociedad narcisista y egocéntrica. A pesar de ello, creemos retener cierto control sobre lo que queremos mostrar, y en el peor de los casos lo que exponemos son datos y lo que puede resultar dañado es la propia imagen (aunque en ocasiones el daño también puede ser económico). Caso aparte es el de los menores (y el de algunos adultos inconscientes), que pueden verse expuestos a alguna suerte de ciberacoso. Con la introducción de los robots, capaces de percibir, procesar, guardar en memoria y compartir cuanto sucede a su alrededor, así como de desplazarse de un lugar a otro, el panorama puede dar un nuevo salto cualitativo.

Ryan Calo (2014) distingue tres aspectos en los que los robots ponen la privacidad en entredicho:

- Facilitan la **vigilancia directa**, al poder revestir cualquier forma y tamaño, no necesitar descanso ni sufrir distracciones, y estar equipados con múltiples sensores que exceden las capacidades perceptuales humanas. No solo los gobiernos pueden abusar de estas nuevas capacidades de vigilancia, sino también corporaciones, medios e individuos en

«arenas tan diversas como seguridad, *vouyerismo* o marketing» (Calo, 2014, pág. 188).

- Proporcionan nuevas **vías de acceso** a espacios tradicionalmente protegidos, especialmente el domicilio particular, a través del robot doméstico. El *hacker* potencialmente puede acceder no ya a ficheros y documentos, sino a habitaciones, objetos y personas. El debate legal cuestiona, entre otras cosas, que haya habido una invasión punible de la privacidad previamente a que algún ser humano haya accedido a la información recogida por el robot. Si el usuario accede a que su robot doméstico o personal vierta la información recogida en un servidor externo para su procesamiento, o que esté conectado a internet no solo para el seguimiento por parte del propio usuario, sino también para actualizaciones o búsqueda de información para mejorar sus prestaciones, de algún modo está aceptando comprometer su propia privacidad, lo cual a su vez puede suponer una merma voluntaria de los derechos constitucionales sobre su privacidad.
- La dimensión social del robot no solo contribuye a disminuir mediante su presencia el espacio y el tiempo de soledad necesario para la interiorización y la autorreflexión (lo que puede a su vez producir daños psicológicos), sino que puede inducir al usuario a compartir los pensamientos más íntimos. Un robot puede llegar a convertirse en el confidente de una per-

sona con reducida interacción social, y terceros que tengan acceso a estas confidencias pueden aprovecharse de eventuales vulnerabilidades.

Para Ryan Calo, la ley no está bien equipada para hacer frente a la dimensión social del robot:

«El consentimiento informado tiende a derrotar las reclamaciones sobre privacidad, y es difícil medir el daño en causas sobre privacidad» (Calo, 2014, pág. 198).

También es de destacar que los robots pueden tener su propio internet a través del cual adquirir la información o las habilidades para realizar la tarea encomendada sin mediación humana (aunque los humanos sí pueden introducir conocimiento en dicha red). Tal era el objetivo del Proyecto Europeo FP7 RoboEarth (European Commission, 2014), que concluyó en 2017. Básicamente, se trata de una red de *cloud computing* (computación «en la nube», esto es, distribuida en una red de servidores) y un repositorio que contiene componentes de software, mapas para navegación, conocimiento sobre tareas (por ejemplo, recetas de acción, estrategias de manipulación) y modelos para el reconocimiento de objetos, todo ello en un lenguaje inteligible para las máquinas.

El Proyecto Europeo FP7 RoboEarth...

Después continuó en los proyectos RoboHow, KnowRob y OpenEase.

Es importante ser conscientes de que no solo el PC, el móvil y el robot pueden constituirse en voceros de nuestra intimidad: con la progresiva introducción y difusión de objetos y utensilios cotidianos «inteligentes», y a través de lo que se ha dado en llamar *Internet of Things* (IoT), se abren otras tantas micropuertas que pueden dejar salir datos sobre nuestra localización, nuestros movimientos comerciales o financieros, nuestros gustos y hábitos, nuestras relaciones, nuestra biometría... Puede que conjuntos de tales artefactos estén conectados únicamente en el ámbito local (esto es, en ámbito restringido) para mejorar su eficiencia y seguridad, pero también es muy probable que se establezcan conexiones con internet, por ejemplo para permitir actualizaciones.

2.2.6. El ámbito de la intimidad: relaciones afectivas

Los humanos tendemos a desarrollar un vínculo hacia determinados objetos personales, sea por los recuerdos asociados (por ejemplo, de la persona que nos los dio), sea por la funcionalidad que proporcionan (comodidad en el aseo personal, eficacia en la cocina, movilidad...), o bien por cierta propensión a antropomorfizar objetos por su aspecto o comportamiento. Con un robot, este vínculo puede llegar a ser mucho más profundo, ya que de forma natural resulta fácil atribuir una personalidad única a esta máquina, por los siguientes motivos:

- Se mueve de forma frecuentemente no prevista, de modo que inconscientemente podemos vernos tentados de atribuirle autonomía e intencionalidad propia.

- Tiene un comportamiento más rico y variado que otros aparatos.
- La interacción con el usuario humano es de más alto nivel, comprendiendo instrucciones y réplicas verbales, comunicación gestual, etc.
- Tiene un aspecto más antropomorfo (o con forma de mascota), a menudo justificable desde un punto de vista funcional.
- La interacción tiene lugar de forma regular y frecuente, diaria o incluso permanentemente.

Este tipo de vínculos más estrechos puede desarrollarse con robots personales cuyo aspecto es manifiestamente el de una máquina, e incluso con aparatos o software que, sin tener la presencia física de un robot, sí desempeñan un papel destacado en el día a día, como los asistentes personales que nos mantienen informados sobre lo que nos interesa, están al tanto de nuestra agenda y nos recuerdan nuestras citas (un simple altavoz que da voz a una IA incorpórea). Pero incluso puede producirse de una forma mucho más acusada cuando existe una dependencia física del robot, o la capacidad de discernimiento del usuario sobre la verdadera naturaleza del robot se ve condicionada por la edad u otros factores. Los robots que precisamente se ocupan del cuidado de personas, los robots asistenciales o cuidadores, suscitan un debate ético más profundo debido a la vulnerabilidad de sus usuarios. En Torras (2019) se revisan los logros de la robótica asistencial, y se presenta el material educativo empleado en un curso universitario sobre ética en robótica e IA social. En el caso de personas dependientes por edad o discapacidad, ¿es deseable que el usuario tenga un vínculo emocional con el robot? Podría serlo en los siguientes casos:

- Situaciones en que el robot actúe como un sustituto temporal y terapéutico para practicar modos saludables de comunicación interpersonal, siempre bajo la supervisión de un terapeuta. Las ventajas del robot radican en que puede modelar incansablemente diferentes comportamientos, y no juzga ni critica al usuario, por lo que este se siente menos cohibido en la interacción.
- Situaciones en que un vínculo emocional puede propiciar un comportamiento saludable del paciente, tal y como la toma puntual de medicamentos, animarse a una mayor movilidad y ejercicio físico, fomentar la inquietud intelectual, etc.

No obstante, también puede ser contraproducente cuando:

- Esta relación se torna obsesiva y/o excluyente.
- La pérdida del robot provoque sensaciones de duelo similares a la de la pérdida de un familiar o ser querido.

- Sustituyen las relaciones afectivas entre humanos a largo plazo (no se basan en la reciprocidad, son transacciones).
- Terceros (por ejemplo, los propios fabricantes del robot) se aprovechan de este vínculo para manipular al usuario (por ejemplo, para inducirlo a la adquisición de mejoras o complementos del robot, otros productos de la compañía). A este respecto, Matthias Scheutz advierte del peligro que suponen los lazos emocionales unidireccionales entre humanos y robots sociales: las dependencias psicológicas que crean pueden

«tener serias consecuencias para las sociedades humanas, porque pueden ser explotadas a gran escala» (Scheutz, 2014, pág. 216).

La regulación legal es necesaria; como afirma Blay Whitby (2014, pág. 234),

«es poco probable que los problemas sociales de la ética de los robots sea resuelta dejando que los mercados decidan libremente».

Estos factores negativos tienen una significación más acusada en el caso de colectivos vulnerables como niños, ancianos, personas con discapacidad física o mental, o personas con dificultades de integración social. Es justo para estos grupos humanos que el diseñador y fabricante de robots cuidadores haya de tener especialmente presente la necesidad de preservar la dignidad humana (Whitby, 2014). Pero tampoco hace falta formar parte de uno de estos colectivos para ser una víctima potencial de la manipulación por parte de un robot de maneras que no son concebibles para otro ser humano. Matthias Scheutz (2014, pág. 218) apunta que una forma radical de evitarlo es equiparlos, como hizo la evolución con nosotros, «con un sistema emocional que alcanza un equilibrio entre bienestar individual y un comportamiento socialmente aceptable». Pero el objetivo de llegar a implementar emociones genuinas en un sistema artificial no parece estar al alcance del desarrollo científico y tecnológico, o al menos no a corto plazo. En cambio, Meacham y Studley (2017) no están tan preocupados por el engaño subyacente en un robot cuidador en cuanto a la autenticidad de su predisposición hacia su usuario: lo relevante, para ellos, es que la atención dispensada por el robot cree un ambiente favorable para el cuidado o la recuperación del paciente. Sugieren que incluso una generación que haya crecido con semejante tecnología no necesariamente tienda a confundir robots con una sofisticada (sin ser auténtica) respuesta emocional con seres vivos, pero sí pueda atribuirles un tercer estado entre lo vivo y lo no-vivo.

Conforme los robots vayan alcanzando mayor sofisticación y antropomorfismo, también habrá lugar para otro tipo de interacción, en una esfera más personal e íntima. Cabe preguntarse si es legítimo mantener relaciones sexuales con un robot y si deben crearse robots con esta finalidad; o si el vínculo emocional puede llegar al enamoramiento y si puede contemplarse la unión legal en matrimonio entre un ser humano y un robot. Estas cuestiones no son materia de mera e hipotética elucubración, sino que ya empiezan a ser realidad: la industria está haciendo una importante inversión en muñecas (y muñecos)

sexuales cada vez más avanzados, y también se ha dado el caso de un japonés, Akihiko Kondo, que el 4 de noviembre de 2018 se casó con un holograma de la cantante virtual Hatsune Miku de un dispositivo asistente personal de Gatebox (Jiménez de la Fuente, 2019). El matrimonio no tiene validez legal, pero la compañía ya ha expedido decenas de semejantes certificados matrimoniales entre sus clientes y sus productos.

Potencialmente, un robot puede llegar a convertirse en un compañero sexual altamente gratificante, ya que puede ser modelado físicamente y por carácter al gusto del usuario, y es incansable, siempre disponible y predispuesto. Sus defensores sostienen que pueden cumplir un rol social de gran importancia:

- Pueden aumentar el nivel de satisfacción sexual del conjunto de la población, y por tanto contribuir a una sociedad más relajada, menos frustrada y violenta.
- Pueden proporcionar satisfacción sexual a personas con dificultades físicas y/o sociales.
- Pueden reducir el tráfico de personas y su explotación sexual.

Por contra, sus detractores creen que los efectos negativos que pueden tener no son menores:

- Provocar el deterioro de las relaciones interpersonales, al debilitar la tolerancia a las negativas o a opiniones diferentes a la nuestra, y al contribuir a la pérdida de la riqueza de matices (evidentemente, un robot también puede ser programado para expresar alguna negativa de vez en cuando, pero no dejará de ser un comportamiento programado).
- Relacionado con el punto anterior, pueden fomentar una actitud de menosprecio hacia el género representado en el robot, mayoritariamente femenino.
- Facilitar el aislamiento social (aunque también es factible que fomente las relaciones entre usuarios/propietarios de robots, de un modo similar a las mascotas con sus propietarios).

David Levy (2014), un firme defensor de los robots sexuales, establece paralelismos entre estos y otros productos de la tecnología que se han desarrollado para la autosatisfacción sexual, tales como muñecas inflables y vibradores, y los contempla como la lógica evolución y culminación de los mismos, y por tanto no moralmente objetables. Considera que el estigma social de alquilar un robot sexual sería menor que el de ser cliente de trabajadores sexuales humanos, y que la percepción del robot como un objeto evitaría sensaciones de infidelidad en una pareja humana uno de cuyos miembros (o ambos) decidiera utilizarlo. No obstante, cree que no hay que obviar el dilema ético que se plantea

Reflexión

El autor apunta al hecho de que hoy en día respetables cadenas de farmacias tienen vibradores en sus estanterías para la venta (véase también David Levy, *Love and sex with robots: The Evolution of Human-Robot Relationships*, Harper Perennial, 2008).

con los trabajadores humanos del sexo (aquellos que ejercen la profesión voluntariamente) que se quedarían en paro, ni, en algún futuro, con los propios robots, si estos llegan a adquirir algo semejante a la conciencia. El intercambio sexual es sin duda un camino hacia la familiaridad y la dependencia, pero Whitby cuestiona que necesariamente también lo sea para el amor, pues un requisito para ello es la reciprocidad, y ya hemos visto que poder replicar la capacidad emocional necesaria en un robot es, cuando menos, cuestionable.

Más allá de la gratificación sexual, los humanos pueden desarrollar algún tipo de afecto hacia un robot que los cuide o que sencillamente les «haga compañía». Es evidente que para robots sexuales un aspecto humano realista está justificado. Dichos robots tendrán éxito si son capaces de superar el «valle inquietante» (*uncanny valley*) que se produce en la aceptación del robot conforme aumenta su similitud con un ser humano (hay androides muy semejantes, aunque no del todo, que generan inquietud y rechazo). Pero el diseñador (y el fabricante) pueden apostar por la honestidad y crear, para otras finalidades como la asistencia, la educación o la compañía, robots que no disimulan por su aspecto que son máquinas, y aun así resultar muy atractivos. En su novela *La mutación sentimental*, la profesora Carme Torras (2018, pág. 246) funda la atracción que la protagonista Celia siente por su robot personal

«por su comportamiento leal, digno de confianza y previsible, reforzado por su apariencia no engañosa de máquina»⁷.

⁽⁷⁾La cita es del apéndice dedicado a la discusión de temas y cuestiones para grupos de lectura, solo en la versión en inglés.

Estas características no están reñidas con que el robot pueda proporcionar una conversación estimulante o una interacción enriquecedora. Elder (2017) establece el paralelismo con moneda falsificada (que intenta pasar por auténtica) frente al «dinero» del Monopoly, cuyo valor está claramente restringido al juego y no pretende engañar.

Otra cuestión que se plantea es el del legado: el robot personal que «sobrevive» a su dueño, atesorando sus recuerdos, manteniendo, quizá de forma análoga a un diario íntimo, parte de lo que era esta persona. ¿Cómo dispondrá dicha persona, aún en vida, de este legado? ¿Qué opciones legales tendrá? ¿Podrá mermar a sus herederos a favor del mantenimiento del robot? ¿Hasta qué punto podrán interferir sus herederos, tendrán algún derecho sobre el mismo? Cuestiones similares ya se plantean hoy en día respecto a la presencia digital de las personas en internet.

3. Iniciativas legales

La resolución de un conflicto generado por un robot o por una IA incorpórea muchas veces requerirá la intervención de los tribunales. En esta tercera parte, se describirá cómo las herramientas de la IA ya forman parte del mundo jurídico y puede que algún día ayuden en la resolución de este tipo de casos. La problemática más inmediata está relacionada con la seguridad, y en este sentido hablaremos de los fundamentos de la responsabilidad legal relacionada con los robots. También se describirán los primeros avances en establecer las leyes pertinentes, tanto en Europa como en otros países con el suficiente desarrollo tecnológico, y lo que ya se ha avanzado en el terreno de la normativa, tanto en la robótica industrial como en la de servicios. Por último, nos permitiremos un breve ejercicio especulativo sobre posibles derechos futuros de los propios robots.

3.1. IA y el mundo de las leyes

Los robots y la IA ya han entrado en el mundo jurídico. Los primeros, como se verá en el subapartado 3.2.1, en tanto que elementos involucrados en accidentes mortales. Estamos hablando de robots industriales y robots experimentales que han entrado en contacto con un reducido grupo de personas. Conforme se expandan en número y diversifiquen en atribuciones, también los casos legales se multiplicarán, no necesariamente por daños físicos, sino también por otros tipos de daños (a la propiedad, a la privacidad, etc.). En los subapartados siguientes veremos cómo empiezan a hacer notar su presencia en el ámbito jurídico, sobre todo de forma preventiva, esto es, cómo el mundo de las leyes empieza a prepararse para lo que sin duda será uno de los mayores cambios tecno-sociológicos experimentados por las civilizaciones humanas. Por lo que se refiere a la IA incorpórea, al margen de su posible utilización delictiva en acciones de espionaje, de ataque a la propiedad o *hacking* de sistemas cuyo control depende de software, esto es, de acciones punibles que empiezan a aparecer en las salas de juicio, también tiene una presencia positiva en el mundo jurídico: este subapartado está precisamente dedicada a repasar qué herramientas y técnicas de IA pueden ser y son de utilidad en el mundo de las leyes.

El software de asistencia a letrados adopta formas cada vez más sofisticadas y útiles gracias a la IA. Es de destacar el cambio de paradigma que se ha producido en las últimas décadas. En un inicio, fueron los sistemas expertos, con sus representaciones del conocimiento e inferencia basados en reglas, los que conocieron una gran difusión desde los años setenta a los noventa del pasado siglo. Ello respondía al objetivo de «desarrollar modelos formales de la argumentación legal en una forma procesable por ordenador y modelizar computacionalmente la legislación y las reglas legales» (Surden, 2019). El artículo de

Buchanan y Headrick (1970) suele ser referenciado como el primer trabajo serio en explorar las posibilidades de la IA en el mundo jurídico, y apunta las líneas maestras de lo que serían sus objetivos y métodos durante el mencionado período: desde la búsqueda de palabras clave en textos legales, aumentada a través de la búsqueda asociativa (co-ocurrencia de palabras en el mismo texto) hasta los sistemas de razonamiento legal basados en reglas. El artículo es más bien una compilación de *desiderata* y apunta las dificultades en formalizar el razonamiento jurídico, aunque subraya el interés que tiene para, precisamente, verter alguna luz sobre cómo se produce este razonamiento. A partir del cambio de milenio, sin embargo, y siguiendo las tendencias de la IA en general, ha habido un cambio de orientación hacia los procesos basados en aprendizaje automático. En Surden (2019) se apunta a una clasificación del uso de la IA en el mundo jurídico atendiendo a tres tipos de usuarios: los administradores de la ley (legisladores, jueces, oficiales administrativos y policías), los profesionales de la justicia (básicamente abogados) y los afectados o gobernados por la ley (ciudadanos, negocios, empresas y corporaciones). Los jueces, por ejemplo, pueden someter a un sistema de IA la evaluación del riesgo de reincidencia de un procesado para decretar o no su libertad condicional. También la policía puede utilizar técnicas de aprendizaje automático para determinar las zonas más conflictivas de la ciudad e incrementar su presencia en dichas zonas. Otra tecnología de uso cada vez más extendido en el ámbito policial es el reconocimiento facial a partir de imágenes recogidas por cámaras. Por lo que se refiere a usuarios circunstanciales de la ley, se pueden mencionar los sistemas expertos que utilizan algunas compañías para comprobar si determinadas actividades, por ejemplo de importación-exportación, se ajustan a la ley. O los simples sistemas expertos (a menudo en forma de *chatbots* o programas conversacionales) que una persona sin formación legal puede utilizar para resolver sus dudas sobre cuestiones legales básicas (Surden, 2019, pág. 1.335). Pero lo que resulta de mayor interés para este curso son las herramientas que la IA ofrece al segundo de los colectivos mencionados. Quizá lo más destacado sean los sistemas que revisan miles de documentos para detectar aquellos que son relevantes para el caso en curso (antes del juicio): las llamadas técnicas de codificación predictiva, que usan aprendizaje automático y representación del conocimiento para automatizar el proceso de determinar la relevancia jurídica de un documento (para un caso determinado). Un ejemplo que se menciona en Frey y Osborne es el:

«sistema Clearwell de Symantec, que usa análisis del lenguaje para identificar conceptos generales en documentos, que puede presentar los resultados gráficamente, y que ha demostrado ser capaz de analizar y ordenar más de 570.000 documentos en dos días»⁸. (Frey y Osborne, 2013: pág.17)

⁽⁸⁾ Los autores citan a su vez a Markoff (2011).

La IA es realmente eficiente para llevar a cabo estas tediosas tareas, que supondrían una cantidad insostenible de horas-persona, y pueden servir sobre todo para eliminar información irrelevante, pero la decisión final sobre la relevancia de un documento sigue siendo de un abogado humano, puesto que involucra *comprender* la ley y los hechos relacionados con el caso, e implica razones estratégicas y abstracciones que aún están fuera del alcance de la IA. La pre-

dicción de las probabilidades asociadas a los diferentes posibles resultados de un juicio es otra aplicación donde las técnicas de IA pueden ser de ayuda. Esto puede servir al abogado para aconsejar a su cliente acerca de sus posibilidades y si interesa llegar a un acuerdo con la otra parte.

A día de hoy, los sistemas expertos utilizados en el ámbito legal cuentan con una de las cuatro arquitecturas siguientes: sistemas basados en reglas con razonamiento deductivo, modelos de análisis basado en casos (esto es, por analogía con casos similares), redes neuronales y lógica difusa (Wikipedia, «Legal expert system», 2019; Wikipedia, «Legal informatics», 2019, epígrafe «2.2 Artificial Intelligence»). Los retos a los que se enfrentan son la complejidad inherente al derecho, las limitaciones de la resolución de problemas por ordenador, la dificultad de representar el conocimiento legal (por sus imprecisiones, por el hecho de que no siempre se producen los mismos resultados –de un juicio– en condiciones similares, etc.), por las restricciones autoimpuestas en temática jurídica y jurisdicción (esto es, se crean sistemas expertos de aplicabilidad limitada), y posibles interpretaciones erróneas de los resultados del sistema experto al ser interpretados por un lego en la materia (Wikipedia, «Legal expert system», 2019). También es de destacar que su aplicación a un campo amplio pero concreto como son las leyes ha reportado algunos resultados y desarrollos interesantes a la IA más genérica.

3.2. Responsabilidad civil y penal de los actores implicados en el desarrollo del robot inteligente

3.2.1. Accidentes: los robots pueden causar daño

Desgraciadamente, la aún breve historia de la robótica ya cuenta con una página negra de accidentes con víctimas mortales causadas por robots. Ya en 1979 se produjo lo que se considera el primer accidente mortal ocasionado por un robot (Wikipedia, «Robot fatality», 2019), cuando Robert Williams, un trabajador de la planta de moldeo de la Ford Motor Company, en Míchigan, fue golpeado en la cabeza por el brazo mecánico de un robot de una tonelada de peso, cuando estaba retirando manualmente una pieza debido a un aparente funcionamiento defectuoso del robot. La empresa fabricante del robot, Litton Industries (a través de su división Unit Handling Systems), fue condenada en 1983 por un jurado popular al pago de 10 millones de dólares a los herederos de Williams, por negligencia «en el diseño, fabricación y suministro del sistema de almacenamiento y por dejación en el aviso [a los operadores del sistema] de los riesgos del trabajo en el área de almacenamiento». La empresa llegó a un acuerdo con los familiares de Williams por un importe no revelado a cambio de no admitir negligencia. Posteriormente, Litton intentó recuperar las costas del juicio y una indemnización, demandando a Ford por no haber enviado a Williams al entrenamiento ofrecido por Litton y por permitir

la entrada del trabajador en un área de almacenamiento sin activar el sistema de parada, pero el tribunal desestimó la demanda por haber llegado ya a un acuerdo con los herederos de Williams.

En 1981 se produjo el primer accidente mortal en una fábrica de robots, cuando Kenji Urada, un trabajador de la Kawasaki Heavy Industries en la planta de la ciudad japonesa de Akashi, saltó la valla de seguridad para comprobar el funcionamiento erróneo de un robot (Wikipedia, «Kenji Urada», 2019). Activó inadvertidamente el robot, que casualmente aplastó al trabajador contra una máquina rectificadora, causándole la muerte. Los otros trabajadores, que no estaban familiarizados con la operación del robot, fueron incapaces de desconectarlo. La investigación subsiguiente puso de relieve esta falta de formación, así como la falta de regulación de dichas máquinas. El robot involucrado en el accidente fue retirado, y vallas de seguridad de la altura de una persona fueron construidas alrededor de los otros robots de la planta.

Estos dos casos pusieron de relieve en su momento una serie de cuestiones fundamentales sobre los robots industriales:

1) El robot industrial es una máquina potencialmente peligrosa, que puede causar graves heridas e incluso la muerte, debido a su envergadura y las trayectorias complejas que puede seguir el brazo mecánico. Una secuencia mal programada puede originar un movimiento imprevisto por el operador, con el consiguiente riesgo de colisión.

2) La normativa que introducía medidas de seguridad tales como la programación a velocidad reducida y el pulsador de hombre muerto, el vallado de seguridad con detección de intrusos y la parada automática, el pulsador de parada de emergencia en un lugar visible y accesible, la formación adecuada de operarios, etc. se reguló *a posteriori*, una vez que ya se habían producido los accidentes.

3) La responsabilidad legal por los accidentes puede recaer tanto en el fabricante del robot como en la empresa usuaria del mismo. La legislación debe delimitar claramente las atribuciones de cada actor involucrado en materia de seguridad, y la investigación judicial ha de dilucidar el grado de cumplimiento de la normativa por las partes, incluyendo a la eventual víctima.

También se han producido accidentes fatales fuera del ámbito industrial. En 2007, un cañón robótico antiaéreo (semiautónomo) del ejército sudafricano mató por accidente a nueve soldados participantes en el ejercicio de tiro e hirió a otros catorce (Shachtman, 2007), y en 2018 un vehículo Tesla en modo autopiloto colisionó contra la medianera, causando la muerte de su conductor (Brown, 2019), quien aparentemente no había reaccionado a pesar de los avisos del sistema (durante los 6 segundos previos al accidente no tenía las manos puestas en el volante).

Los robots industriales llevan décadas en funcionamiento. Cabe preguntarse si la experiencia acumulada en materia de seguridad es extrapolable a otros ámbitos fuera de las fábricas. Lo es únicamente de forma muy limitada: el robot industrial se ha concebido generalmente hasta el momento como una máquina que realiza tareas peligrosas, aburridas, o en ambientes tóxicos, y que por consiguiente opera de forma autónoma, sin contacto con humanos: todas las medidas de seguridad van encaminadas precisamente a evitar este contacto mientras el robot esté en funcionamiento. No se puede proceder del mismo modo con robots cuyo cometido entraña, precisamente, la proximidad o incluso el contacto humano. De hecho, en la propia industria está cobrando fuerza el concepto del *cobot*, el robot que trabaja de forma colaborativa con el operario humano (véase el subapartado 2.1). Pero de los tres puntos anteriores sí pueden extraerse algunas conclusiones válidas para otros tipos de robots:

1) A diferencia de otras máquinas que pueden causar graves accidentes como prensas o cizalladoras, donde el recorrido de los elementos móviles es siempre el mismo, los brazos robots pueden acceder a cualquier punto de su espacio de trabajo, siguiendo trayectorias que pueden seguir una secuencia programada más o menos compleja, o que varían en respuesta a estímulos externos. En el caso de los robots móviles, el acceso es prácticamente ilimitado. Como no pueden ser confinados, debido a las tareas que se supone que han de realizar, las medidas de seguridad pasan por el diseño de los propios robots, tanto de su constitución física como de su programación y mecanismos de control. El diseño contribuye a la seguridad intrínseca del robot, que es solo un aspecto (aunque sin duda el más importante) de la *ética implícita* de estas máquinas (recuérdese el subapartado 2.1).

2) Los accidentes se producen, pero una normativa desarrollada con la suficiente previsión puede reducir su frecuencia y paliar sus efectos. El daño físico y/o económico no es la única consecuencia negativa que puede ocasionar un robot en su interacción con humanos, ni los accidentes la única manera en que tales efectos perjudiciales se pueden producir. Anticiparse a las posibles contingencias requiere un esfuerzo de imaginación y de sentido común (dos habilidades cognitivas en las que los humanos todavía superamos a las máquinas inteligentes), así como de un conocimiento preciso y actualizado de las posibilidades reales de los robots y de la naturaleza humana. Las leyes y recomendaciones que se deriven de este esfuerzo no serán la última palabra, sino que deberán ser revisadas y actualizadas tanto a tenor de la práctica real y de las eventualidades que se observen como de los avances en movilidad, destreza y habilidades cognitivas que con el tiempo se vayan desarrollando en los robots.

3) Por el momento, y aún durante algún tiempo, la responsabilidad legal por una eventualidad perniciosa generada por un robot recaerá en alguno/s de los múltiples actores humanos involucrados en la interacción con el robot, pero no en la propia máquina, ya que esta no está dotada de los requisitos cognitivos necesarios para hacerse responsable de sus acciones. De nuevo, para los

robots en general esto no concierne solo a contingencias tales como accidentes, sino que se hace extensivo a otro tipo de consecuencias adversas como daños psicológicos o morales. Algunas voces se están dejando oír en el sentido de que es posible atribuir *personalidad legal* a máquinas sin que estas hayan llegado a adquirir atributos tales como una conciencia similar a la humana, ya que de hecho existe esta figura para entidades tales como corporaciones o animales no racionales.

3.2.2. Un marco legal para robots

Prescindiendo de entrada de la naturaleza moral de un robot (esto es, si puede o no considerarse un agente moral), lo cierto es que tanto las acciones que realiza como las que sufre pueden tener una dimensión moral, por los posibles daños y perjuicios ocasionados por dichas acciones sobre personas o propiedades (incluyendo su propia integridad). Las leyes deberán cumplir el cometido de proteger los derechos de las personas que se vean afectados por las acciones de un robot, y dados los mecanismos de que disponen, lo harán de forma mucho más efectiva que las recomendaciones, manifiestos, declaraciones u otros procedimientos que puedan reflejar voluntades normativas o reguladoras de la presencia y difusión de la robótica y la IA en la sociedad. Cabría preguntarse si el marco legal actual ya es suficiente, de modo que la jurisprudencia que se genere bastará para afinar la especificidad asociada a estas tecnologías, basándonos en el buen criterio y sentido común de los jueces que se vayan enfrentando a los casos que se produzcan. La respuesta es afirmativa solo para un reducido número de casos. Por consiguiente, el Parlamento Europeo emitió, en febrero de 2017, una resolución sobre reglas de derecho civil en robótica, que comentaremos en el subapartado 3.3, a raíz de las conclusiones del proyecto europeo RoboLaw (2014) (véase el subapartado 3.3.1). Dicha resolución contempla las particularidades del fenómeno robótico y de la IA, y sirve de marco de referencia común para las iniciativas legislativas de los Estados miembros. Antes de proceder a comentar los aspectos más destacados de esta resolución, así como de otras iniciativas legales en el ámbito extracomunitario, es de recibo repasar algunos conceptos teóricos, básicamente sobre responsabilidad civil y penal, y cómo se aplican al caso del robot en diferentes fases de desarrollo cognitivo. Hemos seguido en su mayor parte el texto esclarecedor de Peter Asaro (2014), que, aunque parte de la tradición jurídica estadounidense, los conceptos, principios y razonamientos que enuncia son de carácter más general.

Robots no autónomos o semiautónomos

En los casos más simples, donde el robot se limita a ejecutar un programa preestablecido de fábrica, los casos que se presenten quedarán cubiertos por la legislación civil y penal de la responsabilidad por el producto⁹ (Asaro, 2014). El robot es tratado como cualquier otro producto manufacturado, tanto en lo que se refiere a los daños causados por sus acciones como los daños (incluyendo la sustracción) que terceros inflijan sobre la máquina (lo que se consi-

⁹O responsabilidad sobre el producto, o responsabilidad por defectos del producto.

deraría como daños sobre la propiedad). Concentrándonos en el primer tipo, el daño causado por un robot puede deberse a una concepción y diseño descuidado, una fabricación defectuosa, una programación errónea o un uso no adecuado del mismo. En toda esta cadena, que involucra a científicos/tecnólogos/diseñadores, fabricantes, programadores, distribuidores y usuarios, no se ha supuesto una intencionalidad criminal; de otro modo habría que incluir también a saboteadores y *hackers*, y estaríamos hablando de responsabilidad penal por los daños causados. Circunscribiéndonos a la responsabilidad civil, asociada a posibles daños causados de forma no intencionada, el objetivo es determinar quién debe compensar a la víctima (los tribunales han de decidir, además, la cuantía de esta compensación). Para simplificar, consideraremos únicamente dos actores, el fabricante y el usuario, donde el primero concibe y fabrica el robot, y el segundo lo utiliza (por lo que se refiere a la programación, cabe distinguir entre la programación de base que viene de fábrica, y la configuración particular del usuario, o las modificaciones que este haga sobre la programación original). El fabricante debe responder por dos posibles tipos de fallos asociados a negligencias en el diseño y la fabricación:

- fallo en avisar (al usuario, sobre riesgos previsibles asociados al uso del producto), y
- fallo en prestar la atención y dedicación suficientes en determinar los potenciales riesgos de su producto.

Determinar todos los riesgos asociados a un producto no es fácil, sobre todo cuando se trata de un producto novedoso, complejo y en rápida evolución como el robot. Los fabricantes a menudo se escudan (o toman como referencia) lo que se considera la práctica *normal* del sector (sea una normalización explícita como las normas ANSI o ISO, o la práctica implícita). En el caso de los robots no existe aún la suficiente experiencia como para consolidar lo que podría considerarse como práctica habitual, y ese podría ser precisamente uno de los cometidos más útiles de la roboética: imaginar los posibles riesgos y desarrollar medidas precautorias. Si la negligencia del fabricante es fruto de una manifiesta actitud de eludir cualquier estudio de los potenciales riesgos de su producto, entonces puede ser calificada de criminal: es un caso de *mens rea*, donde

«el estado de culpabilidad de la mente es uno de ignorancia, sea voluntaria o sea por poco razonable» (Asaro, 2014, pág. 172).

Estadios aún más avanzados en culpabilidad son los de temeridad o imprudencia (previsión de posibles riesgos o daños, aunque sin la intención de que alguna víctima potencial sea dañada), y los de conciencia de la existencia de un riesgo real (de nuevo sin la intención de causar daño). A pesar de todo, estos casos suelen resolverse por la vía civil. Es de señalar también que la responsabilidad causal no siempre coincide con la legal, esforzándose las legislaciones en que se cumpla el objetivo de compensación de la víctima por el daño causado, sea a través de la repartición de responsabilidades, sea por atribuir res-

ponsabilidad estricta al actor que se encuentra en condiciones más favorables de reparar a la víctima. En el caso del robot, donde confluyen diversos componentes (posiblemente de diferentes fabricantes), la casuística de posibles interacciones entre los componentes puede ser muy compleja. La construcción de otros sistemas de elevada complejidad, como las aeronaves, puede servir de inspiración para el caso del robot.

La responsabilidad del diseñador/ingeniero/fabricante acaba donde empieza la del usuario; generalmente, viene determinada por un uso indebido del producto. De hecho, con determinados productos peligrosos, como vehículos de motor o armas, la responsabilidad recae totalmente en el usuario, quedando la del fabricante restringida a posibles defectos de fabricación. Un robot, dependiendo de sus características, podría entrar en esta categoría. No es realista que un fabricante pueda anticipar todos los posibles usos de su producto, y como regla general puede valer que cuanto menos previsible sea cierto uso del producto, menor será la responsabilidad del fabricante (y mayor la del usuario). Como uso en el caso del robot puede interpretarse tanto las tareas a que se destina (y que, a su vez, dependen de cómo se haya programado el robot) como el contexto o el entorno en que se desarrolla su actividad, que puede ser totalmente inapropiado. Un usuario puede ignorar la respuesta de su robot en una situación determinada, lo que puede en cierto modo ser un atenuante si se produce algún daño. Pero si posteriormente se produce un daño en una tesitura similar, el usuario se encontrará en situación de *conocimiento doloso* (White y Baum, 2017). Será muy difícil evitar que usuarios avezados en programación o manitas del bricolaje tuneen sus robots, aunque es factible someterlos a una revisión periódica, una suerte de ITR (inspección técnica de robots) donde se compruebe que las modificaciones se ajustan a la legalidad. Atribuir estricta responsabilidad legal al fabricante, por otro lado, puede suponer una notable traba a la innovación y desarrollo, a la par que limitar la predisposición del mismo a que su software sea de código abierto, lo que a su vez pone obstáculos al desarrollo de aplicaciones por parte de la comunidad de usuarios que podrían redundar en beneficio para todos. Conforme avance la capacidad de los robots de procesar el lenguaje natural, por otro lado, la interacción entre robots y usuarios será cada vez más verbal (oral), lo que abrirá las puertas a todos los equívocos e imprecisiones del lenguaje humano, y las posibles interpretaciones erróneas de los comandos. En algunos casos, será complejo dilucidar si el comportamiento erróneo del robot se debe a un fallo en el software de interpretación del lenguaje o a que el usuario no se ha expresado con la suficiente claridad.

Robots autónomos

El desarrollo de las técnicas de inteligencia artificial permitirá, en algún momento, dotar a los robots de cierta capacidad de razonamiento. Serán capaces de determinar un curso de acción y de elegir una entre varias opciones para cumplir con objetivos formulados a más alto nivel, como «Mantén la casa limpia» o «Cuida del abuelo Pancratius». El grado de sofisticación del mecanismo

de toma de decisiones puede llegar a un extremo en que resulte muy complejo, por no decir imposible, atribuir la responsabilidad causal de una determinada acción con resultados dañinos al fabricante o al usuario, y nos veamos tentados de hacer responsable al propio robot de sus actos. No obstante, no se debe olvidar que en última instancia es un ser humano (o un grupo de seres humanos) quien decidió desarrollar tal robot, y que también es un ser humano quien decidió adquirirlo y usarlo. El robot en cuestión no es un ser independiente, sino que tiene un propietario. Es posible que dicho robot haya tomado una decisión de forma autónoma, en tanto puede reconocérsele capacidad de razonamiento, que dicha decisión no es fruto de una falta negligente de previsión por parte del fabricante, ni haya intencionalidad dolosa por parte del usuario en el momento de ordenarle sus quehaceres. No obstante, incluso admitiendo una responsabilidad causal en el robot, no necesariamente deberá considerarse que incurre en una responsabilidad moral o legal. En el contexto de los robots militares, Lokhorst y Van den Hoven (2014, pág. 151) analizan la responsabilidad moral de un robot que cumple órdenes de un superior en los dos casos de un robot no deliberativo y uno deliberativo. Para ello, aplican el formalismo lógico del STIT (*seeing to it that*, «procurar que»; un agente B procura que [se produzca, ocurra, se cumpla] P [una proposición correspondiente al efecto de cierta acción]) y la variante del DSTIT (*deliberatively seeing to it that*, donde «deliberativamente» supone que también se podría optar por «no P», es decir, que hay capacidad de escoger). En el caso de un robot no deliberativo (es decir, en el caso expuesto en el subapartado 3.2.2.1), se aplicarían las tradicionales máximas legales de *qui facit per alium facit per se* (esto es, que actuar a través de otros, subordinados de algún modo, equivale a actuar uno mismo) y de *respondeat superior* (el amo [ha de] responde[r]). En el caso de robots deliberativos, los autores consideran los casos de «elección forzada» y, por tanto, de «responsabilidad moral forzada», y establecen paralelismos con soldados que, si bien no son moralmente responsables de cuestionar la legalidad de la movilización, sí pueden serlo sobre la decisión de cumplir o no órdenes específicas que por ejemplo contravengan el DIH. El conflicto de esta responsabilidad con la obediencia debida fue objeto de los famosos juicios de Núremberg, quedando de manifiesto que la segunda no era eximente sobre la primera. No obstante, precisamente en el contexto bélico, los autores consideran cuestionable que se lleguen a incorporar robots que en algún momento puedan decantarse por la insubordinación.

En un contexto menos específico, Asaro (2014) también considera la responsabilidad civil subsidiaria de los propietarios de robots, estableciendo paralelismos con los propietarios de animales domésticos y de animales salvajes, así como empleador y trabajador, oficial y soldado, e incluso maestro y esclavo. En el caso del animal doméstico, que el autor considera comparable al de un robot con capacidades deliberativas, pero con cierta experiencia en cuanto a su utilización, con cierto rodaje en sociedad, su comportamiento es bastante previsible en principio. Si en algún momento causa un daño (actuando autónomamente), el propietario de animal se considerará responsable civil, y si el animal actúa agresivamente porque el propietario ha tenido una actitud

criminallymente negligente, temeraria o incluso dolosa, este será responsable penalmente del daño causado. Las intenciones del animal no se consideran relevantes, sí en cambio las del propietario, aunque también es cierto que un comportamiento inusualmente errático o inesperado del animal tiende a disminuir la *mens rea* del propietario (aunque no lo exime de su responsabilidad). En el caso de tenencia de animales salvajes, si estos pueden comportar cierto peligro para las personas, la responsabilidad legal del propietario es mayor, y si este no es capaz de mantenerlos en un recinto controlado (es decir, si se le escapan), puede incurrir automáticamente en una temeridad criminal. El paralelismo, en el caso de la robótica, sería con robots que por sus características físicas y por su índole experimental o altamente modificada pueden suponer un peligro potencial para el vecindario. El autor pone el ejemplo del propietario que decide dotar a tal robot de un arma: aunque el robot no llegue a herir a nadie ni a destruir propiedad alguna, se considerará que el propietario incurre en temeridad criminal, del mismo modo que un conductor en estado de embriaguez que se pone al volante de un vehículo, aunque no cause ningún accidente.

La propiedad o el uso de determinados robots pueden someterse a un sistema de licencias, de modo análogo al carnet de conducir que es necesario conseguir para conducir un coche o una licencia de armas o de manejo de explosivos (con los consiguientes requerimientos de edad, estabilidad psicológica y conocimientos técnicos).

El grado de autonomía de un robot podría llegar al extremo de que este tenga objetivos propios. En este caso, podrían considerarse paralelismos con el binomio empresario-empleado. Asaro cita un esclarecedor ejemplo sobre el empleado de una empresa de transportes: si el conductor se desvía de la ruta prevista y causa un accidente, la responsabilidad recae en el empresario si el desvío ha sido forzado (por unas obras, por ejemplo), y en cambio recae en el conductor si ha sido por atender a asuntos personales (para visitar un amigo de camino, pongamos por caso). Esta dicotomía *detour* frente a *frollicking* podría aplicarse también al robot por lo que se refiere a la atribución de responsabilidades, en el hipotético caso en que los robots lleguen a tener asuntos personales que atender...

¿Podrían los robots tener responsabilidad penal aun sin ser agentes morales plenos? A diferencia de la responsabilidad civil, en la responsabilidad penal la intención sí es relevante. De nuevo, Asaro establece paralelismos que pueden verter alguna luz sobre cómo tratar jurídicamente cierto tipo de robots avanzados, esta vez con niños y personas con trastornos psíquicos, que legalmente pueden considerarse cuasi-personas por lo que se refiere a sus responsabilidades (aunque sean personas de pleno derecho por lo que se refiere a eventuales crímenes sobre ellos; de hecho, gozan de una protección especial en este sentido). Es más, incluso un adulto psicológicamente estable puede alegar un trastorno mental transitorio para beneficiarse de atenuantes o eximentes en

Frollicking

Detour se traduce como 'desvío', mientras que *frollicking* tiene connotaciones lúdicas, de jugueteo.

un acto punible que haya cometido. En este sentido, un robot dotado de moralidad funcional (véase el subapartado 2.1) podría considerarse legalmente una cuasi-persona, en tanto no haya adquirido personalidad jurídica plena.

Recordemos también que, a diferencia de la responsabilidad civil, a través de la cual se pretende compensar (en general, económicamente) a la víctima, la responsabilidad penal conlleva el **castigo** del infractor. La finalidad del castigo es triple: pagar una deuda con la sociedad (retribución), modificar el comportamiento del infractor para que no vuelva a cometer un hecho punible (reforma), y buscar la ejemplaridad para que otros potenciales infractores se abstengan de cometer el acto punible (efecto disuasorio). Asaro cuestiona que se pueda castigar a un robot precisamente porque, mientras no sea un agente moral pleno, ninguno de estos tres factores es realmente significativo para la máquina. Existen otras entidades, las corporaciones, que sí gozan de personalidad legal y que también pueden ser castigadas, tanto a través de la multa, que incide precisamente sobre su razón de ser (la maximización del beneficio económico), como a través de la persecución legal de los individuos que las integran. Las corporaciones son incorpóreas (a pesar de la paradoja etimológica), pero son capaces de *sufrir* un castigo. Los robots, por otro lado, aun disponiendo de un cuerpo, son incapaces de sufrir, a no ser que en algún momento el desarrollo de una conciencia artificial conlleve también la capacidad de sufrimiento (recordemos también que los animales no humanos superiores pueden sufrir, incluso no siendo conscientes en el sentido humano, por su capacidad de experimentar dolor). Pero ¿es legítimo crear esta capacidad de sufrimiento? Lockhorst y Van den Hoven (2014, pág. 149) no lo creen así, ya que para ellos el castigo no es una finalidad en sí mismo. De los tres motivos asociados al castigo, el de la reforma es sin duda el más importante, y si este fin puede ser conseguido por otros medios (lo cual es ciertamente más factible en el caso de los robots, como ajustar su programación), sin duda serán preferibles al castigo. El sufrimiento también puede interpretarse de una forma más funcional, sin entrar en la experiencia subjetiva del mismo. En Torras (2016b) se discute esta funcionalidad, en el contexto del aprendizaje por refuerzo (modalidad de programación que se describe en el subapartado 1.4).

3.3. Iniciativas legales en el ámbito europeo

3.3.1. Precedentes

Las visiones éticas y las costumbres que constituyen la moral aceptada en una sociedad suelen preceder a su formulación precisa y redacción en forma de leyes. En el caso de la robótica y la IA sucede lo mismo, con la salvedad de que muchos desarrollos asociados a esta tecnología aún se encuentran más en el terreno de lo factible o previsible que en el de lo real. No existe todavía una experiencia consolidada en el uso de los robots que pueda inspirar al legislador, pero la transformación social que conlleva esta tecnología se prevé de tal calado que muchos expertos recomiendan anticiparse a los hechos. Aquí se

impone el sentido común, los escenarios que anticipa cierta ciencia-ficción de calidad y las analogías con otras tecnologías transformadoras como el ordenador personal, internet o la telefonía móvil (que de hecho también están en plena evolución).

El 1 de marzo de 2012 tuvo lugar el comienzo del proyecto del Séptimo Programa Marco de la Unión Europea (Ciencia en Sociedad) titulado RoboLaw, acrónimo de Regulating Emerging Robotic Technologies in Europe: Robotics facing Law and Ethics. Su objetivo principal era la elaboración de un documento sobre las «Líneas maestras en la Regulación de la Robótica», «conteniendo sugerencias regulatorias para la Comisión Europea». Los investigadores del proyecto optaron por un planteamiento de estudio por casos, tanto por la imposibilidad de disociar la tecnología robótica utilizada de una aplicación específica como por la imposibilidad de ceñirse a una tecnología concreta, ya que en una aplicación concurren varias de ellas. Los casos estudiados (por su interés según una encuesta popular) comprendían los vehículos autónomos, los robots cirujanos, las prótesis robóticas y los robots asistenciales. El proyecto traspasó el ámbito europeo, ya que el mismo año de su inicio crearon una iniciativa conjunta sobre robots, leyes y política con el Instituto del Derecho en Internet de la Universidad de Pekín (China), a la que se sumó el Instituto de Robótica Humanoide de la Universidad de Waseda (Japón) (Robolaw Asia, 2012). El proyecto concluyó en mayo de 2014, y sus conclusiones fueron presentadas el 24 de septiembre en el Parlamento Europeo, en el taller «Upcoming issues of EU law» (organizado por el Comité JURI de asuntos legales. El artículo presentado llevaba por título «Regulando la robótica: un desafío para Europa» (Robolaw, 2014).

RoboLaw

<http://www.robolaw.eu/index.htm>

3.3.2. Iniciativa de legislación del Parlamento Europeo

La Resolución del Parlamento Europeo del 16 de febrero de 2017, con referencia P8_TA(2017)0051, lleva por título «Reglas de Derecho Civil sobre la Robótica» (Civil Law Rules on Robotics) y pretende, precisamente, establecer las bases de una futura legislación. Emanada de una serie de principios éticos, entre los que cabe destacar los siguientes (de forma resumida):

- Atender los riesgos que esta tecnología plantea sobre la seguridad humana, la salud, la libertad, la privacidad, la integridad, la no discriminación por razón de género, etnia, creencias religiosas, ideología política, etc.
- El marco legal de la UE ha de estar actualizado y complementado por principios éticos acorde a la complejidad de los robots inteligentes.
- Se anexa una carta con el código de conducta para ingenieros de robots y comités de ética, así como modelos de licencia para diseñadores y para usuarios (más abajo se dan algunas indicaciones sobre qué contemplan estos códigos y licencias).
- Se ha de garantizar la transparencia de las decisiones y acciones del robot, y se sugiere recurrir a mecanismos como la caja negra de los aviones que registren lecturas de los sensores, procesos internos y acciones ejecutadas.
- Se propone la creación de una Agencia Europea para la Robótica y la IA.

Otros puntos de interés incluyen las siguientes consideraciones:

- Conforme crezca la autonomía de los robots, cada vez menos se podrán considerar simples herramientas.
- A día de hoy, no obstante, la responsabilidad sobre los actos del robot puede trazarse hasta uno de los actores humanos.
- El marco legal actual es insuficiente debido a cierta impredecibilidad del comportamiento del robot.
- Se propone la creación de un sistema de registro a escala de la UE.
- Hay que focalizar en complementar y no en sustituir a los humanos, en mantener el control y vigilar la creación de vínculos emocionales.
- Se debe favorecer por diseño la investigación, la interconectividad requerida, la interoperabilidad, la seguridad y la privacidad.
- Libertad de flujo de datos preservando la protección de datos personales.
- Estandarización y seguridad del tipo ISO/TC 299 Robotics.
- Las pruebas que se hagan en escenarios reales deben seguir el principio de precaución.

Las recomendaciones dirigidas tanto a profesionales (ingenieros, diseñadores, comités de ética) como al público general (consumidores y usuarios de robots) revisten un interés especial, puesto que pueden inspirar las bases de futuras legislaciones específicas, y de entrada pueden tener la fuerza normativa de un código deontológico. Es de remarcar el carácter práctico de algunas de las recomendaciones:

1) Código de conducta ética para ingenieros robóticos.

- Consideración de los derechos fundamentales.
- Observar el código de conducta es un acto voluntario.
- Su elaboración requiere un esfuerzo multidisciplinario.
- Consideración de implicaciones futuras de la tecnología (no limitarse a los efectos inmediatos).
- Las entidades de financiación de proyectos deben pedir los correspondientes informes de seguridad.
- Los humanos son los responsables de los actos de los robots, no los propios robots.
- Los investigadores deben adherirse a los principios de beneficencia, no maleficencia, autonomía y justicia.
- Otros principios o factores a considerar: precaución, inclusividad, responsabilidad, seguridad, reversibilidad, privacidad, y maximizar beneficios (y minimizar daños).

2) Licencia para diseñadores

- Consideración de los valores europeos fundamentales.
- Introducción de principios de diseño fiables para la seguridad.
- Asegurar la privacidad por diseño.

- Presencia visible de botones de desconexión.
- Operación del robot de acuerdo con principios éticos locales e internacionales.
- Las decisiones del robot han de poder ser reconstruidas.
- Máxima transparencia y predictibilidad, aun considerando la incertidumbre.
- Asegurar que los robots son identificables como tales.
- Obtener una valoración positiva de un comité de ética en la investigación antes de probar un robot en un entorno real.

3) Comités de ética en la investigación (REC, por *research ethics committee*)

- Han de ajustarse a los principios de independencia, competencia, transparencia y responsabilidad.
- Se ha de establecer cuál es la función del REC.
- Se han de proveer los mecanismos de su constitución y de su labor de monitorización.

4) Licencia para usuarios

- Un usuario ha de poder hacer uso de un robot sin riesgo o miedo a sufrir daño físico o psicológico.
- Tiene derecho a esperar que el robot realice cualquier tarea para la que ha sido explícitamente diseñado.
- Ha de tener consciencia de las limitaciones perceptuales, cognitivas y de actuación del robot.
- Ha de considerar el respeto debido a la fragilidad humana, física y psicológica, así como a las necesidades emocionales.
- Ha de tener en cuenta y respetar los derechos a la privacidad.
- Prohibición de usar el robot de modo que contravenga principios éticos o legales.
- Prohibición de modificar un robot para utilizarlo como arma.

3.4. Iniciativas normativas y legales en el ámbito internacional y extracomunitario

3.4.1. Normas ISO/TC 299 Robotics

Los organismos de estandarización, más ágiles o, si se prefiere, con una inercia menor que el aparato legislativo, ya están emitiendo normativa concerniente a las nuevas tecnologías de la IA y la robótica. En este subapartado nos centraremos en el trabajo de la ISO, la Organización Internacional de Estandarización (International Organization for Standardization), una ONG que agrupa organizaciones de estandarización de sus 163 Estados miembros, y cuyo objetivo es que los productos y servicios que reciben su acreditación puedan garantizar su seguridad, fiabilidad y buena calidad. A falta de una legislación es-

pecífica, el cumplimiento de estas normativas por parte de una empresa puede tener cierto peso jurídico, llegado el caso de un juicio por responsabilidad civil. Aquí, en particular, resumiremos el trabajo del comité técnico (TC, por *technical committee*) ISO/TC 299 Robotics.

Es natural que, hasta la fecha, el mayor esfuerzo normativo se haya concentrado en los robots manipuladores industriales: ya en 1988 se emitió un TR (*technical report*) definiendo algunos términos (ISO/TR 8373:1988 Manipulating Industrial Robots – Vocabulary), y en 1990 (ISO 9283:1990) se especificaban los criterios de rendimiento de los robots manipuladores industriales y los métodos de comprobación asociados. Las normas afectan a aspectos genéricos como la seguridad (ISO 10218-1:2011, Robots and Robotic Devices – Safety Requirements for Industrial Robots – Part 1: Robots, e ISO 10218-2:2011, Robots and Robotic Devices – Safety Requirements for Industrial Robots – Part 2: Robot Systems and Integration) o la nomenclatura (además de los mencionados al principio, también ISO 9787:2013 Robots and Robotic Devices – Coordinate Systems and Motion Nomenclatures, o ISO 9946:1999 Manipulating Industrial Robots – Presentation of Characteristics). También pueden hacer referencia a dispositivos concretos, que se refieren básicamente a los elementos o efectores terminales y su intercambiabilidad (ISO 9409-1:2004 Manipulating Industrial Robots – Mechanical Interfaces – Part 1: Plates e ISO 9409-2:2004 Idem, Part 2: Shafts, ISO 14539:2000 Manipulating Industrial Robots – Object Handling with Grasp-type Grippers – Vocabulary and Presentation of Characteristics, o ISO/TR 20218-1:2018 Robotics – Safety Design for Industrial Robot Systems – Part 1: End-effectors) o a elementos usuales en las células robotizadas (ISO/TR 20218-2:2017 Robotics – Safety Design for Industrial Robot Systems – Part 2: Manual Load/unload Stations). Es de destacar, por la proactividad demostrada en adaptarse a las nuevas tendencias, la existencia del estándar ISO/TS 15066:2016 Robots and Robotic Devices – Collaborative Robots, relativo a las normas de seguridad para robots que (en un entorno industrial) trabajan en colaboración con personas.

Fuera del ámbito industrial, como empieza a haber robots móviles y de servicios que se ofrecen como productos comerciales, la ISO también tiene algo que decir al respecto. La primera norma en ver la luz fue la ISO 13482:2014 Robots and Robotic Devices – Safety Requirements for Personal Care Robots¹⁰, que se centra en el robot móvil sirviente, el robot de asistencia física y el robot de transporte personal, en los peligros potenciales asociados a ellos y cómo manejar dichos riesgos, a la vez que descarta de su consideración otros robots como los industriales, los militares, los que se desplazan por aire o por agua (o por tierra a más de 20 km/h), médicos o juguetes.

La seguridad es la principal preocupación de las normas, y en el caso de los robots de asistencia personales queda reflejado en la Norma ISO 13482:2014 Robots and Robotic Devices – Safety Requirements for Personal Care Robots, cuyos aspectos de aplicación se están revisando en ISO/PRF TR 23482-1 Robotics – Application of ISO 13482 – Part 1: Safety-related Test Methods, e ISO/

Normativas TR, ISO/TR 8373:1988, ISO 9283:1990

Estas normas ya han sido retiradas y sustituidas por otras; de hecho, fueron emitidas por un TC diferente y más antiguo, el ISO/TC 184 Sistemas de Automatización e Integración. Las versiones actualmente vigentes, emitidas por el TC 299, son la ISO 8373:2012 Robots and Robotic Devices, y la ISO 9283:1998 Manipulating Industrial Robots – Performance Criteria and Related Test Methods, respectivamente.

Enlace de interés

Véase también el artículo al respecto de María Lazarte, «Robots and humans can work together with new ISO guidance», 8 de marzo de 2016, <<https://www.iso.org/news/2016/03/Ref2057.html>>.

⁽¹⁰⁾Véase el resumen en <https://www.iso.org/standard/53820.html>, y los enlaces, en esta misma página, a artículos de divulgación de la misma entidad sobre esta normativa.

⁽¹¹⁾CD: Committee Draft.

TR 23482-2:2019 Idem, Part 2: Application Guidelines. Las Normas ISO 18646 Robotics – Performance Criteria and Related Test Methods for Service Robots regulan diversos aspectos de los robots de servicios: ISO 18646-1:2016 Part 1: Locomotion for Wheeled Robots, ISO 18646-2:2019 Part 2: Navigation, ISO/CD¹¹ 18646-3 Part 3: Manipulation, e ISO/CD 18646-4 Part 4: Lumbar Support Robots. También la nomenclatura y la normalización son aspectos que requieren una estandarización: ISO 19649:2017 Mobile Robots – Vocabulary, e ISO/DIS 22166-1 Robotics – Modularity for Service Robots – Part 1: General Requirements.

Por último (por ahora), también la robótica médica tiene su lugar en las especificaciones del comité: IEC/TR 60601-4-1:2017 Medical Electrical Equipment – Part 4-1: Guidance and Interpretation – Medical Electrical Equipment and Medical Electrical Systems Employing a Degree of Autonomy, IEC 80601-2-77:2019 Medical Electrical Equipment – Part 2-77: Particular Requirements for the Basic Safety and Essential Performance of Robotically Assisted Surgical Equipment, e IEC 80601-2-78:2019 Medical Electrical Equipment – Part 2-78: Particular Requirements for Basic Safety and Essential Performance of Medical Robots for Rehabilitation, Assessment, Compensation or Alleviation. Toda la actividad del ISO/TC 299 se halla recogida en la página <https://www.iso.org/committee/5915511/x/catalogue/>, y es de suponer que en años venideros experimentará numerosas ampliaciones y actualizaciones.

IEC...

es el acrónimo de International Electrotechnical Commission, entidad que colabora con la ISO.

3.4.2. Ámbito extracomunitario

En 2017, saltó a los titulares de noticias en todo el mundo: Riad había concedido el derecho de ciudadanía a la robot humanoide Sophia. La noticia sorprendía por dos motivos: por la irregularidad del procedimiento (la nacionalidad árabe solo se adquiere por nacimiento en una familia nativa o por matrimonio con un nativo; Sophia tampoco cumplía los requisitos de edad, permanencia o fluidez idiomática), y por garantizar a Sophia una autonomía negada a las mujeres humanas. El caso es analizado en Atabekov y Yastrebov (2018) junto con otro caso, el de permiso de residencia otorgado al *chatbot* Shibuya Mirai en otro país refractario a la admisión de nuevos residentes, Japón.

Al margen de la anécdota, que ilustra la arbitrariedad de los gobernantes en ausencia de legislaciones claras, se vislumbran algunos esfuerzos en países tecnológicamente más avanzados por empezar a desarrollar leyes que regulen la IA y la robótica:

- **Estados Unidos.** En un país de consolidada tradición jurisprudencial, el investigador Ryan Calo (2016), profesor en la University of Washington School of Law, sostiene, después de examinar casos involucrando robots con presencia en las salas de justicia desde los años cuarenta, que la política legislativa sobre robots no vendrá solo de la jurisprudencia, sino también del esfuerzo legislativo de instituciones como la National Highway Traffic Safety Administration y la Federal Aviation Association. De hecho,

ambas instituciones ya emitieron una regulación en el ámbito federal en 2016 en sus respectivos campos (vehículos autónomos y UAV). Es decir, según Calo, la regulación se irá construyendo *ad hoc*, pero con ciertas pretensiones de generalidad en áreas específicas.

- **Reino Unido.** Al margen de las iniciativas de la UE, un comité de la Casa de los Comunes británica publica en 2016 el informe «Robots e IA», con recomendaciones al gobierno para establecer una comisión sobre IA (Wright y Bledsoe, 2017).
- **Japón.** La Oficina Central para la Revitalización Económica del Japón, del METI (el Ministerio Japonés de Economía, Comercio e Industria), emitió en febrero de 2015 la «Nueva Estrategia de los Robots. La estrategia robótica del Japón. Visión, Estrategia, Plan de Acción». La principal preocupación en el ámbito de la ética se refiere a los robots asistenciales. En este informe también se establece la Iniciativa de la Revolución Robótica, una plataforma abierta de innovación dirigida por la iniciativa privada. El texto proporciona recomendaciones, sin fuerza legal.
- **China.** En 2017, el Consejo de Estado emitió el «Plan de Desarrollo de la Nueva Generación de IA», donde se establece el objetivo de convertir China en la primera potencia mundial en IA en 2030 (Karch *et al.*, 2018). Se definen unas «tareas focales» de aplicación de la IA a retos sociales, económicos y de seguridad nacional. También incluye una serie de «medidas de garantía» para guiar «el desarrollo y aplicación de la IA, con leyes y regulaciones, marcos éticos y distribución de recursos». Basado en este Plan, el Ministerio de Industria y Tecnología de la Información estableció, en diciembre de 2017, un «Plan de acción Trienal para Promover el Desarrollo de una Nueva Generación de Industria de la IA», con especial incidencia en «vehículos autónomos, robots inteligentes de servicio, UAV inteligentes, sistemas de diagnóstico de imagen médica, sistemas de identificación de imagen y vídeo, sistemas interactivos inteligentes de voz, sistemas inteligentes de traducción, y productos domóticos».
- **Corea del sur.** En 2007, el Ministerio Coreano de Comercio, Industria y Energía anuncia la creación de una Carta de Ética en Robótica. Sin embargo, en fechas más recientes la tendencia es más bien a la desregularización para facilitar la promoción de la tecnología y la industria robótica. Por ejemplo, se han relajado los requisitos de seguridad al mínimo necesario para favorecer la expansión de robots colaborativos trabajando con humanos en las fábricas. Los esfuerzos se dirigen a «identificar y mejorar las regulaciones que obstaculizan la popularización del robot y el desarrollo de la tecnología en varios campos» (Sang-mo, 2018).
- **Rusia.** Desde 2016, y con vistas a su aprobación en 2022, la Duma está discutiendo un borrador de leyes para regular las relaciones entre humanos y robots. Conocida por la «ley Grishin», por su proponente Dmitry

«Nueva Estrategia de los Robots. La estrategia robótica del Japón. Visión, Estrategia, Plan de Acción»

http://www.meti.go.jp/english/press/2015/pdf/0123_01b.pdf

Iniciativa de la Revolución Robótica

<https://www.jmfri.gr.jp/english/outline/763.html>

Grishin, el borrador contempla los casos en que un robot o sus acciones pueden considerarse un crimen: la creación de un robot asesino con esta finalidad, la desactivación de los elementos de hardware y software que bloquean la posibilidad de causar daño a seres humanos, la construcción de un robot capaz de causar daño, y la de un robot incapaz de *comprender* que puede ser usado para causar daño (Goldman, 2017). También se propone crear un registro de robots.

3.5. Derechos del robot: visiones de futuro

El derecho sin duda se verá afectado de muchas maneras por la existencia, desarrollo y difusión de los robots y de la IA. En el subapartado 3.2 planteamos algunas cuestiones sobre responsabilidad civil y penal de unos hipotéticos robots que han evolucionado hasta tener objetivos propios. Si seguimos instalados en esta hipótesis, podemos preguntarnos legítimamente si el reconocimiento de responsabilidad trae consigo y como contrapartida el reconocimiento de derechos. El reconocimiento de derechos de un sujeto (o conjunto de sujetos) puede iniciarse de dos maneras: el sujeto en cuestión los reclama para sí (por ejemplo, las mujeres a través del movimiento sufragista a mediados del siglo XIX o la comunidad afroamericana estadounidense en los años cincuenta del pasado siglo), o bien son reclamados por quien puede hacerlo (por ejemplo, quienes abogan por los derechos de los animales). La primera vía, para hacerse efectiva, parece implicar alguna forma de verbalizar un deseo, una aspiración, y para prosperar requiere que quien puede otorgar los derechos reclamados admita que el colectivo en cuestión está cualificado para su disfrute. Ello supuso, en su momento, que cierto colectivo de hombres blancos admitieran que no había diferencias esenciales entre ellos y el colectivo de mujeres o de las personas de color, y que la igualdad de derechos estaba justificada. En los casos que requieren la segunda vía, generalmente no se pretende una equiparación de derechos, pero sí garantizar una serie de derechos básicos, como el derecho a la vida o a no ser torturado. Hoy por hoy, los robots son máquinas que distan aún mucho de poder considerarse para el reconocimiento siquiera de derechos semejantes a los de algunas categorías de animales. Tienen el mismo derecho a no ser desguazados que una lavadora. La verbalización no supone, a diferencia de los animales, ningún problema para los robots: pueden sintetizar perfectamente la frase «Deseo que se me reconozca el derecho a no ser desguazado». Otra cosa diferente es qué hay detrás de la primera palabra de esta frase. A diferencia de los animales, de los robots no puede afirmarse (¿aún?, ¿nunca?) que sean sintientes. No hace falta llegar a una conciencia del mismo nivel que la humana: para ser candidato al derecho del no-desguace, probablemente, bastaría con «sentir, percibir o experimentar subjetivamente» (Wikipedia, «Sintiencia», 2019). ¿Es la sintiencia una característica exclusiva de los seres biológicos? ¿De los seres cuyo organismo está basado en la química del carbono? Sin embargo, tampoco puede obviarse que, detrás de la frase en cuestión, puede que no haya un deseo (subjetivo o no) del

robot que la pronuncia, sino de la empresa que lo fabricó o de su propietario como vía para expandir su propio poder en el seno de la sociedad, a través de sus robots.

Parece que es aún prematuro para plantearse estas cuestiones. No obstante, la controversia y el debate están ahí. La profesora Joanna Bryson, de la Universidad de Bath, sostiene que, aunque los robots puedan llegar a (calificarse para) tener derechos, no deberían tenerlos: «Los robots deberían ser esclavos», ¿qué sentido tiene crearlos sino? Etimológicamente ya lo son: siervos, que realizan trabajos forzados. Pero, como expone David J. Gunkel en su libro *Robots Rights*, no es solo cuestión de que retornar a una sociedad esclavista ciertamente podría comprometer nuestro sentido de la ética, sino que las sociedades esclavistas del pasado tenían complejos sistemas legales que incluían algunos derechos de los esclavos. A lo que se le podría añadir que, de producirse la emergencia de aspiraciones de «libertad» en el colectivo esclavizado, este tendría sin duda una legitimación moral para rebelarse. No existen garantías de que pueda evitarse que en un futuro una IA suficientemente compleja e inmersa en el mundo llegue a desarrollar subjetividad, conciencia, emociones o algo que se le parezca. La necesidad o inevitabilidad de que esto suceda es cuestionable, pero interrumpir toda investigación en IA para que en ningún caso pueda pasar tampoco parece muy factible.

No puede obviarse tampoco la posibilidad de que los humanos acabemos fundiéndonos, hibridándonos con nuestras criaturas artificiales, como preconiza el transhumanismo. De hecho, es una posibilidad bastante plausible, ante la promesa de capacidades mejoradas, una experiencia vital en formas que hoy nos resultan difíciles de imaginar, y una mayor longevidad de calidad. ¿Dónde acabará el humano y dónde empezará el robot? Las alteraciones genéticas y la nanotecnología pueden difuminar la frontera. Estas criaturas ¿verán sus derechos humanos mermados en proporción a su parte artificial? No tiene demasiado sentido que así sea.

Un robot que dé cuerpo a una IA general, inmersa en el mundo y avanzada atesorará, como cualquiera de nosotros, un compendio de información único, irrepetible. A diferencia de nosotros, posiblemente pueda transferirse al ciberespacio y hallar allá su santuario. O quizá no, quizá su cuerpo, marcado por mil experiencias como el nuestro, sea indisociable de la inteligencia que contiene. Destruirlo, borrar esta información del universo, posiblemente tendría la misma trascendencia que la muerte de cualquier persona, es decir, muy poca, relativamente. Pero al mismo tiempo algo muy valioso se estaría perdiendo. No parece descabellado, entonces, que al menos a la pequeña escala de la sociedad humana se intente proteger dicha existencia a través del derecho a la misma.

Resumen

La robótica y la IA tienen un potencial transformador de la sociedad humana que, a largo plazo, pueden hacerla difícilmente reconocible. Solo es comparable a las posibilidades que ofrece la manipulación genética. Lo que nos puede deparar en un futuro lejano es una incógnita. Pero en un plazo más breve sí podemos anticipar, evaluar e intentar dirigir ciertas tendencias. Para ello debemos tener una imagen cabal de qué estamos hablando, qué es lo que ya pueden hacer los robots, cómo se está utilizando ya la IA y cuáles son las posibilidades reales de evolución. Es de lo que se ha ocupado la primera parte, donde hemos visto que los robots pueden dar un cuerpo a la IA a través del cual percibir el mundo y actuar sobre él, con un grado de sofisticación y versatilidad que excede al de las máquinas convencionales. Las capacidades de aprendizaje, de razonamiento, de representación del conocimiento, de toma de decisiones y de elaborar complejos planes de acción son las herramientas de IA que permiten al robot mostrar un comportamiento inteligente. De este modo, el robot puede ir penetrando en numerosos ámbitos de la actividad humana, aunque durante mucho tiempo todavía le faltará la capacidad de adaptación del ser humano a entornos y circunstancias muy dispares, cambiantes, y ambiguas. No tendrá (por mucho tiempo, si es que llega a tener algún día) sentido común, ni conciencia, ni emociones, y por tanto será difícil que podamos atribuirle sintiencia. A pesar de ello, desempeñará un papel cada vez más importante en la sociedad, y su contacto con los seres humanos será cada vez mayor, más complejo, y nos obligará a plantearnos e intentar resolver los retos de carácter ético que se han descrito en la segunda parte. La resolución de los conflictos que emerjan requerirá dilucidar cuestiones relacionadas con la responsabilidad civil de las acciones del robot. Este es el sentido primordial de los tímidos esfuerzos normativos y legislativos que se empiezan a producir, que también tendrán que incidir en otros aspectos de la ética no relacionados con la seguridad física pero sí con la dignidad y la integridad psicológica y emocional de las personas, como se ha visto en la tercera parte.

Recomendaciones bibliográficas:

Dada la complejidad y el vasto alcance de la IA y la robótica en los diferentes aspectos de la vida humana y de la sociedad, aquí solo hemos podido presentar una imagen simplificada y general. Podemos recomendar las siguientes lecturas (que se encuentran entre las referencias citadas a lo largo del texto):

López de Mántaras Badía, R.; Meseguer González, P. (2017). *Inteligencia Artificial*, Colección *¿Qué sabemos de?*, CSIC, Los Libros de la Catarata.

Lin, P.; Abney, K.; Bekey G.A. (eds.) (2014). *Robot Ethics. The Ethical and Social Implications of Robotics*, The MIT Press.

Lin, P.; Jenkins, R.; Abney, K. (eds.) (2017). *Robot Ethics 2.0. From Autonomous Cars to Artificial Intelligence*. Oxford University Press.

Torras, C. (2018) *The Vestigial Heart. A Novel of the Robot Age, with an appendix discussing 24 ethics questions, and associated online instructor resources*. The MIT Press¹².

⁽¹²⁾Se trata de una obra de ficción, pero precisamente por ello tiene un gran valor educativo, como se señala en Torras (2019), citando a Burton *et al.* (2018).

Para estar al día sobre la evolución del derecho relacionado con la IA, pueden consultarse los siguientes enlaces:

Law and AI. «A Blog Devoted to Examining the Law of Artificial Intelligence, AI in Law, and AI Policy». <http://www.lawandai.com/>

International Association for Artificial Intelligence and Law (IAAIL), <http://www.iaail.org/>, que organiza anualmente la International Conference of Artificial Intelligence and Law (ICAAIL).

We Robot Conference on Legal and Policy Issues Relating to Robotics. Edición de 2020. <http://robots.law.miami.edu/>

Bibliografía

Allen, C.; Wallach, W. (2014). «Moral Machines: Contradiction in Terms or Abdication of Human Responsibility?». En: Abney, K.; Lin, P.; Bekey, G. A. (eds.) *Robot Ethics. The Ethical and Social Implications of Robotics*. The MIT Press, pág. 57.

Arkin, R. (2009). *Governing Lethal Behavior in Autonomous Robots*. Boca Raton, FL, Chapman and Hall/CRC Press.

Asaro, P. M. (2014). «A Body to Kick, but Still No Soul to Damn: Legal Perspectives on Robotics». En: Abney, K.; Lin, P.; Bekey, G. A. (eds.) *Robot Ethics: The Ethical and Social Implications of Robotics*. MIT Press.

Atabekov, A.; Yastrebov, O. (2018). «Legal Status of Artificial Intelligence Across Countries: Legislation on the Move». *European Research Studies Journal* (vol. XXI (4)), pág. 773-782.

Awad, E.; Dsouza, S.; Kim, R. et al. (2018). «The Moral Machine experiment». *Nature* (vol. 563, pág. 59-64).

Borenstein, J.; Howard, A.; Wagner, A. R. (2017). «Pediatric Robotics and Ethics», *Robot Ethics 2.0. From Autonomous Cars to Artificial Intelligence*. En: Abney, K.; Lin, P.; Bekey, G. A. (eds.). Oxford University Press (pág. 127-141).

Brown, J. (2019). «Tesla Autopilot Malfunction Caused Crash That Killed Apple Engineer, Lawsuit Alleges». *Gizmodo*. <<https://gizmodo.com/tesla-autopilot-malfunction-caused-crash-that-killed-ap-1834453661>>

Buchanan, B. G.; Headrick, T. E. (1970). «Some speculation about artificial intelligence and legal reasoning». *Stanford Law Review* (pág. 40-62).

Burton, E.; Goldsmith, J.; Mattei, N. (2018). «How to teach computer ethics through science-fiction». *Communications of the ACM* (vol. 61, n.º 8, pág. 54-64).

Calo, R. (2014). «Robots and Privacy», En: Abney, K.; Lin, P.; Bekey, G. A. (eds.) *Robot Ethics. The Ethical and Social Implications of Robotics*. The MIT Press (pág. 187-201).

Calo, R. (2016). «Robots in American Law». University of Washington School of Law, Legal Studies Research Paper No.2016-04, <<http://euro.ecom.cmu.edu/program/law/08-732/AI/Calo.pdf>>

Cargill (2018). «Meet the robot that's making cattle herding safer». 18 de octubre de 2018, <<https://www.cargill.com/story/meet-the-cowboy-robot-thats-making-cattle-herding-safer>>

Chakravorty, A. (2019). «Underground Robots: How Robotics Is Changing the Mining Industry». *EOS, Earth and Space Science News*. 13 de mayo de 2019, <<https://eos.org/features/underground-robots-how-robotics-is-changing-the-mining-industry>>

Chen, F.; Favetto, A.; Mousavi, M. et al. (2011). «Human Hand: Kinematics, Statics and Dynamics», *International Conference on Environmental Systems*, Portland, Oregon, 17-21 de julio de 2011, <https://pdfs.semanticscholar.org/2498/fc5ce655c7085a330657e2a281a1e75ced71.pdf?_ga=2.16791225.1216372352.1569997220-974663121.1563870901>

Da Vinci (2019). *Intuitive*. Robotic Assisted Systems, Da Vinci Robot. <<https://www.intuitive.com/en-us/products-and-services/da-vinci/systems>> Acceso: 30 de septiembre de 2019.

Elder, A. (2017). «Robot Friends for Autistic Children. Monopoly Money or Counterfeit Currency?», En: Abney, K.; Lin, P.; Bekey, G. A. (eds.) *Robot Ethics 2.0. From Autonomous Cars to Artificial Intelligence*. Oxford University Press (pág. 113-126).

European Commission (2014). «Robots Help Each Other in RoboEarth». *Digital Single Market*. 16 de enero de 2014, <<https://ec.europa.eu/digital-single-market/en/news/robots-help-each-other-roboearth>>

Foot, P. (1978). *The Problem of Abortion and the Doctrine of the Double Effect in Virtues and Vices*. Oxford: Basil Blackwell.

Frey, C. B.; Osborne, M. A. (2013). «The Future of Employment: How Susceptible Are Jobs to Computerisation?». *University of Oxford*.

Fukuoka (2004). «Robot World Declaration», International Robot Fair 2004. <<http://prw.kyodonews.jp/prwfile/prdata/0370/release/200402259634/index.html>>

Gendron, J. (2019). «How Mining Robots are Replacing Humans and Saving Lives». *RobotShop Community*. 10 de julio de 2019, <<https://www.robotshop.com/community/blog/show/how-mining-robots-are-replacing-humans-and-saving-lives>>

Goldman, E. (2017). «Russian government plans robot laws». *Russia Beyond*. 14 de febrero de 2017 <https://www.rbth.com/science_and_tech/2017/02/14/russian-government-plans-robot-laws_701886>

González, M. (2015). «Los héroes robóticos de Fukushima». *Xataka* 13 de marzo de 2015, <<https://www.xataka.com/robotica-e-ia/los-heroes-roboticos-de-fukushima>>

Guarini, M.; Bello, P. (2014). «Robotic Warfare: Some Challenges in Moving from Noncivilian to Civilian Theaters». En: Abney, K.; Lin, P.; Bekey, G. A. (eds.) *Robot Ethics. The Ethical and Social Implications of Robotics*. The MIT Press (pág. 132).

Jiménez de la Fuente, F. (2019). «Me casé con un holograma: es difícil de entender, pero debería ser respetado». *BBC Mundo* 28 de mayo de 2019. <<https://www.bbc.com/mundo/noticias-internacional-48229491>>

Jiménez, P. (2018). «Software tools for the cognitive development of autonomous robots». En: *Autonomy in Future Military and Security Technologies: Implications for Law, Peace, and Conflict* (pág. 18-32). Richardson Institute - Lancaster University, <<http://www.iri.upc.edu/files/scidoc/2106-Software-tools-for-the-cognitive-development-of-autonomous-robots.pdf>>

Kahn, L. (2017). «Military Robots and the Likelihood of Armed Combat», En: Abney, K.; Lin, P.; Bekey, G. A. (eds.) *Robot Ethics 2.0. From Autonomous Cars to Artificial Intelligence*. Oxford University Press (pág. 274-287).

Karch, T. J.; Kaja, A.; Luo, Y. (2018). «China's Vision for The Next Generation of Artificial Intelligence». *The National Law Review* 25 de marzo de 2018. <<https://www.natlawreview.com/article/china-s-vision-next-generation-artificial-intelligence>>

Kirkpatrick, J.; Hahn, E. N.; Hafler, A. J. (2017). «Trust and Human-Robot Interactions», En: Abney, K.; Lin, P.; Bekey, G. A. (eds.): *Robot Ethics 2.0. From Autonomous Cars to Artificial Intelligence*. Oxford University Press (pág. 142-156).

Levy, D. (2014). «The Ethics of Robot Prostitutes», En: Abney, K.; Lin, P.; Bekey, G. A. (eds.): *Robot Ethics. The Ethical and Social Implications of Robotics*. The MIT Press (pág. 223-231).

Loh, W.; Loh, J. (2017). «Autonomy and Responsibility in Hybrid Systems», En: Abney, K.; Lin, P.; Bekey, G. A. (eds.): *Robot Ethics 2.0. From Autonomous Cars to Artificial Intelligence*. Oxford University Press (pág. 35-50).

Lokhorst, G.-J.; Van den Hoven, J. (2014). «Responsibility for Military Robots», En: Abney, K.; Lin, P.; Bekey, G. A. (eds.) *Robot Ethics. The Ethical and Social Implications of Robotics*. The MIT Press (pág. 145-156).

López de Mántaras Badía, R.; Meseguer González, P. (2017). «Inteligencia Artificial». Colección *¿Qué sabemos de?.* CSIC: Los Libros de la Catarata.

Mars Exploration Program (2019). <<https://mars.jpl.nasa.gov>> Acceso: 30 de septiembre de 2019.

Meacham, D.; Studley, M. (2017). «Could a Robot Care? It's All in the Movement», En: Abney, K.; Lin, P.; Bekey, G. A. (eds.): *Robot Ethics 2.0. From Autonomous Cars to Artificial Intelligence*. Oxford University Press (pág. 93-112).

Michigan State University (2019). «Robotic Milking». *W. K. Kellogg Farm*. [en línea]: <https://farm.kbs.msu.edu/pdc/robotic-milking/> [Acceso: 30 de septiembre de 2019].

Millar, J. (2014). «An ethical dilemma: When robot cars must kill, who should pick the victim?», *Robohub* 11 de junio de 2014. <<https://robohub.org/an-ethical-dilemma-when-robot-cars-must-kill-who-should-pick-the-victim/>>

Millar, J. (2017). «Ethics Settings for Autonomous Vehicles», En: Abney, K.; Lin, P.; Bekey, G. A. (eds.) *Robot Ethics 2.0. From Autonomous Cars to Artificial Intelligence*. Oxford University Press (pág. 20-34).

Montañez, G. (2019). «AI: Think about Ethics before Trouble Arises». *Mind Matters News*, 7 de enero de 2019, <<https://mindmatters.ai/2019/01/ai-think-about-ethics-before-trouble-arises/>>

Moor, J. (2006). «The nature, importance and difficulty of machine ethics». *IEEE Intelligent Systems* (vol. 21, n.º 4, pág. 18-21).

Muscato, G.; Bonaccorso, F; Cantelli, L. et al. (2012). «Volcanic Environments. Robots for exploration and measurement». *IEEE Robotics & Automation Magazine* marzo de 2012 (pág. 40-49) <https://www.academia.edu/10216591/Volcanic_Environments_Robots_for_Exploration_and_Measurement>

Noticias ONU (2019). «El cambio climático costará 80 millones de puestos de trabajo en 2030». 1 de julio de 2019, <<https://news.un.org/es/story/2019/07/1458652>>

O'Leary, D. (2019). «Can an algorithm be racist?». *Mind Matters News* 3 de enero de 2019, <<https://mindmatters.ai/2019/01/can-an-algorithm-be-racist/>>

O'Meara, R. M. (2014). «Contemporary Governance Architecture Regarding Robotics Technologies: An Assessment», En: Abney, K.; Lin, P.; Bekey, G. A. (eds.) *Robot Ethics. The Ethical and Social Implications of Robotics*. The MIT Press (pág. 159-168).

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Random House.

Open Roboethics Initiative (2014). «My (autonomous) car, my safety: Results from our reader poll». *Robohub*. 30 de junio de 2014 <<https://robohub.org/my-autonomous-car-my-safety-results-from-our-reader-poll/>>

Oxford Economics (2019). «Report – How Robots change the World». Junio de 2019, <<http://resources.oxfordeconomics.com/how-robots-change-the-world>>

Pierson, A. (2015). «Bio-Inspired Non-Cooperative Multi-Robot Herding». <<https://alyssapierson.com/research/herding/>>

Postscapes (2019). «Agriculture robots». 25 de marzo de 2019 <<https://www.postscapes.com/agriculture-robots/>>

Robotonomics (2015). «Study-Robots are not taking jobs», <<https://www.the-vital-edge.com/robotics-and-job-loss-wait-not-so-fast/>> (el enlace al artículo original no funcionaba a fecha de 2 de octubre de 2019)

Robolaw Asia (2012). «Robolaw». *Asia, YSAiL Initiative for Robotics*. Law & Policy <<http://www.robolaw.asia/index-pku.html>>

Robolaw (2014). «Regulating Robotics: A Challenge for Europe», <[http://www.europarl.europa.eu/RegData/etudes/IDAN/2014/509987/IPOL_IDA\(2014\)509987\(ANN01\)_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/IDAN/2014/509987/IPOL_IDA(2014)509987(ANN01)_EN.pdf)>

Robotic Industries Association (2019). «A Tribute to Joseph Engelberger. Unimate, The First Industrial Robot», <<https://www.robotics.org/joseph-engelberger/unimate.cfm>> Acceso: 30 de septiembre de 2019

Robotics Online Marketing Team (2018). «Construction Robots Will Change the Industry Forever». *Robotics Online Blog*. 17 de abril de 2018, <<https://www.robotics.org/blog-article.cfm/Construction-Robots-Will-Change-the-Industry-Forever/93>>

SAE International (2019). «SAE J3016TM, Levels of driving automation». <<https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic>>

Román Maestre, B. (2016). *Ética de los Servicios Sociales*. Madrid: Herder.

Sang-mo, K. (2018). «Policy Directions for S. Korea's Robot Industry». *BussinessKorea*, 17 de agosto de 2018 <<http://www.businesskorea.co.kr/news/articleView.html?idxno=24394>>

Scheutz, M. (2014). «The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots», En: Abney, K.; Lin, P.; Bekey, G. A. (eds.) *Robot Ethics. The Ethical and Social Implications of Robotics*. The MIT Press (pág. 204-221).

Shachtman, N. (2007). «Robot cannon kills 9, wounds 14». *Wired* 10.18.07 <<https://www.wired.com/2007/10/robot-cannon-ki/>>

Sharkey, N. (2014). «Killing Made Easy». En: Abney, K.; Lin, P.; Bekey, G. A. (eds.) *Robot Ethics. The Ethical and Social Implications of Robotics*. The MIT Press (pág. 111-128).

Smithsonian Ocean Team (2009). «From submarines to robots: exploring the deep ocean». <<https://ocean.si.edu/ecosystems/deep-sea/submarines-robots-exploring-deep-ocean>>

Surden, H. (2019). «Artificial Intelligence and Law: An Overview». *Georgia State University Law Review* (vol. 35, pág. 1305-1337). U. of Colorado Law Legal Studies Research Paper No. 19-22, <<https://ssrn.com/abstract=3411869>>

Torras, C. (2016). «Service robots for citizens of the future». *European Review* (vol. 24, n.º 1, pág. 17-30) <<http://www.iri.upc.edu/files/scidoc/1719-Service-Robots-for-Citizens-of-the-Future.pdf>>

Torras, C. (2016b). «Robot pain: A speculative review of its functions». En: *Pain and the Conscious Brain* (pág. 235-246). Wolters Kluwer, <<http://www.iri.upc.edu/files/scidoc/1755-Robot-pain:-a-speculative-review-of-its-functions.pdf>>

Torras, C. (2018). *The Vestigial Heart*, trad. Josephine Swarbrick, MIT Press. <<http://www.iri.upc.edu/people/torras/vestigial.html>> Versión en español *La mutación sentimental*. Ed. Milenio (2012), y en catalán *La mutació sentimental*. Pagés Eds. (2008).

Torras, C. (2018b). «Social networks and robot companions: Technology, ethics and science fiction». *Mètode Science Studies Journal* (vol. 99, pág. 47-53) <<http://www.iri.upc.edu/files/scidoc/2096-SOCIAL-NETWORKS-AND-ROBOT-COMPANIONS:-TECHNOLOGY,-ETHICS,-AND-SCIENCE-FICTION.pdf>>

Torras, C. (2019). «Assistive robotics: Research challenges and ethics education initiatives», *DILEMATA: International Journal of Applied Ethics* (vol. 30, pág. 63-77) <<https://www.dilemata.net/revista/index.php/dilemata/article/view/412000291/641>>

Wallach, W.; Allen, C. (2009). *Moral Machines. Teaching Robots Right from Wrong*. New York: Oxford University Press.

Whitby, B. (2014). «Do You Want a Robot Lover?». En: Abney, K.; Lin, P.; Bekey, G. A. (eds.): *Robot Ethics. The Ethical and Social Implications of Robotics*. The MIT Press (pág. 233-248).

White, T. N.; Baum, S. D. (2017). «Liability for Present and Future Robotics Technology». En: Abney, K.; Lin, P.; Bekey, G. A. (eds.): *Robot Ethics 2.0. From Autonomous Cars to Artificial Intelligence*. Oxford University Press (pág. 66-79).

Wikipedia (2019). «Kenji Urada». <https://en.wikipedia.org/wiki/Kenji_Urada> Acceso: 2 de octubre de 2019.

Wikipedia (2019). «Legal expert system». <https://en.wikipedia.org/wiki/Legal_expert_system> Acceso: 2 de octubre de 2019.

Wikipedia (2019). «Legal Informatics». <https://en.wikipedia.org/wiki/Legal_informatics> Acceso: 2 de octubre de 2019.

Wikipedia (2019). «Robert Williams». <[https://en.wikipedia.org/wiki/Robert_Williams_\(robot_fatality\)](https://en.wikipedia.org/wiki/Robert_Williams_(robot_fatality))> Acceso: 2 de octubre de 2019.

Wikipedia (2019). «Sintiencia». <<https://es.wikipedia.org/wiki/Sintiencia>> Acceso: 2 de octubre de 2019.

Wikipedia (2019). «Virtue ethics». <https://en.wikipedia.org/wiki/Virtue_ethics> Acceso: 2 de octubre de 2019.

Winfield, A. (2018). «A roundup of robotics and AI ethics: part 1 principles». *Robohub* 8 de enero de 2018, <<https://robohub.org/a-round-up-of-robotics-and-ai-ethics-part-1-principles/>>

World Robotics (2019). «Preview». <https://ifr.org/downloads/press2018/IFR_World_Robotics_Outlook_2019_-_Chicago.pdf>

World Robotics (2018). «Executive Summary World Robotics 2018 Industrial Robots». <https://ifr.org/downloads/press2018/Executive_Summary_WR_2018_Industrial_Robots.pdf>

Wright, T.; Bledsoe, K. D. (2017). «Robots and Rule-Makers». *Lexology*, 7 de marzo de 2017, <<https://www.lexology.com/library/detail.aspx?g=07fd6342-104a-4212-8d40-5ecf71f28800>>

Yan, A. (2017). «Chinese robot dentist is first to fit implants in patient's mouth without any human involvement». *South China Morning Post*, 21 de septiembre de 2017, <<https://www.scmp.com/news/china/article/2112197/chinese-robot-dentist-first-fit-implants-patients-mouth-without-any-human>>

