



INTELIGENCIA ARTIFICIAL: UN PANORAMA DE ALGUNOS DE SUS DESAFÍOS ÉTICOS Y JURÍDICOS

Universitat de Girona

Alumno: David Álvaro Pascual
Tutor/a: Maria del Carmen Vázquez Rojas

ÍNDICE

1. Introducción.....	3
2. Cuestiones terminológicas y conceptuales	5
2.1 Inteligencia artificial. Concepto, características y funcionamiento	5
2.2 Inteligencia artificial/ Robótica	8
2.3 Ética/Moral.....	9
3. Aplicaciones de la inteligencia artificial al mundo jurídico	12
3.1 Sistemas expertos jurídicos. Noción, función y características.....	13
3.2 Modelos de sistemas expertos jurídicos	14
4. Cuestiones éticas de la inteligencia artificial.....	18
4.1 La moral ante el progreso tecnológico.....	18
4.2 Implantación de la inteligencia artificial de modo seguro	22
4.3 Adquisición de derechos y deberes por parte de máquinas y robots.....	23
4.4 Sistemas de inteligencia artificial más inteligentes que los humanos	25
4.5 Problemas sociales estructurales a raíz de la inteligencia artificial	25
4.6 Los 23 principios de Asilomar.....	28
4.7 Análisis del algoritmo COMPAS: Al límite de la ética	33
5. Nuevos problemas jurídicos a partir de la inteligencia artificial y la robótica	35
5.1 Regulación actual sobre robótica e inteligencia artificial	35
5.2 Derecho internacional y armas autónomas	41
5.3 Derecho de daños y robótica	42
5.3.1 Interrogantes jurídicos de los vehículos autónomos	44
5.4 Derecho laboral y de la seguridad social y robótica.....	46
5.5 Derecho penal y robótica	49
6. Conclusiones	52

1. Introducción

De un tiempo a la actualidad, la humanidad ha tenido el sueño de crear máquinas que sean capaces de ejecutar las tareas más pesadas y tediosas de las personas, e incluso de tomar decisiones de importancia en la vida de las mismas, con el pretexto de hacerles la vida más sencilla y cómoda, objetivo que parece estar en todo avance tecnológico. El derecho no escapa a ese campo de desarrollo. Así pues, como todo cambio que implique progreso, es necesario tener una serie de precauciones que permitan tener un control de dicho progreso, para no convertirlo en contraproducente, y no permitir que lo que ha de ser una ayuda para la sociedad se convierta en un lastre. Pero, ¿cuál es la realidad a la que nos enfrentamos en estos días?

En el año 1997, la supercomputadora “Deep Blue”, desarrollada por IBM, derrotó al campeón del mundo de ajedrez Garri Kasparov, considerado el mejor ajedrecista de todos los tiempos. Puede parecer un dato superfluo, pero lo cierto es que era tan sólo un aviso de todo lo que estaba por venir. Dos décadas después, disponemos de invenciones tecnológicas extraordinarias. Muy recientemente, la empresa automovilística “Tesla” ha comenzado a fabricar coches autónomos, que, aunque todavía en fase de desarrollo, están dotados de sistemas que ya hoy día permiten realizar trayectos sin conductor.¹ Por su parte, el famoso sistema de computación cognitiva “Watson”, también de “IBM”, ya se utiliza para ayudar a los médicos a elegir el tratamiento más adecuado ante varios tipos de cáncer.² Sin embargo, las estimaciones de los expertos para el futuro son aún más impresionantes. Por ejemplo, un estudio del matemático y futurólogo británico Ian Pearson, prevé la posibilidad de que en el año 2050 se pueda prácticamente erradicar la prostitución y la trata de mujeres mediante la creación de robots con fines sexuales.³

Ante un panorama tan inconcebible para nuestros hábitos y estilos de vida aún convencionales en muchos aspectos, surge la necesidad de que la humanidad se haga ciertas preguntas y, ante las respuestas que resulten, habrá de actuar el derecho. Tales preguntas tienen un carácter ético o moral muy pronunciado. En función de la postura que adoptemos ante el inminente cisma tecnológico que se avecina, se tendrán que modificar las normas de una manera o de otra. Este trabajo pretende abordar algunas cuestiones que conciernen a esa etapa de desarrollo tecnológico en la que nos encontramos, pero sobre todo en cuestiones que afectaran a nuestro futuro, mismas que, inevitablemente tienen un alto contenido jurídico. En primer lugar, se expondrá lo que nos puede ofrecer la inteligencia artificial aplicada a la toma de decisiones judiciales, mediante los llamados sistemas jurídicos expertos. En tal escenario surgen al menos dos preguntas: ¿son dichos sistemas fiables tecnológicamente? Y, ¿en el caso de que lo sean, ¿cómo debemos implantarlos?

¹ (28 de diciembre de 2016) “El Autopilot de Tesla evita un accidente”. *El Mundo*. Recuperado de : <http://www.elmundo.es/motor/2016/12/28/5863ede6e5fdea6c0e8b4669.html>

² (28 de junio de 2015) “Watson contra el cáncer” *Cromo*. Recuperado de: <http://www.cromo.com.uy/watson-contra-el-cancer-n656130>

³ Fresneda, C. (1 de octubre de 2015) “2050: la revolución de los robots sexuales” *El Mundo*. Recuperado de: <http://www.elmundo.es/ciencia/2015/10/01/560c1d4be2704e996b8b459a.html>

Por otra parte, y como ya se ha apuntado, hay que tener en cuenta ciertas consideraciones éticas y morales ante la inteligencia artificial y la robótica –conceptos, por cierto, que no son sinónimos, como posteriormente se expondrá-. ¿Qué es lo que debemos preservar a toda costa de lo que nos hace humanos? ¿Qué consecuencias para nuestro modo de concebir la realidad puede tener esta revolución futura? ¿Cómo deberíamos relacionarnos con esos sistemas artificiales para que sean un provecho y no un problema? Ante estas preguntas surge un debate marcadamente filosófico en el que aquí se tratará de hacer hincapié. Existen varias posturas ante el fenómeno del progreso tecnológico, todas ellas defienden bienes jurídicos o maneras de concebir la vida humana merecedoras de ser respetadas y, por ello, no se hará aquí un juicio de valor a ninguna de ellas. No obstante, se tratará de exponer las ventajas y desventajas de todas las posiciones de manera objetiva.

Finalmente, un desarrollo tecnológico de tal magnitud requiere de ordenamientos jurídicos nacionales e internacionales eficaces y seguros para el bienestar de la humanidad. Ante el gran desconocimiento, incluso de los juristas más expertos, del impacto que dicho desarrollo puede suponer, nace la necesidad bien de adaptar nuestras normas existentes al paradigma venidero o de poner en juego nuevas normas que traten el régimen jurídico tanto de sistemas de inteligencia artificial como de robots. Ello suscita la necesidad de modificar el ordenamiento en cuestiones tan importantes como la responsabilidad civil por los daños que aquéllos causen, el tratamiento que se le ha de otorgar a la propiedad intelectual de creaciones hechas por los mismos o la garantía de los derechos fundamentales, con tal de que las personas no hayan nunca de verse obligadas a renunciar, por ejemplo, a determinados puestos de trabajo. En ese sentido, es menester que las instituciones de gobierno nacionales y supranacionales empiecen a tratar el tema lo antes posible; de momento, queda aún lejano el momento en que los ordenamientos jurídicos estén dotados plenamente de estas garantías, aunque no obstante en algunas zonas del mundo –Japón, Corea del Sur o Taiwán- ya existen las primeras normativas que regulan estos asuntos. Por lo que a nuestro entorno respecta, ha habido ya algunas referencias a este tema por parte de las instituciones de la Unión Europea y ya hay autores especialistas en el tema que lo han abordado.

En definitiva pues, nos encontramos ante uno de los desafíos más importantes de la historia de la humanidad: adaptar a la realidad de las personas una de las tecnologías con más potencial para hacer el bien. Para ello, la ciencia dispone ya de la materia prima necesaria y el derecho y la filosofía –junto con otras disciplinas como la economía o la sociología- tienen que moldearla para construir un mundo más próspero. Se tratará aquí de dar una buena visión general de esas cuestiones, sin profundizar exhaustivamente en ninguna de ellas.

2. Cuestiones terminológicas y conceptuales

Este trabajo es de carácter jurídico-filosófico. Como es sabido, la filosofía suele abordar algunos temas con bastante contenido técnico-científico, siendo éste uno de esos casos. Un trabajo que trate sobre inteligencia artificial, como todo trabajo que incluye cuestiones científicas –aunque no sea un trabajo del ámbito científico- debe contener, en aras de facilitar su comprensión, algunas definiciones previas, tanto de conceptos técnicos complejos como de términos que inicialmente pueden parecer similares, pero tienen sus matices. Así pues, este apartado está dedicado a la noción y características de diversos conceptos que se tratarán en el trabajo, sin profundizar en exceso en ellos, pues el objetivo es clarificarlos para la comprensión de estas páginas. En consecuencia, cuando se profundice en los diversos apartados sobre alguna cuestión, no se explicarán estos conceptos controvertidos, sino que se hará de forma previa en este apartado, como una especie de glosario introductorio.

2.1 Inteligencia artificial. Concepto, características y funcionamiento

El concepto de inteligencia artificial puede tener varias definiciones plausibles. Antes de centrarnos en él, es conveniente hacer una definición del concepto de inteligencia. El término *inteligencia* proviene del latín “*inteligere*”, y se puede definir como “*la facultad de comprender las relaciones entre los hechos y las cosas*”⁴. Podemos afirmar, en consecuencia, que la comprensión y la capacidad de comunicación de lo comprendido son los rasgos básicos de la inteligencia. Así, una persona con más inteligencia que otra es capaz de relacionar más eficientemente los conocimientos que posee para resolver un problema. Es menester distinguir este término con otros que se suelen confundir en el uso diario como la sabiduría o la cultura. Estos últimos conceptos hacen referencia al conjunto de conocimientos que una persona acumula. Por ende, una persona con muchos conocimientos acumulados será sabia o culta – depende del matiz subjetivo que le otorguemos al término cultura- pero ello no quiere decir necesariamente que esos conocimientos estén relacionados con la inteligencia de esa persona. Tampoco debe confundirse el concepto de inteligencia con la memoria, siendo ésta la capacidad mnemotécnica para recordar una gran cantidad de datos, de manera que una persona puede ser capaz de recordar muchos datos, pero no tener la inteligencia suficiente para utilizar esos datos a efectos de resolver un problema, lo cual sería, ahora sí, una manifestación de la inteligencia. Para acabar de perfilar el concepto de inteligencia es necesario distinguir dos teorías acerca de la misma. Por una parte, la teoría que engloba todas las manifestaciones de inteligencia en el mismo concepto, de modo que todas ellas, aunque en ámbitos diferentes, dependan del mismo tipo de habilidad. Por otra parte, hay una gran cantidad de autores y expertos que afirman que existen varios tipos de inteligencia y que cada una de ellas es independiente de las otras, de manera que una persona puede tener una gran inteligencia lingüística pero poca inteligencia lógico-matemática. Según los autores, se catalogan estos tipos de inteligencia desde 8 a 12 clases diferentes.⁵ Como no es objeto de este trabajo, es suficiente con mencionarlo.

⁴ Hachette encyclopédiques, 1994

⁵ Para profundizar en la materia se puede consultar a Gardner, H. (1993) “Inteligencias múltiples: la teoría en práctica.” Nueva York. Hellinger, B y Olvera, A. (2010) “Inteligencia transgeneracional: Sanando las

Una vez dada esta noción de inteligencia, podemos empezar a desgranar el concepto de inteligencia artificial, un término, por cierto, que nació en Dartmouth (EE.UU) en 1956, en el marco de una conferencia científica. La inteligencia artificial persigue el fin de utilizar las máquinas para determinar las funciones cognitivas de las personas e intentar emularlas. Así pues, es una rama de la informática que pretende reproducir las funciones cognitivas humanas tales como el razonamiento o la toma de decisiones (signos de inteligencia). También la podemos definir como la rama de la ciencia computacional que investiga y desarrolla programas cuyos productos finales, de ser atribuibles a un ser humano presupondrían procesos mentales inteligentes.⁶ En otras palabras, un sistema con inteligencia artificial tiene la capacidad de llevar a cabo procesos, que si los llevara a cabo una persona, diríamos de esa persona que es inteligente. Cuando un coche autónomo reconoce la voz de una persona, interactúa con ella y le lleva a su destino, en realidad está utilizando los conocimientos que tiene adquiridos para resolver un problema –en este caso llevar al pasajero a su destino-. Si en vez de un coche autónomo, una persona minusválida le pide a otra que le lleve al mismo destino, la persona encargada de llevarlo utilizara los conocimientos y conceptos adquiridos –por ejemplo, lo que significa llevar o transportar, la ruta que debe seguir, cómo debe agarrar la silla de ruedas, etc-, es decir, llevará a cabo procesos inteligentes. La única diferencia pues entre la inteligencia humana y la inteligencia artificial es, redundantemente, la artificialidad de la segunda, que es conseguida mediante técnicas y sistemas artificiales, al contrario que la inteligencia humana, o incluso la animal, que son producidas biológicamente.

En cualquier intento por definir el concepto de inteligencia artificial es clásico mencionar al celeberrimo científico y matemático inglés Alan Turing, quien, además de enormes contribuciones a la ciencia, por lo que aquí concierne, ideó el test de Turing, propuesto en 1950⁷ y la función del cual es discernir si una máquina tiene inteligencia o no. La inteligencia artificial, debido a su compleja naturaleza, se compone de varias ramas o maneras de desarrollarse. Cada una de estas ramas son simplemente diferentes metodologías mediante las cuales se pretende resolver un problema. Algunas de ellas, las más tradicionales, tienen ya un largo recorrido de investigación, mientras que en los últimos lustros se están desarrollando nuevas ramas, las denominadas de vanguardia. Se expondrán a continuación de manera somera el concepto y las características solamente de las que tienen más relación y conexión con las aplicaciones jurídicas en general.⁸

- a) **Redes neuronales:** se trata de una de las ramas de la inteligencia artificial en la que se ha puesto más énfasis en los últimos años y que más repercusión y

heridas del pasado. Constelaciones familiares.” Grupo Cudec Y Gardner, H. (2011) “La inteligencia reformulada”. Paidós Ibérica.

⁶ Cáceres Nieto, E. (26 de octubre de 2010). Sistema EXPERTIUS: Sistema experto para la ayuda a la decisión judicial en la tradición romano-germánica. XIV Congreso Iberoamericano de Derecho e Informática. IJ-UNAM. Conferencia realizada en Monterrey, N.L, México.

⁷ Turing, A. (1950). *Computing machinery and intelligence*. Oxford University Press on behalf of the Mind Association.

⁸ Además de las mencionadas, existen otras ramas de la inteligencia artificial (logaritmos genéticos, agentes inteligentes, lógica difusa, o procesamiento de lenguaje natural)

aplicaciones está teniendo actualmente. Las redes neuronales tratan de emular la forma física en que funciona nuestro cerebro, es decir mediante el aprendizaje, contraponiéndose a la manera clásica que tienen de trabajar las computadoras, es decir, la programación. Para simplificar el concepto se tratará de explicar con una metáfora de Oliver Nabani.⁹ Imaginemos que tenemos una habitación con una multitud de personas dentro y una tarea que éstas han de resolver. Supongamos que, para resolver este problema, algún input externo nos proporciona tres números. Supongamos además que cada persona en la habitación cumple la función de una neurona, cada cual tiene asignada una habilidad determinada –una sabe sumar, otra restar, multiplicar, realizar logaritmos, y así con varias operaciones diferentes. Otras personas de la habitación tienen la habilidad de dar órdenes al resto para que realicen las operaciones asignadas. Imaginemos que el input externo nos dice que a partir del número 4 y del número 5 debemos obtener el número 9. Ahora, una persona encargada de dar órdenes emitirá una orden a la persona encargada de restar para que efectúe restas con el número 4 y 5. De esa orden y posteriores operaciones obtenemos los resultados 1 y -1 (5-4 y 4-5). Al no tener éxito mediante dicha operación –no hemos obtenido el 9 que nos demanda el input externo- lo que hacen las personas de la sala (neuronas) es reestructurarse, rediseñarse físicamente para efectuar otra operación, y dar órdenes a las neuronas que efectúan operaciones para que mediante las operaciones pertinentes logren sacar el resultado que desea el input externo. Después de sacar resultados negativos que resultan de multitud de operaciones, en algún punto las personas de la sala obtienen el 9. Seguidamente a estas personas les llegan nuevos inputs con diferentes valores y diferentes resultados deseados a los cuales se llega mediante una suma. En algún momento, a base de entrenar a las neuronas –las personas en la sala de nuestra metáfora- llegan a la conclusión de que el algoritmo correcto es una suma. Esta sencilla metáfora llevada a escalas de complejidad que pueden variar, es la forma en que nuestras neuronas aprenden y a la cual tratan de reproducir las redes neuronales. De hecho, la primera red neuronal que funcionó fue un sistema al que se le daban fotografías y ésta debía de identificar si en ellas aparecía un gato o no. Al sistema se le iban introduciendo infinidad de fotos diferentes de gatos y tras cada error o acierto se le comunicaba, por lo que tras muchísimos errores –cada vez menos- era capaz de identificar en cualquier foto si aparecía o no un gato. Por tanto, este sistema es capaz de tomar decisiones futuras en función de experiencias pasadas, es decir en función del aprendizaje. Un ejemplo de aplicación de las redes neuronales es el reconocimiento facial de la red social *Facebook*¹⁰.

⁹ Oliver Nabani. (28 de mayo de 2016). *Redes Neuronales*. <https://www.youtube.com/watch?v=6Q9nZM30IBE>

¹⁰ (24 de marzo de 2014) “El reconocimiento facial de Facebook, tan preciso que da miedo”. ABC. Recuperado de: <http://www.abc.es/tecnologia/redes/20140321/abci-facebook-reconocimiento-facial-preciso-201403211737.html>

- b) **Sistemas expertos:** Según la British Computer Society, se define un sistema experto como “la implementación en un sistema de computación de una base de conocimiento especializada y versada, de modo que la máquina puede ofrecer consejo inteligente o tomar una decisión inteligente sobre una función del proceso”. Los sistemas expertos son computadoras cuyo fin es tratar de modelar la forma de pensar de los expertos mismos en los distintos ámbitos de la sociedad. Ahora bien, ¿por qué surge la necesidad de crear sistemas expertos artificiales si ya tenemos expertos humanos?¹¹ En primer lugar los expertos humanos son escasos, son muy caros (salario, coste de formación, mantenimiento, etc.), tardan mucho en desarrollar su conocimiento y además están limitados por sus defectos humanos –cansancio, estrés, presión-. Por su parte, los expertos artificiales son potencialmente permanentes –si no se averían o destruyen, lo que desde luego se puede evitar- mientras que un experto humano, cuando muere, se lleva todo su conocimiento con él. Además, son más económicos que los expertos humanos y es posible modelar una gran cantidad de conocimiento en muy poco tiempo. Así pues, un sistema experto es un sistema computacional capaz de emitir respuestas a problemas, que si las atribuyéramos a humanos, presupondrían procesos heurísticos.¹²
- c) **Robótica:** al ser un concepto tan extenso y propenso a ser confundido con el de inteligencia artificial, se expondrá en el apartado siguiente.

2.2 Inteligencia artificial/ Robótica

Para abordar este trabajo es preciso distinguir entre los conceptos de robótica e inteligencia artificial. Para empezar, hemos de definir tanto sociológicamente como jurídicamente qué es un robot, o en términos generales, qué es la robótica. Sociológicamente, la robótica sería todo aquello capaz de sustituir al ser humano en la realización de las tareas que éste ejecuta o de llevar a cabo tareas para las cuales aquél es incapaz, mediante procesos mecanizados y programados, y que se integra en los procesos productivos de una sociedad, pudiendo incluso interactuar con los miembros de ésta. La robótica como ciencia, es la que estudia el diseño y construcción de máquinas capaces de desempeñar las tareas mencionadas. Jurídicamente es algo más complejo definir a la robótica, pues el concepto no aparece en prácticamente ninguna legislación moderna. Podríamos definirla como el conjunto de mecanismos que ejecutan o desarrollan actividades programadas previamente por un operador humano, las cuales deben siempre perseguir objetivos no contrarios al ordenamiento jurídico. En este

¹¹ Cáceres Nieto, E. (26 de octubre de 2010). Sistema EXPERTIUS: Sistema experto para la ayuda a la decisión judicial en la tradición romano-germánica,. XIV Congreso Iberoamericano de Derecho e Informática. IJ-UNAM. Conferencia realizada en Monterrey, N.L, México.

¹² La heurística, término con etimología griega, se traduce en la capacidad de cualquier sistema para realizar instantáneamente soluciones o innovaciones positivas para sus fines, como podría ser la capacidad para solventar problemas mediante la creatividad, o el pensamiento lateral. Según el matemático húngaro George Pólya, la base de la heurística está en la experiencia de resolver problemas y ver cómo otros lo hacen.

trabajo partiré de la clasificación del concepto de robótica en 4 niveles efectuada por Nisa Ávila, cada uno con un grado de sofisticación superior al anterior.¹³

Nivel 1 → Lo constituyen los sistemas inteligentes programados, es decir, máquinas que llevan integrados sistemas robóticos que ayudan a sus dueños o personas que los utilicen a efectuar tareas automatizadas, pero que no son en sí mismos robots porque no son más que un pequeño sistema robótico integrado dentro de otro no robótico, donde predomina este último.

Nivel 2 → Robots no autónomos, que serían todos aquellos que en su totalidad a nivel de construcción industrial son robots y se dedican a realizar tareas programadas simples sin asistencia humana, y que en mayor o menor medida son capaces de tomar decisiones mecánicas *vinculadas exclusivamente a su tarea ante* imprevistos o situaciones de contingencia. Por ejemplo, los robots aspiradora, que al chocar contra un objeto, prosiguen su ruta.

Nivel 3 → Robots autónomos, los cuales tienen capacidad de desarrollar tareas encomendadas complejas, listando las mismas, priorizando y tomando decisiones con libertad dentro de su ámbito de trabajo, con base en unos objetivos encomendados sin asistencia humana, con autonomía suficiente para que dada una orden por un operador humano pueda llevar a cabo sus funciones según lo considere más adecuado para el cumplimiento de su objetivo, y teniendo siempre como principio elemental la protección de productos o maquinarias y de usuarios o humanos como

Nivel 4 → La inteligencia artificial es el último nivel y el más sofisticado dentro del progreso en robótica. Los sistemas dotados de inteligencia artificial son aquellos sistemas mecánicos que perciben el ambiente externo por sí mismo sin necesidad de órdenes preprogramadas externas, con capacidad de captar y percibir diferentes circunstancias que sucedan en su entorno y con capacidad para desplazarse voluntariamente. Los sistemas de inteligencia artificial deben de obedecer tanto órdenes de otros sistemas de inteligencia artificial como las órdenes de los humanos que interactúan con él.

Explicadas estas características se puede concluir que pueden existir tanto robots con inteligencia artificial como inteligencia artificial sin robots, ya que el robot siempre tiene un soporte físico mientras que la inteligencia artificial se basa en bytes y algoritmos, es decir, es un tipo de software.

2.3 Ética/Moral

En un trabajo en que haya contenido relativo a la inteligencia artificial, sea cual sea su ámbito de aplicación, es inevitable hacer referencias a la ética, pues no en vano la

¹³ Nisa Ávila, J.A. (30 de marzo de 2016) "Robotica e Inteligencia Artificial ¿Legislación social o nuevo ordenamiento jurídico?". *El derecho*. Recuperado de: http://tecnologia.elderecho.com/tecnologia/internet_y_tecnologia/Robotica-Inteligencia-Artificial-legislacion-social-nuevo-ordenamiento_11_935305005.html

creación de máquinas en cierto modo autosuficientes puede afectar, y mucho, la manera de vivir y de pensar de las personas. En primer lugar, es conveniente desgranar o especificar las diversas connotaciones sobre el término “ética” y como sus características.

El concepto de ética suele referirse a todo sistema normativo con pretensión de regular nuestra conducta que se nos impone, no por razones legales, sino por el valor intrínseco que este sistema tiene para los humanos. Así, si se nos impone, podemos decir que la ética tiene un carácter prescriptivo. No obstante, a diferencia del derecho, el cual suele tener una jurisdicción bien delimitada en la cual es aplicable, la ética suele ser aplicable a la humanidad en general –más allá de la idiosincrasia y el pensamiento colectivo de cada comunidad- y no se circunscribe a un territorio o jurisdicción concreta –si bien sí que podemos hablar de la moral social-. Además, la sanción o consecuencia al incumplimiento de sus pautas, que actúa como motivo de disuasión en la conducta no consiste en privaciones de libertad, derechos o patrimonio, sino que suele consistir en la desaprobación o rechazo hacia esa conducta de los demás individuos. Otra buena definición de ética sería la de considerarla como “*la ciencia filosófica que se encarga de reflexionar sobre los comportamientos morales del ser humano con el objetivo de realizar valoraciones genéricas que puedan ser universalizables*”¹⁴

Respecto al término moral, este se utiliza habitualmente como sinónimo de ética, pues las diferencias son meramente para distinguir ambos conceptos a nivel de su estudio, dado que la ética suele referirse a aspectos formales (teorías o métodos) y la moral suele englobar el contenido u objeto (normas o principios morales concretos).¹⁵ Por ejemplo, para distinguir su significado en una oración, podríamos decir que la ética es el conjunto de teorías y métodos que determinan qué conductas son o no son morales. También se puede afirmar que la ética no pretende respondernos a las preguntas acerca de qué está bien o mal, si no analizar las diferentes morales y determinar los criterios que éstas utilizan para catalogar una acción como buena o mala. Sea como fuere, el objeto de este trabajo no requiere de precisiones doctrinales sobre los matices del término.

No obstante, es menester mencionar la manera habitual de clasificar el concepto de moral, distinguiéndola en 3 tipos o categorías: la moral universal o crítica, la moral social o positiva y la moral individual. Para los que la sostienen la moral universal, sólo existen unos criterios únicos para determinar la corrección de las acciones humanas. Así, desde el punto de vista interno de una persona que sostiene la moral universal, solamente existirán esas pautas morales, las universales e inmutables, por lo que todo aquello que se aleje de dichas pautas será considerado inmoral. A su vez, desde el punto de vista externo, pueden existir (y de hecho existen) diferentes contenidos de lo que constituiría la moral universal- véase todo el conjunto de religiones, por ejemplo, cada una con su moral universal diferente a las otras-. El ejemplo anterior, sin embargo, no implica que la moral haya de identificarse necesariamente con una religión. Respecto a la moral

¹⁴ Castrillón Gómez, O.D., Rodríguez Córdoba, M. y Leyton Castaño, J.D. “Ética e inteligencia artificial ¿Necesidad o urgencia?”. *International Institute of Informatics and Systemics*. Recuperado de: <http://www.iiis.org/CDs2008/CD2008CSC/CISCI2008/PapersPdf/C054TM.pdf>

¹⁵ Fabris, A. ,Bartolommei, S. y Datteri, E. “Ética y robótica”. *Eikasia revista de filosofía*. Recuperado de: <http://www.revistadefilosofia.org/19-11.pdf>

social o positiva, podemos decir que son los criterios de corrección que están vigentes en una sociedad, comunidad o colectivo de personas concreto, en un momento determinado de la historia. Para una persona que sostenga la moral universal, solamente reconocerá una moral social en cuanto siga las pautas morales de dicha universalidad moral. En cambio, para las personas que sostienen y conforman la moral social de una determinada comunidad de personas, no existe una moral universal, pues éstas son conscientes de que sus criterios son válidos solamente para ellas, aceptando que pueden existir y existen otros colectivos con criterios morales diferentes. Por último, se encuentra la moral individual, que conforma el conjunto de pautas morales de un único individuo. La dificultad para determinar la frontera entre estas distinciones conceptuales es elevada, pues la moral individual puede compartir criterios de una determinada moral social, pero alejarse en otros. Así se puede decir que la moral social es la más trascendente, y que es más trascendente cuanto más amplio sea el número de personas que la sigan. Por ejemplo, si consideramos la moral española, es decir la que está vigente en el estado español, encontramos a su vez otras morales sociales menores dentro de ella. Sin embargo, lo que conforma la moral social española es todo aquello en que coinciden las distintas sociedades más pequeñas que conforman el estado español. Es decir, parece bastante claro que, por ejemplo, Cataluña y la Comunidad de Madrid difieren en ciertas cuestiones morales, y que esas diferencias son las que las caracterizan como morales sociales distintas, mas coinciden en la gran mayoría de cuestiones morales, las cuales –las que coinciden- forman, como se ha dicho, la moral social española. Dejando de lado la moral social, cada individuo tendrá su propia moral individual. La mayoría de individuos que forman el tejido de una determinada moral social compartirán la mayoría de patrones e ideales morales de la misma, pero pueden además tener su propia moral individual que difiera en algunos aspectos de esa moral social o positiva. No obstante, la moral individual, desde un punto de vista externo, parece ser secundaria y está subordinada a la moral social del entorno del individuo, pues los criterios de un individuo por sí solo no pueden tener más peso que criterios compartidos por muchos individuos. Ahora bien, desde el punto de vista interno, la moral individual es más importante que la social para cada persona.

Precisamente por esta mutabilidad de los patrones morales, no es tan sencillo responder a las preguntas de carácter ético o moral en relación al auge de la inteligencia artificial, ya que lo que hasta ahora se había considerado totalmente indeseable por la humanidad, podría en un futuro ser plausible. Un hipotético ejemplo sería la creación de ciborgs – un ser híbrido mitad humano mitad máquina -, concepto hasta ahora reservado a las películas de ciencia ficción, y que a algunos les puede resultar ridícula tan sólo la idea de plantearse. No obstante, hay varios expertos, como el magnate de la tecnología Elon Musk que ya han planteado no sólo la posibilidad, sino la necesidad de que en épocas venideras los humanos deban aumentar sus capacidades físicas o mentales mediante la implantación de sistemas artificiales.¹⁶ Este último dato sólo es un ejemplo de la sensibilidad con la cual hay que afrontar de manera ética la inteligencia artificial o la robótica, pues cuando hay bienes jurídicos vitales en ponderación,

¹⁶ (17 de febrero de 2017) Kharpal. A. "Elon Musk: Humans must merge with machines or become irrelevant in AI age. CNBC. Recuperado de: <http://www.cnbc.com/2017/02/13/elon-musk-humans-merge-machines-cyborg-artificial-intelligence-robots.html>

cualquier decisión pese a tener sus virtudes también tendrá sus consecuencias negativas.

3. Aplicaciones de la inteligencia artificial al mundo jurídico

Como es sabido, el derecho, en su acepción equivalente a actividad (ser jurista) requiere de actividad intelectual y cognitiva. Por ello, su relación simbiótica con la inteligencia artificial tiene sentido. Sin embargo, no es algo nuevo el hecho de utilizar la inteligencia artificial para los propósitos del mundo jurídico. De hecho, se llevan varios años utilizando algunas técnicas y sistemas de inteligencia artificial para facilitar la vida a los operadores jurídicos.

Los primeros avances en IA aplicada al derecho iban dirigidos a simples procedimientos, tales como la aplicación de reglas sencillas a hechos no controvertidos. No obstante, recientemente este campo de la investigación científica se ha centrado en áreas más complejas y desafiantes que conciernen a los procesos legales y se ha hecho énfasis sobretudo en la toma de decisiones judiciales. De momento, ningún mero engranaje de reglas o patrones algorítmicos pueden emular a la perfección el modelo humano de toma de decisiones judiciales, pues no en vano el juez tiene la ardua tarea de solucionar controversias de las cuales suelen depender el patrimonio, los derechos o la libertad de las personas. La toma de decisiones judiciales requiere una unión entre conocimientos legales y habilidades cognitivas. Muchos conceptos presentes en la toma de decisiones judiciales tienen una vaguedad que solamente pueden interpretar y resolver los humanos; pensemos por ejemplo en los conceptos, según el ordenamiento español, de “diligencia de un buen padre de familia”, o “moral pública”. En efecto, conceptos como estos tienen, como muchos otros, una vaguedad inherente a ellos, únicamente superable mediante la discrecionalidad de personas humanas, instituidas como jueces. En ese sentido, el razonamiento judicial combina diversas habilidades cognitivas, como la valoración de hechos –es decir, decidir lo que sucedió a partir de las pruebas presentadas-, la interpretación de textos –es decir, atribuirle un significado al enunciado normativo creado por el legislador para identificar la norma-, o la creación de analogías –es decir, dar una solución a un caso no reglado a partir de otro sí reglado considerando las semejanzas en sus aspectos relevantes-.

Los proyectos de desarrollo de inteligencia artificial aplicada a la toma de decisiones judiciales se han alejado de la idea utópica de crear un juez robot que resuelva casos reales y se ha enfocado en el desarrollo de herramientas prácticas y analíticas que sirvan de ayuda al juez humano en sus actividades de resolución de casos. No obstante, el hecho de agregarle elementos proporcionados por sistemas de inteligencia artificial a la discrecionalidad judicial humana, no está exento de peligros, –como más adelante se verá- a pesar de sus beneficios.

Los juzgadores, en sus tareas, ejecutan actividades del tipo: evaluar la credibilidad de un testigo, el valor probatorio de unas pruebas, o en especial en el orden penal, el balance entre piedad y severidad a la hora de aplicar una pena, valorando las circunstancias

internas y externas del caso, dentro del rango discrecional que la ley les otorga.¹⁷ Por ello, el componente humano es clave en los razonamientos que se dan en la resolución de casos particulares. Así pues, ¿de qué maneras puede contribuir la inteligencia artificial al progreso en el campo de la toma de decisiones judiciales? Pese a que he afirmado que ningún sistema computacional sería capaz de emular el razonamiento humano de manera idéntica, actualmente y más todavía en un futuro a medio plazo, la inteligencia artificial puede llegar a imitar el comportamiento humano con un grado muy avanzado de similitud. En esa línea se encuentran los sistemas expertos, que no son otra cosa que programas que imitan el comportamiento humano. Utilizan información previamente introducida a dicho sistema y emiten una solución con base en dicha información.

Como se ha dicho, la inteligencia artificial no se está desarrollando actualmente con la idea de sustituir a un juez humano. En ese sentido existen diversos grados de colaboración aplicada al derecho entre la máquina y el humano. En primer lugar, encontramos los sistemas que sustituyen al experto. En segundo lugar, se encuentran los sistemas que identifican el problema y los sistemas que ayudan a identificarlo, y, en tercer lugar, tenemos los sistemas que ayudan a la toma de decisiones.¹⁸

3.1 Sistemas expertos jurídicos. Noción, función y características

Los sistemas expertos jurídicos (en adelante SEJ) pretenden situar la inteligencia artificial al servicio del derecho. De hecho, son las herramientas de inteligencia artificial más utilizadas en el derecho. En el apartado de terminología se ha expuesto ya en qué consistía un sistema experto, siendo éste en síntesis un entorno lógico con capacidad de gestionar una base de conocimientos propia, resolver problemas específicos, producir nuevos conocimientos y explicar su razonamiento.¹⁹ Los sistemas expertos jurídicos no son simples sistemas de gestión, sino que son capaces de razonar en un entorno. Además, se rigen por reglas jurídicas, tratando de encontrar una solución mediante reglas válidas. Más importante es, sin embargo, que estos sistemas expertos jurídicos son capaces de emitir un resultado válido y jurídicamente justificable, así como todas las etapas del razonamiento. De esta manera, el usuario que utiliza el SEJ debe poder consultar y responder a las razones del sistema.²⁰

Al ser sistemas encargados de contribuir a la resolución de conflictos, hemos de distinguir varias clases de conflictos. En primer lugar, los conflictos que tienen una solución predeterminada, como por ejemplo resolver un crucigrama. En segundo lugar, tenemos los conflictos que no tienen una solución predeterminada, pero ésta es determinable mediante ciertas reglas, como por ejemplo resolver un problema matemático. En tercer lugar, existen los conflictos divergentes, a los cuales se les puede aplicar diversas soluciones, y es aquí donde quedarían englobadas las controversias jurídicas. Así pues, el sistema jurídico experto es capaz de solucionar una controversia

¹⁷ Sartor, G. y Branting, K. (1998) "Judicial Applications of Artificial Intelligence". Wolters kluwer

¹⁸ Martínez Bahena, G.C. (2012). "La inteligencia artificial y su aplicación al campo del derecho." *Alegatos* (núm 82) p.828-846. Recuperado de: <http://www.corteidh.or.cr/tablas/r30570.pdf>

¹⁹ Bourcier, D. (2003). "Inteligencia artificial aplicada al derecho". Barcelona (España). Pompeu Casanovas (Editorial UOC) p.70

²⁰ Bourcier, D. (2003). "Inteligencia artificial aplicada al derecho". Barcelona (España). Pompeu Casanovas (Editorial UOC) p.71

jurídica de diferentes maneras. A diferencia de los sistemas de ayuda documental jurídica, que simplemente son bases de datos que suelen contener jurisprudencia, legislación y doctrina, estos sistemas intentan emular la manera en la que los juristas interpretan o aplican el derecho. Ahora bien, ¿cómo interactúa este sistema con el experto humano? Como ya se ha apuntado, estos sistemas no están pensados actualmente para resolver controversias por sí mismos, sino que requieren de una relación con el usuario. El sistema, a través de su base de datos de conocimiento jurídico –legislación, jurisprudencia, doctrina relevante- permite al motor de inferencia ejecutar patrones de secuencias de razonamiento para encontrar solución a los problemas planteados. No obstante, para esto último se requerirá la colaboración del usuario, que ha de proporcionar la información adecuada y relevante al sistema para que este pueda llegar a una solución. Es decir, las cargas cognitivas siempre se repartirán entre el humano y la máquina.²¹ El sistema dispone también de una base de hechos que almacena los hechos y datos particulares de un caso pendiente de resolución. Esta base se nutre constantemente con el razonamiento del sistema. Además, está dotado de una interfaz en la que las preguntas y respuestas se hacen de acuerdo con los casos planteados y que gestiona las funcionalidades del diálogo.²²

Mediante esas características, los SEJ contribuyen a crear sistemas judiciales más seguros. Todo sistema judicial humano es en cierto modo inconsistente. Esta noción de inconsistencia implica que ante mismas circunstancias fácticas podemos tomar decisiones distintas. Llevado a la práctica jurídica, nos podemos encontrar (y de hecho nos encontramos habitualmente) con sentencias contradictorias para casos completamente análogos. La consecuencia de esto es obviamente la inseguridad jurídica que provoca. Un sistema experto puede ayudar a corregir en buena medida este fenómeno y contribuir así a tejer un sistema judicial más seguro y previsible. Todo sistema experto jurídico se elabora siguiendo una serie de pasos. En primer lugar es necesario determinar el ámbito experto sobre el cual versará el sistema. En segundo lugar, hay que determinar qué tipo de problemas resolverá, para así determinar también las cargas cognitivas para el sistema y para el usuario. Después, se identifica un grupo de expertos, quienes se encargan de colaborar mediante el vertido de información al sistema –conceptos, definiciones, criterios de interpretación-. Por último, los programadores crean un prototipo de sistema capaz de representar ese conocimiento hacia el exterior y utilizarlo en casos reales.

3.2 Modelos de sistemas expertos jurídicos

Para analizar los distintos modelos de sistemas expertos jurídicos se ha de dejar claro de antemano qué es un modelo. Podemos definir un modelo como “la representación dirigida a producir una acción y/o verificar una hipótesis.”²³ De esta manera, un modelo

²¹ Martínez Bahena, G.C. (2012). “La inteligencia artificial y su aplicación al campo del derecho.” *Alegatos* (núm 82) p. 828-846 Recuperado de: <http://www.corteidh.or.cr/tablas/r30570.pdf>

²² Bourcier, D. (2003). “Inteligencia artificial aplicada al derecho”. Barcelona (España). Pompeu Casanovas (Editorial UOC) p. 72

²³ Bourcier, D. (2003). “Inteligencia artificial aplicada al derecho”. Barcelona (España). Pompeu Casanovas (Editorial UOC) p. 108

lo que nos aporta es la capacidad de representar los fenómenos que ocurren en el mundo clasificándolos de forma abstracta. H.L.A. Hart nos explicaba qué es un modelo con una sencilla historia:²⁴

Un padre trata de enseñar a su hijo que al entrar a la iglesia hay que quitarse el sombrero. El niño puede aprender este modelo de conducta de dos maneras:

- 1- El padre dice: "Cualquier hombre y cualquier joven deben quitarse el sombrero al entrar en una iglesia." Este enunciado nos lleva a establecer una norma general del tipo "Si entras a la iglesia entonces debes quitarte el sombrero"
- 2- El padre, al entrar a la iglesia y quitarse el sombrero dice: "Observa, ésta es la forma adecuada de comportarse en estas circunstancias".

En la primera forma de aprendizaje se lleva a cabo el modelo de la regla y en el segundo el modelo del ejemplo. Así pues, vemos como lo que trata de hacer un modelo es construir una generalización, es decir, reducir los fenómenos que ocurran en el mundo a una forma de actuar o de resolver los problemas que se plantean. Dicho esto, se han expuesto previamente las clases de sistemas expertos en general, más a su vez encontramos varios tipos de modelos de sistemas expertos jurídicos:

- **Sistemas expertos jurídicos basados en reglas de producción y el modelo de sistemas expertos jurídicos positivista explícito subyacente**: estos sistemas son muy simples, se basan en la reducción de todos los problemas jurídicos a la fórmula lógica "si p entonces q" ($p \rightarrow q$), de manera que la base de datos del sistema identifica los supuestos normativos, a la vez que le son introducidos los supuestos fácticos, y el sistema deduce una conclusión (q) a partir de la subsunción del supuesto fáctico concreto dentro de los supuestos normativos previstos. Las objeciones o críticas a este tipo de sistemas son evidentes. En primer lugar, se pueden dar casos fácticos no contemplados por la norma, tras lo cual la máquina no podría encontrar solución. Además, la máquina habría de enfrentarse a los grandes problemas interpretativos que genera la textura abierta del lenguaje, sumado al hecho de que la legislación cambia constantemente, y tras cada cambio legislativo habría de reescribirse toda la base de datos. También es criticable la reducción de la solución de la controversia a un proceso mecánico, ya que los jueces reales para deducir conclusiones de la premisa mayor (regla) y la premisa menor (supuesto fáctico) realizan operaciones creativas o valorativas para seleccionar una u otra regla como premisa, o interpretar una premisa de manera diferente a la que en principio apunta la regla. Esto último, debido a que, por ejemplo, la regla podría ser válida en el ordenamiento jurídico pero no vigente,²⁵ es decir, que puede haberse promulgado conforme a los criterios adecuados para ser considerada regla jurídica, pero ser poco probable que dicha regla vaya a ser aplicada, debido a otros factores, como podría ser el conflicto de una ley o un artículo determinado en el caso concreto con principios del derecho que tengan más

²⁴ Bourcier, D. (2003). "Inteligencia artificial aplicada al derecho". Barcelona (España). Pompeu Casanovas (Editorial UOC) p. 107-108

²⁵ Para ahondar más en los conceptos de validez y vigencia de las normas: Bulygin, E. (1966) Sentencia judicial y creación de derecho. *La ley (núm 124)* p. 355-369

peso que la regla, caso en el cual este sistema no estaría suficientemente dotado para solucionar la contradicción de manera efectiva.

- **Modelo constructivista:** bastante más plausible en los aspectos en que los sistemas anteriores eran criticados. Este modelo de sistema experto jurídico fue desarrollado en México, en concreto por el doctor Enrique Cáceres Nieto, del Instituto de Investigaciones Jurídicas de la UNAM. Huelga decir que todo modelo de SEJ tiene en más o menos medida cierta dependencia con alguna de las concepciones del derecho, por eso, más que idear o desarrollar, lo que en realidad hacen los expertos humanos en la creación de SEJ es seleccionar una concepción del derecho plausible para ellos e intentar aplicarla al razonamiento del sistema, con la inestimable colaboración de informáticos e ingenieros. Este modelo concibe la norma jurídica no como el enunciado explícito formado por palabras, sino como el resultado de haber interpretado ese enunciado. De ese modo, el sistema en realidad copia la concepción realista del derecho. Al igual que hacen los juristas, primeramente el sistema identifica el derecho positivo – es decir, los símbolos que forman palabras que a su vez forman el enunciado jurídico – y a partir de ahí realiza diferentes interpretaciones de acuerdo a diferentes criterios dogmáticos, doctrinales o jurisprudenciales que tiene almacenados en su base de datos. Sería la actividad que según Riccardo Guastini llamamos interpretación cognitiva, esto es, la atribución de todos los significados posibles a un mismo enunciado jurídico. En este caso el sistema experto jurídico determinaría esos posibles significados que constituirían la norma en esencia mediante pautas para esa interpretación aceptadas en la comunidad jurídica. Esto nos lleva a aceptar la afirmación de que a sujetos diferentes –jueces- que se les presenta la misma realidad empírica –hechos- y positiva – enunciado normativo- pueden realizar constructos –normas- diferentes, debido a motivos irracionales.²⁶ Tal afirmación nos acerca a las tesis realistas y escépticas, y por tanto legítima a hacer plausible el famoso aforismo que dice que *“el derecho es lo que los jueces dicen que es”*, lo cual a su vez pone de manifiesto que este tipo de sistema experto jurídico es el que más se acerca a la realidad de lo que constituye la actividad de un juez.
- **Modelo conexionista:** los modelos conexionistas tratan de emular el funcionamiento biológico del cerebro. En esencia, se basan en las redes neuronales, ya explicadas anteriormente. Los sistemas que utilizan modelos conexionistas lo que en realidad hacen es crear una “micro-estructura de cognición”²⁷. Las aplicaciones jurídicas del modelo conexionista se basan en ayudar a tomar una decisión, primeramente mediante la circunscripción de la operación de juicio a determinados factores y resultados posibles. Posteriormente se formalizan todos los casos resueltos similares al problema que se ha de resolver, tras lo que se entrena al sistema dándole esos casos

²⁶ Según Guastini (2015) en “Un enfoque realista acerca del derecho y el conocimiento jurídico (p.1) a esta operación consistente en seleccionar una interpretación de la norma de entre todas las posibles, se le denomina *“interpretación decisoria”*

²⁷ Bourcier, D. (2003). “Inteligencia artificial aplicada al derecho”. Barcelona (España). Pompeu Casanovas (Editorial UOC) p. 120

resueltos, y, además, se le van dando casos nuevos. El usuario le dirá al sistema si éste va acertando en sus decisiones o no, por lo que se actualiza la base de ejemplos. Tras cada nuevo ejemplo el sistema mejora, por lo que reduce su margen de error. Hoy en día ya hay múltiples aplicaciones a tareas jurídicas que se basan en modelos conexionistas.

- **Modelo de razonamiento legal basado en casos**: este sistema, ideado por Kevin Ashley, de la Universidad de Pittsburg²⁸, parte de la premisa de que un caso es un conjunto de circunstancias empíricas que constituye una controversia que se puede resolver mediante una solución de instancia de tipo (analógicamente). El ser humano suele recuperar información similar de experiencias anteriores para resolver conflictos, de manera que adapta la información de un caso similar al caso nuevo. Por ello, este sistema requiere un esquema de representación de hechos, un esquema de comparación para evaluar las semejanzas relevantes de los casos anteriores con el caso actual, y un mecanismo para comparar casos y obtener premisas creadas por analogía.²⁹ El sistema dispone de una gran base con casos resueltos y cuando se le presenta un nuevo caso realiza una búsqueda en la base para ver si encuentra alguno similar, de manera que analiza cuánto se parecen ambos casos y, mediante la aplicación de operadores de modificación sobre la solución del caso comparado, puede obtener una solución para el caso nuevo.
- **“Split-up”, o sistema experto basado en reglas y en redes neuronales**: ideado por Andrew Stranieri en la Universidad de Melbourne. Tiene una aplicación muy concreta, que es la de asesorar jurídicamente sobre la forma en que se deben distribuir los bienes después de la disolución del vínculo matrimonial, así como determinar la guarda y custodia de los menores, y actúa catalogando y ordenando jerárquicamente los factores que se utilizan en Australia para determinar la distribución de bienes y la guarda y custodia. Esta jerarquía se determinó con la ayuda del conocimiento de expertos que tuvieron en cuenta factores como las posibilidades de supervivencia futura de cada cónyuge, las obligaciones económicas de los mismos, los recursos financieros, la contribución de los cónyuges a las tareas domésticas mientras duró el matrimonio. Los programadores de este SEJ estudiaron la jurisprudencia australiana relevante, y de ahí fueron insertando al SEJ los valores que obtenían de cada sentencia, generando éste la jerarquía de los factores a tener en cuenta para la disolución del matrimonio. Este sistema puede ser de gran utilidad tanto como para mediadores – en su labor de evaluar la jerarquía de los datos relevantes de cada parte para llevar su tarea de mediación-, como para abogados – para explorar escenarios hipotéticos argumentando de una u otra manera- e incluso jueces – para tener más información fiable y poder llegar a una sentencia equitativa.

²⁸ Alevan, V (2003) “Using background knowledge in case-based legal reasoning: a computational model and an intelligent learning environment”. *Artificial Intelligence*.

²⁹ Alevan, V (2003) “Using background knowledge in case-based legal reasoning: a computational model and an intelligent learning environment”. *Artificial Intelligence*.

4. Cuestiones éticas de la inteligencia artificial

Con la inteligencia artificial ha surgido la denominada “*roboética*”, que es el conjunto de criterios o teorías con los que se pretende dar respuesta a los problemas éticos que plantea la creación y el uso de robots, y que se proyectan en sus fabricantes y usuarios e, incluso, en los propios robots.³⁰ En las próximas líneas se tratará de poner sobre la mesa distintas cuestiones que atañen a la disciplina de la ética relacionada con las vicisitudes que puede traer consigo la tecnología y más en concreto, la inteligencia artificial y la robótica. Pese a que en un primer contacto con algunas cuestiones pueden parecer futuristas en exceso o incluso de ciencia ficción, lo cierto es que están más cerca que nunca y por eso hemos de estar preparados y poder ofrecer respuestas satisfactorias ante ellas.

4.1 La moral ante el progreso tecnológico

Fue el escritor de ciencia-ficción Isaac Asimov quien, en 1942, en un relato titulado “*Runaround*” dentro de su libro “*Yo, robot*”, ideó las 3 leyes de la robótica.³¹ Posteriormente, ha habido diversos intentos de perfeccionar esas leyes, de añadir otras nuevas y de regular a nivel ético la relación entre humanos y máquinas. En todo caso, es necesaria una regulación jurídica lo antes posible, para lo cual hay que reflexionar sobre las pautas a establecer para solventar los problemas que surgen a partir del avance de la tecnología.³² Para ello, lo esencial es determinar el marco de referencia ético desde donde afrontar el conflicto. En este sentido, es posible destacar 4 posiciones doctrinales sobre cómo afrontar los avances tecnológicos desde el punto de vista ético:³³

- A) Bioconservadurismo:** esta posición defiende la idea de mantener la identidad humana y su dignidad, es decir conservar siempre y bajo toda circunstancia aquello que nos hace humanos. Se considera desde esta perspectiva que el avance de la tecnología entraña peligros para la humanidad y su supervivencia, y se prefiere preservar las cualidades individuales del ser humano a toda costa. Se considera que el acto de crear vida o de imitar la vida humana es antinatural y, por tanto, siempre negativo. Un eventual auge de la tecnología podría hacer perder la noción de la ética a las personas. Se considera que el sentido de la vida yace en parte en nuestras limitaciones y dejaría de tener sentido un mundo

³⁰ De Asís, R. (2013). “El marco ético de la robótica”. *Huri-age. El tiempo de los derechos*. Núm 8. Recuperado de: http://earchivo.uc3m.es/bitstream/handle/10016/19290/marco_asis_PTD_2013.pdf?sequence=1

³¹ Leyes de la robótica de Isaac Asimov: 1- *Un robot no puede hacer daño a un ser humano ni directamente ni a través de su intervención*. 2.- *Un robot debe obedecer las órdenes que le den los seres humanos, a menos que las mismas entren en conflicto con el primer mandamiento*. 3.- *Un robot debe salvaguardar su propia existencia, a menos que su autodefensa se contradiga con el primer y el segundo mandamiento*

³² Así, la “roboética”, o de manera más general, la “tecnoética”, pretenden aunar las distintas tesis que existen acerca de las pautas morales del ser humano ante el fenómeno robótico y tecnológico, de la misma manera que sucede con otras disciplinas seguramente más conocidas, como la bioética, que se dedica a proporcionar los principios para la conducta más adecuada del hombre en relación a la vida.

³³ De Asís, R. (2013). “El marco ético de la robótica”. *Huri-age. El tiempo de los derechos*. Núm 8. Recuperado de: http://earchivo.uc3m.es/bitstream/handle/10016/19290/marco_asis_PTD_2013.pdf?sequence=1

donde éstas no existieran. Esta concepción de la ética respecto al progreso reclama siempre la máxima precaución y el establecimiento de límites sólidos a toda invención que pueda modificar el estilo de vida en comunidad de las personas concebido hasta hoy en día. Esta posición, como es obvio, se opondría a los sistemas expertos jurídicos de ayuda en la toma de decisiones judiciales, por ejemplo.

B) Neurodeterminismo: este planteamiento identifica la moral con la biología. Nuestra estructura ética, y por tanto también biológica, nos hace animales únicos de entre todos los del planeta tierra. Dentro de esta corriente hay varias posiciones. Hay quienes consideran que no existe el libre albedrío tan típico de los humanos, ni la autonomía de éstos, sino que el “yo” es simplemente un estado del cerebro. Esta concepción basaría su ética en los datos cerebrales, al contrario que la ética predominante durante toda la historia de la humanidad, basada en que las personas sanas, en un estado normal, son responsables de sus acciones y podemos orientar su conducta hacia uno u otro lugar. Para esta posición, no obstante, el comportamiento humano no sería únicamente innato, sino que también estaría influenciado socialmente. Varias de estas posiciones ya han tenido su repercusión en el derecho, cuestionando aspectos jurídicos tan básicos para la vida en sociedad como la responsabilidad penal.³⁴

C) Transhumanismo:³⁵ esta concepción pretende aprovechar la tecnología para hacer más próspera la humanidad, construyendo seres humanos más capaces, con vidas más cómodas y felices. El transhumanismo defiende que la tecnología ha de permitir al ser humano luchar para sobrevivir, e incluso posturas extremistas pretenden llegar a un punto de fusión entre máquina y ser humano. Defiende también la difusión abierta de todos los avances potenciales y de los objetivos que se pueden lograr con ellos, para que cada generación de humanos decida qué avances se aplican y cuáles no. En esta corriente existen diferentes grados, habiendo posiciones más moderadas y otras más progresistas, existiendo personas que sostienen que el ser humano debe abandonar los rasgos característicos que lo hacen débil y utilizando los rasgos de las máquinas que más les benefician. Otras posiciones tratan de conciliar el imparable avance de la tecnología con los valores propios de la Ilustración, defendiendo que los individuos son más felices cuanto más margen de autonomía tienen. Otras enfocan el fin de todo avance en el prolongamiento de la vida y calidad humana físicamente hablando. El filósofo británico David Pearce aporta dos argumentos clave en favor del transhumanismo: superar el dolor físico y mental y elegir el grado de satisfacción en la vida humana.³⁶

³⁴ Para profundizar en un tema tan interesante como el determinismo se recomienda leer a González Lagier, D. (2016) “Sobre los presupuestos filosóficos de las pruebas neurocientíficas (A propósito de «Proceso y neurociencia. Aspectos generales» de M. Taruffo)” *Jordi Ferrer y Carmen Vázquez, Debatiendo con Taruffo*. Marcial Pons.

³⁵ El primero en utilizar el término “*transhumanismo*” fue el biólogo Julian Huxley en 1957.

³⁶ De Asís, R. (2013). “El marco ético de la robótica”. *Huri-age. El tiempo de los derechos*. Núm 8. Recuperado de: http://earchivo.uc3m.es/bitstream/handle/10016/19290/marco_asis_PTD_2013.pdf?sequence=1

Esta corriente, no obstante, se enfrenta a fuertes críticas. Por un lado, caemos en el peligro de construir una sociedad en la que el mérito y el esfuerzo –valores que también son característicos de la realización personal de un individuo- sean abandonados. Además, otro gran peligro es el aumento de la brecha social de desigualdad. Es decir, que la tecnología más avanzada solamente llegue a los ricos. Dicho argumento es fácilmente sostenible echando una ojeada a la historia de los avances tecnológicos en la humanidad, donde la brecha de derechos y calidad de vida entre el que tiene y el que no tiene poder adquisitivo se ha mantenido a pesar de las innovaciones. También es criticable la potencial destrucción de la identidad humana y su dignidad.

En conclusión, la posición ética del transhumanismo defiende que el valor moral del ser humano está contenido en lo que hace y no en los rasgos propios de su especie. Para defender esta posición se suele utilizar la falacia circular que afirma que el progreso para la humanidad no puede ser malo moralmente, ya que mejora al ser humano en conjunto y, por lo tanto, también mejora su moral.

D) Humanismo abierto desde los derechos humanos: examinadas las tres concepciones anteriores, se puede concluir que todas ellas tienen algo de verdad, pero ninguna de ellas en conjunto parece tener una argumentación completa y convincente sobre cómo actuar respecto al progreso de la inteligencia artificial. Así pues, lo más plausible es quedarse con las razones más convincentes de cada manera de pensar. Del bioconservadurismo podemos quedarnos con la necesidad de establecer siempre una línea, una que no debemos cruzar si no queremos perder todo aquello que nos hace humanos. Así, lo más adecuado sería tener siempre presente los derechos humanos. No obstante, existen varias concepciones acerca de los derechos humanos, así como diferentes formas de dar contenido e interpretar cada uno de los derechos que conformarían ese conjunto. Aun así, podemos aislar un substrato esencial en toda teoría de los derechos humanos: la colocación del ser humano en el centro de trascendencia de todo suceso empírico, procurando en la medida de lo posible una vida digna para todas las personas. De esta manera, se pueden considerar como criterios limitadores del poder tecnológico al menos la autonomía y la satisfacción de las necesidades –en especial de las necesidades vitales básicas- de las personas. En otras palabras, el progreso siempre ha de quedar limitado allí donde invada el ámbito mínimo de autonomía de las personas que libremente quieran ejercerla, y también debe restringirse en la medida en que no sea útil para las necesidades humanas. El concepto de utilidad, en este caso, se utiliza como utilidad positiva y generalizada –en el sentido utilitarista de Bentham y Mill³⁷- es decir, que la innovación sólo será útil si sirve para mejorar las vidas de un número de población considerable, ya que, si nos refiriéramos al concepto de útil en sentido estricto, nos sería imposible encontrar ningún avance humano que no tenga una aplicación determinada, sea ésta benévola o no y se aplique a muchos o a pocos individuos.

³⁷ Jeremy Bentham expone su doctrina utilitarista en “Introducción a los principios de moral y legislación” (1789) y su discípulo utilitarista, John Stuart Mill continúa con esa doctrina en “El utilitarismo” (1863).

Ahora bien, esta teoría de los derechos humanos ha de complementarse con la idea de progreso de las tesis neurodeterministas y transhumanistas, dando el máximo protagonismo posible al progreso siempre que cumpla con los límites mencionados. Esta teoría debería aplicarse con la mente puesta en el futuro y en las generaciones venideras, pues la noción de progreso y de vida digna varía sustancialmente si, en vez de pensar en nuestra muerte como el fin de todo, actuamos con la idea de que la vida en nuestro planeta se alargue y prospere todo lo posible.

Finalmente, esta teoría no debe olvidarse de incluir los conceptos de diversidad y discapacidad en la construcción del concepto de vida digna. Esto significa que, desgraciadamente, muchos de los avances tecnológicos se hacen pensando en un estándar determinado de persona o en personas que lleven un determinado estilo de vida, sin tener en cuenta a las personas con discapacidad o a aquellas personas que eligen vivir de un modo diferente. Imaginemos una tribu indígena del Amazonas a la que, hipotéticamente, y en aras de un supuesto progreso, se le decidiera cambiar su modo de vida e introducirles ciertas mejoras tecnológicas. Este ejemplo nos suscita cierto reproche moral, y precisamente por eso se debe aplicar la noción del respeto a la diversidad en todo cambio tecnológico, pues seguramente la noción de vida digna de esa tribu indígena incluye vivir del modo en el que lo están haciendo hasta el momento, y por tanto habría de concedérseles una autodeterminación suficiente para decidir si quieren formar parte de la globalización tecnológica o no. Respecto a la discapacidad, imaginemos que, llegado cierto momento, se decide llevar a cabo métodos eugenésicos para mejorar los genes de las personas y así evitar que nazcan a partir de ese momento personas con síndrome de Down.³⁸ En este ejemplo vemos cómo surge un espinoso debate que confronta posiciones transhumanistas y tesis bioconservaduristas, y en el cual la interpretación que se le dé a los derechos humanos juega un papel clave. En ese sentido, parece plausible la interpretación que se ha realizado en la Declaración de la Internacional de Personas con Discapacidad, en la que se expone literalmente: *“Queremos dejar claro que las personas con discapacidad no nos oponemos a la investigación médica que tenga por objeto un auténtico tratamiento o el alivio del dolor. A lo que nos oponemos es a la limpieza genética impulsada por motivos de rentabilidad y de eficiencia social, basada en prejuicios contra las personas con discapacidad y llevada a cabo en nombre de la curación o el tratamiento”*. Habiendo visto las 4 concepciones morales más relevantes ante el fenómeno tecnológico, me centraré en dar cuenta de algunos de los más importantes desafíos que la inteligencia artificial supone para filósofos y juristas.

³⁸ El “Institut d’Estudis Catalans” define la eugenesia como la “teoría que postula que la constitución genética de los humanos se puede perfeccionar mediante el control de su reproducción”. Es decir, la eugenesia supone la modificación de los genes humanos en aras a la supresión de las discapacidades y defectos que se puedan dar en la gestación.

4.2 Implantación de la inteligencia artificial de modo seguro

La cada vez más cercana implantación de los sistemas de inteligencia artificial o de robots de forma permanente en nuestra sociedad hace surgir una serie de problemas, a los cuales hay que aportar una serie de precauciones para combatirlos. Imaginemos, por ejemplo, que tuviéramos un sistema de inteligencia artificial encargado de toma de decisiones judiciales menores, como por ejemplo, en procesos monitorios de poca cuantía. Pongamos que ese sistema, casualmente o no, raramente fallara a favor de personas de origen magrebí, de modo que cuando en un caso de reclamación de cantidad una de las partes fuera magrebí, el sistema fallara en la mayoría de los casos a favor de la contraparte. ¿Qué haríamos ante estas estadísticas? ¿Puede el algoritmo del sistema fallar y, por tanto, ser totalmente discriminatorio y por ende, inmoral e ilegal? ¿Podría algún experto en estos sistemas hackear dicho algoritmo para modificarlo a su antojo? Ante estos eventuales problemas es necesario tomar las máximas precauciones antes de la implementación de estos sistemas tanto al derecho como a otras áreas sociales que requieren tareas cognoscitivas. Por eso, sería conveniente que se tuvieran en cuenta al menos las siguientes medidas:³⁹

- A) Control e inspección de los sistemas mediante expertos humanos que puedan auditar o inspeccionar el correcto funcionamiento de dichos sistemas periódicamente.
- B) Los algoritmos también deben ser susceptibles de inspección y ser periódicamente inspeccionados.
- C) Predictibilidad y previsibilidad. Esta propiedad trasladada al derecho vendría a convertirse en el principio de legalidad con todas sus manifestaciones, lo que significa en su esencia seguridad jurídica, es decir que todos los ciudadanos sepan cuáles son los criterios mediante los cuales estos sistemas actúan y toman decisiones, para que los ciudadanos puedan tener un plan de vida más seguro y previsible. Es posiblemente la garantía más importante que ha de cumplirse.
- D) Incorruptibilidad. Los algoritmos de los sistemas han de ser inexpugnables ante manipulaciones, es decir, lo más seguros posibles ante injerencias externas en los mismos con fines perversos.
- E) Responsabilidad por los fallos del sistema. Cuando el sistema inteligente comete un error que tiene consecuencias para las personas y sus derechos han de establecerse parámetros para determinar con claridad quién es el responsable —el diseñador, el fabricante de los componentes, el encargado de su ensamblaje, los usuarios, etc.

Habiendo visto las pautas básicas para una implantación segura de la inteligencia artificial, abordaremos algunas cuestiones relacionadas con ésta que conllevan cierta problemática ética.

³⁹ Bostrom, N. y Yudkowsky, E. (2011) "The Ethics of Artificial Intelligence". *Cambridge Handbook of Artificial Intelligence*. William Ramsey and Keith Frankish (Cambridge University Press) Recuperado de: <http://www.nickbostrom.com/ethics/artificial-intelligence.pdf>

4.3 Adquisición de derechos y deberes por parte de máquinas y robots

Otro tipo de problemas surgen cuando tratamos de situarnos en un contexto futuro donde las máquinas estén desarrolladas de tal manera que nos tengamos que plantear otorgarles un estatus moral. Respecto a este último concepto, la filósofa estadounidense especializada en ética aplicada, Frances M. Kamm, sugirió la siguiente definición:⁴⁰

X tiene estatus moral = dado que X es consciente moralmente de su propio derecho, es permisible realizar acciones para su bien.

Por un lado, está claro que los humanos tienen un estatus moral. Por otro, también está claro que un palo de madera no tiene un estatus moral. El estatus moral, pues, es la cualidad que convierte en legítimas ciertas acciones ante dicho sujeto –relacionarse con él, por ejemplo- y en ilegítimas otras acciones –dañarlo, matarlo-. Algunos paralelismos sencillos acerca del problema del estatus moral de las máquinas sería el estatus moral de un embrión en un debate acerca de la ética en el aborto, o el estatus moral de los animales cuando se experimenta con ellos, o incluso el estatus moral de una persona en estado vegetal ante la perspectiva de la eutanasia. Actualmente es ampliamente aceptado por la comunidad científica, filosófica y jurídica que los sistemas de inteligencia artificial no disfrutan de un estatus moral. Es por ello que podemos utilizarlos con absoluta libertad y sin límites intrínsecos, es decir, sólo hay límites en su uso en cuanto puedan afectar a los derechos de un ser humano. Aunque actualmente no se les otorgue estatus moral, no están claros cuáles son los criterios que deberían establecerse para determinar cuándo un individuo –humano o no- es susceptible de tener estatus moral. Se suelen proponer los siguientes dos criterios:⁴¹

- a) **Sensibilidad:** Esto es, la capacidad del individuo de reaccionar a fenómenos mediante la capacidad de sufrir o sentir dolor, físico o emocional.
- b) **Sapiencia:** El conjunto de actividades relacionadas con la inteligencia superior – la humana- como la autoconciencia y la capacidad para comprender las consecuencias de sus actos.

En el caso de los animales se ha reconocido que tienen sensibilidad y, por tanto, en cierta medida tienen estatus moral, pero, por otra parte, no tienen sapiencia, lo que hace que sólo los humanos tengan el estatus moral más elevado. Lo mismo podríamos decir de los embriones –aunque científicamente mucho más discutible y controvertido- o de personas con discapacidades mentales muy severas. Se han establecido dos criterios secundarios para abarcar aún más los individuos susceptibles de tener estatus moral. Estos serían la pertinencia de un individuo a un colectivo que un su generalidad sí tiene

⁴⁰ Tal y como se cita en: Bostrom, N. y Yudkowsky, E. (2011) “The Ethics of Artificial Intelligence”. *Cambridge Handbook of Artificial Intelligence*. William Ramsey and Keith Frankish (Cambridge University Press) Recuperado de: <http://www.nickbostrom.com/ethics/artificial-intelligence.pdf>

⁴¹ Bostrom, N. y Yudkowsky, E. (2011) “The Ethics of Artificial Intelligence”. *Cambridge Handbook of Artificial Intelligence*. William Ramsey and Keith Frankish (Cambridge University Press) Recuperado de: <http://www.nickbostrom.com/ethics/artificial-intelligence.pdf>

estatus moral –los humanos- y la existencia de una relación que vincule al individuo en cuestión con otro individuo que tenga por sí solo estatus moral.⁴² De ese modo, aunque una persona con discapacidades mentales severas no llegue a tener la cualidad de sapiencia, sí que tiene un estatus moral equivalente a una persona sin esas discapacidades por el hecho de pertenecer al colectivo de la humanidad. Por otro lado, si el individuo tiene una relación afectiva con una persona que por ella misma tiene un estatus moral completo, ese individuo, según el criterio anterior, debería tener también el mismo estatus moral. Este último criterio es más difícil de sostener moralmente, pues la existencia de un término tan subjetivo como “relación afectiva” no debería modificar el estatus moral de un individuo. Un ejemplo sería la relación entre una persona y su mascota. El hecho de que la persona tenga un estatus moral completo y ésta ame a su mascota, no puede implicar que automáticamente la mascota goce exactamente del mismo estatus moral que el humano. Tal implicación hipotética nos llevaría a absurdos tales como no saber qué bien jurídico tiene más relevancia o más valor en un eventual conflicto, si la vida de una persona o la vida de un animal. Obviamente, eso no quiere decir que la integridad física –e incluso moral- de los animales no hayan de ser protegidas ética y legalmente ante cualquier lesión, si no que no podemos equiparar tal protección a la que recibe la especie humana. Toda esta argumentación se torna moralmente compleja, dado el afecto que siente la comunidad en general por los animales domésticos, pero si extrapolamos el argumento a un sistema inteligente con las mismas características, es decir, que no tenga sapiencia pero mantenga una suerte de relación afectiva con una persona humana⁴³, advertimos fácilmente que no es plausible otorgar un estatus moral en el mismo nivel que el otorgado a los humanos a un sistema de inteligencia artificial que no tiene sapiencia ni sensibilidad, sólo porque su propietario haya trabado amistad, o incluso romance con él.

Otra cosa a tener en cuenta es el hecho de que no es plausible imponer un criterio de procedencia a un individuo para determinar su estatus moral, es decir, que no es adecuado negarle el estatus moral a un sujeto *sólo* porque este haya sido diseñado y fabricado artificialmente. Este principio de no discriminación no implica que haya de otorgársele estatus moral a todo sistema de inteligencia artificial, sino ponderar adecuadamente teniendo en cuenta los criterios de la misma manera que ponderaríamos para cualquier otro individuo, incluido los humanos. Así pues, en un futuro a corto o medio plazo, debemos abrir un debate ético, cuyos resultados determinaran una u otra actuación jurídica, pues otorgar estatus moral a un individuo o colectivo de individuos que antes no lo tenían ha de comportar un cambio en el tratamiento jurídico de dichos individuos, y que por tanto ha de plasmarse en el derecho positivo.

⁴² Bostrom, N. y Yudkowsky, E. (2011) “The Ethics of Artificial Intelligence”. *Cambridge Handbook of Artificial Intelligence*. William Ramsey and Keith Frankish (Cambridge University Press) Recuperado de: <http://www.nickbostrom.com/ethics/artificial-intelligence.pdf>

⁴³ Para los más escépticos, se recomienda el visionado de la película “Her”-, Película norteamericana dirigida por Spike Jonze y protagonizada por Joaquin Phoenix que relata las vivencias de un hombre que mantiene una relación afectiva con su teléfono móvil, dotado de un sistema de inteligencia artificial.

4.4 Sistemas de inteligencia artificial más inteligentes que los humanos

Irving J. Good ideó el concepto hipotético de la “explosión de inteligencia”.⁴⁴ Este concepto se refiere a un hipotético futuro tecnológico en que los sistemas de inteligencia artificial sean conscientes de su propia naturaleza y puedan a su vez, rediseñarse y crear otros sistemas más inteligentes aún, y así sucesivamente. Esta hipótesis circular no sólo sería aplicable a las máquinas, sino también a las personas. Imaginemos una persona que pudiera incrementar su cociente intelectual mediante un programa informático. Llegado el momento, podría utilizar su inteligencia mejorada para crear un programa que la mejorase aún más. También sería posible un escenario en el que se consiguiera aumentar radicalmente la velocidad en la que piensa un humano, aumentando la velocidad a la que son capaces de trabajar sus redes neuronales. Entonces, un milenio entero de pensamiento para un humano normal equivaldría al proceso cognitivo llevado a cabo en pocas horas por un humano con la inteligencia acelerada.

Así pues, es innegable que la posibilidad de “superinteligencias” presenta varios desafíos éticos. Ahora bien, como se ha dicho, la moral predominante de la humanidad en una época determinada puede variar y ser diferente en otra época. Por ejemplo, hace siglos en prácticamente todo el mundo no se consideraba inmoral la esclavitud, o hace no tantos años en muchos países del mundo las mujeres aún no tenían derecho a voto. ¿Quién nos asegura que las civilizaciones venideras no reprobaran nuestra moral en muchas cuestiones que a nosotros nos parecen aceptables? Al ser estos valores tan mutables en el tiempo, parece que la cuestión radica en programar los sistemas de inteligencia artificial con capacidad de discernir entre el bien y el mal según los estándares morales vigentes en cada coyuntura de la historia. Esto último resulta una obviedad, si se tiene en cuenta que un sistema de inteligencia artificial va entrenando o mejorando su ratio de acierto en la resolución de problemas a medida que se le presentan más, de manera que esa es su capacidad de aprendizaje. Así pues, si el sistema está programado para interpretar datos y variables con unos patrones morales determinados –en teoría, los vigentes en cada momento de la historia- a medida que estos patrones cambien, el sistema debería advertir esos cambios y por tanto reajustar su ética a la de las personas humanas.

4.5 Problemas sociales estructurales a raíz de la inteligencia artificial

A nivel ético, amén de los problemas que ya se han planteado, existen otros inconvenientes a nivel de la humanidad en su conjunto que *a priori* nos hacen preguntarnos si realmente merece la pena la implementación de la inteligencia artificial con tanta intensidad como nos permiten los avances tecnológicos. En estos casos, conviene tener un prisma que nos permita ver los problemas, las posibles soluciones si es que las hay, y a partir de ahí hacer una evaluación moral acerca de la conveniencia o no de tal implementación, o de qué sectores o ámbitos de nuestras vidas deberían

⁴⁴ Good, I.J. (1965) “Speculations Concerning the First Ultra-intelligent Machine” *Advances in computers*. p.31-88

permanecer vírgenes ante la penetración de la inteligencia artificial. A continuación, daré una lista no exhaustiva de todos los problemas que se pueden dar, trata tan solo unos cuantos ejemplos, los que considero más relevantes.

- **Desempleo.** Una de las grandes preocupaciones de los estados acerca del impacto de la inteligencia artificial es el desempleo. Hay economistas que mantienen una perspectiva negativa respecto al impacto de la inteligencia artificial en la tasa de desempleo, la cual obtiene una respuesta por parte de ciertos futurólogos e investigadores que afirman que, en efecto, si bien se perderán muchísimos puestos de trabajo, estos serán reemplazados por otros tantos empleos que requerirán menor esfuerzo, menos horas y más remuneración para los trabajadores. En cualquier caso, es conveniente que todos los estados a todos los niveles institucionales y territoriales tratasen de estudiar estadísticamente el impacto en cada sector para tratar de solucionar los problemas de desempleo. Se abordará el desempleo desde una perspectiva de conflicto jurídico más adelante.⁴⁵
- **Tecnoadicción y tecnodependencia.** En una sociedad cada vez más dependiente de la tecnología y adicta a ella, sería acertado tratar de establecer mecanismos en todos los ámbitos que restrinjan el uso de la tecnología en ciertos sectores y fomenten también valores no tecnológicos. Fundamentalmente son los legisladores en materia educativa los que más cuidado y responsabilidad tienen en el proceso de construir una sociedad que, aunque dependa de la tecnología, no subordine esa dependencia a su capacidad para ser feliz y desarrollar un plan de vida. Las leyes de educación de todos los estados deberían incluir la tecnología en los métodos docentes en los cuales se haya demostrado ya el impacto positivo de aquélla. Sin embargo, se debería formar a las personas de manera que sepan que no siempre ha sido así, y que existen objetos, actividades de ocio y fuentes de información que no tienen un componente tecnológico avanzado.
- **Seguridad.** Con la afición a la seguridad nos referimos a la lesión del derecho a la vida y a la integridad física y moral, que en nuestro ordenamiento está recogido como derecho fundamental en el art. 15 CE. En el apartado siguiente se examinarán algunos principios que ya se han propuesto como base para evitar daños de los sistemas inteligentes a los derechos de las personas.⁴⁶
- **Afectación a los derechos fundamentales y derechos humanos.** Actualmente ya hay un incipiente debate acerca de la dudosa moralidad de algunos avances tecnológicos que permiten realizar acciones que penetran en ciertos ámbitos de privacidad de las personas que hasta ahora eran inexpugnables. El famoso “Big Data”, por ejemplo, que permite recolectar los datos de navegación en la red de usuarios para utilizarlos a la hora de trazar un perfil sociológico mediante el cual ofrecernos el producto más adecuado. La legalidad vigente permite que los operadores económicos utilicen nuestros datos personales –qué páginas web visitamos o a qué le damos “like” en las redes sociales- para bombardearnos y

⁴⁵ Infra p. 45

⁴⁶ Infra p. 27

convertirnos cada vez más en zombis del consumismo. Por otra parte, la inteligencia artificial permite ya –y con expectativas de mayor exactitud- poner en marcha acciones como el reconocimiento facial, es decir dispositivos, por ejemplo drones, que vuelan por el entorno y son capaces de reconocer los rostros de las personas, mientras simultáneamente alguien dirige el aparato. Si pueden reconocer caras, con mayor facilidad pueden grabar conversaciones e incluso identificar a los participantes en ellas. En este sentido, aunque no es inteligencia artificial, hay que mencionar la cada vez más estandarizada geolocalización –por ejemplo en los teléfonos móviles-, que permite a prácticamente cualquier persona saber dónde está cualquier otra en cualquier momento.

Todas estas utilidades tecnológicas y algunas otras chocan de lleno con nuestra concepción de algunos derechos fundamentales reconocidos en la Constitución Española y derechos humanos reconocidos en la Declaración Universal de Derechos Humanos, como el derecho al honor, a la intimidad personal y a la propia imagen, el derecho a la vida y a la integridad, el derecho a la inviolabilidad de la vivienda o el derecho al secreto de las comunicaciones. En ese sentido, el artículo 18.4 CE establece que *“la ley limitará el uso de la informática para garantizar el honor y la intimidad personal y familiar de los ciudadanos y el pleno ejercicio de sus derechos”*, artículo que se encuentra desarrollado legalmente por la Ley 34/2002, de 11 de julio, de Servicios de la Sociedad de la Información y de Comercio Electrónico, así como por la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal. Dicha ley otorga al ciudadano los derechos de oposición, acceso, rectificación y cancelación de los datos de carácter personal. No obstante, y debido a la creciente masividad de datos de carácter personal que cada día se generan y almacenan, provocada por el auge tecnológico –incluido el de la inteligencia artificial- desde el ámbito europeo se ha pretendido regular este asunto más eficazmente. Por ello, el próximo 25 de mayo de 2018 entrará en vigor el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos). Al ser un Reglamento, tiene aplicación directa y por tanto los estados de la UE tendrán muy poco margen para legislar a su gusto. En líneas generales, se amplía el concepto de “datos de carácter personal”, incluyendo en éste datos identificativos de la actividad de una persona en la web, como las *cookies* o la dirección IP. Además, se regulan nuevos derechos, como el derecho de portabilidad de los datos de un prestador a otro, el derecho al olvido integrado en la cancelación o la supresión de datos, y se obligará a la mayoría de las empresas a contar con un delegado de protección de datos en plantilla.⁴⁷ Habiendo visto algunos problemas morales que puede generar la inteligencia artificial, abordaré ahora una serie de

⁴⁷ Para profundizar más en la protección de datos, y en concreto en las novedades del nuevo Reglamento Europeo se recomienda la lectura del blog del doctor Pere Simón Castellano “Derecho al olvido y protección de datos”. Recuperado de: <http://blog.peresimon.com>

principios expuestos en una conferencia científica y que pueden ser de gran utilidad para conformar el marco ético de la mencionada tecnología.

4.6 Los 23 principios de Asilomar

Continuando con los aspectos éticos y morales, en enero de 2017 se celebró en Asilomar (California) una conferencia organizada por el Future of Life Institute⁴⁸ en la cual se establecieron 23 principios o recomendaciones cuyo objetivo es minimizar los riesgos que puede generar la inteligencia artificial y lograr aprovechar su increíble potencial para beneficiar a la humanidad y no para dañarla. Dichos principios han sido apoyados por más de 1200 científicos expertos en el campo de la inteligencia artificial, entre ellos nombres de gran calado como Stephen Hawking o Elon Musk. Esta serie de estándares o directrices son clave porque son los que determinarán las actualizaciones que debemos implantar en nuestros ordenamientos jurídicos. Dicho de otra manera, estos principios morales, por la enjundia y prestigio de los expertos que los han consensuado, además de por ser bastante acertados, podrían servir como guía a partir de la cual confeccionar las reformas jurídicas. A continuación, se expondrán y desarrollaran brevemente los principios consensuados en dicha conferencia, los cuales se dividen en tres apartados. El primero de ellos tiene que ver con los principios a seguir para maximizar los beneficios de la investigación en inteligencia artificial. El segundo tiene que ver con los principios y valores que deberían estar presentes en el diseño e implementación de la inteligencia artificial. El tercero está referido a los riesgos que conlleva dicha tecnología y las precauciones necesarias para superar dichos riesgos.

Conflictos en la investigación

- 1- **Objetivo de la investigación:** La inteligencia artificial no debe desarrollarse sin dirección alguna, sino que debe enfocarse como un medio para conseguir un fin, el del beneficio para la humanidad.
- 2- **Financiación de la investigación:** La inversión en inteligencia artificial es básica para potenciar su capacidad de generar beneficios y minimizar su capacidad de daño. Así, es clave el apoyo financiero de estados y organizaciones internacionales para resolver cuestiones espinosas como las posibles manipulaciones en los sistemas inteligentes o malos funcionamientos en los mismos, o adaptar nuestros ordenamientos jurídicos para favorecer que esta tecnología crezca de manera segura.
- 3- **Conexión entre la ciencia y la política:** Sería muy importante una conexión simbiótica entre políticos e investigadores, de manera que los primeros marquen las directrices y principios que han de conformar el marco jurídico y los segundos

⁴⁸ El Future of Life Institute (FLI) es una organización de expertos científicos voluntarios dedicada a la investigación y desarrollo de soluciones frente a los riesgos existenciales a los que se enfrenta la humanidad con el avance de las tecnologías, en especial frente a los riesgos que puede suponer la Inteligencia Artificial. Para conocer más sobre dicha organización, se puede visitar su página web: <https://futureoflife.org/team/>

puedan trabajar y avanzar dentro de dicho marco. Este principio viene a conformar lo que se denomina deferencia hacia la ciencia, y supone que sea ésta la que lleve el peso y marque el camino del progreso de la humanidad y la política esté subordinada a los resultados que la ciencia obtenga.⁴⁹

- 4- **Cultura de la investigación:** La investigación debería desarrollarse siguiendo unos principios de buena fe, confianza mutua entre investigadores y transparencia.
- 5- **La investigación no es una carrera o competición:** Los investigadores en inteligencia artificial deben trabajar siempre poniendo por delante ciertos estándares de seguridad, y no deben tratar de ser pioneros para tener reconocimiento individual si eso conlleva renunciar al desarrollo seguro de la tecnología.

Ética y valores

- 6- **Seguridad:** Los sistemas de inteligencia artificial deben siempre ser seguros y fiables a lo largo de toda su vida útil, así como verificables. Además, su aplicación a un ámbito de la realidad debe ser factible.
- 7- **Transparencia en los errores:** Si un sistema de inteligencia artificial causa un daño a las personas y sus derechos, debería ser posible llegar a la raíz del fallo que causó el daño. Este es uno de los principios más básicos y uno de los hándicaps técnicos más complejos. Se habla desde muchos sectores de aplicación de la inteligencia artificial –como el de los coches autónomos- de la necesidad de una caja negra –como la famosa caja negra de los aviones- que detecte e informe del origen del fallo del sistema.
- 8- **Transparencia judicial:** Cualquier controversia jurídica resuelta por un sistema de inteligencia artificial debe suministrar una argumentación que debe de ser auditable por una autoridad humana competente. Este principio es otro de esenciales y que tiene mucha relevancia en el tema que aquí se trata. En efecto, parece que aun teniendo en cuenta las previsiones de futuro que implican la existencia de sistemas expertos con un margen de error en la toma de decisiones cada vez más estrecho, no podemos confiar una decisión judicial firme que implique cosa juzgada a uno de estos sistemas. Mientras exista ese margen de error, por ínfimo que sea, parece que no es plausible que tengan la última palabra estos tipos de sistemas, por las consecuencias obvias que podría comportar. Si bien es cierto que actualmente los jueces humanos se equivocan, el hecho de confiar una decisión a un sistema experto debería, mientras éste

⁴⁹ Para profundizar en la concepción deferencialista hacia la ciencia, se recomienda leer a Schauer F. (2010) "Can Science Be Good Evidence?" *Cornell Law Rev.* p.1191-1220. Además, se pueden encontrar argumentos en contra de la deferencia hacia la ciencia en Darnaculleta Gardella, M., Esteve Pardo, J. y Spiecker gen. Döhmman, I. (2015) "Estrategias del derecho ante la incertidumbre y la globalización". Marcial Pons.

tenga margen de error, acompañarse con la posibilidad de que esa decisión fuera revisable por un juez humano. El motivo de esto, no obstante, no obedece a criterios racionales, pues como se ha dicho los jueces humanos también se equivocan –y mucho- y posiblemente una decisión tomada con base en un algoritmo pudiera tener un ratio de acierto más elevado. Bien al contrario, parece que el hecho de rechazar una decisión tomada por un sistema de inteligencia artificial en última instancia obedece a nuestros prejuicios internos, que le otorgan más valor y legitimidad a una decisión de este tipo que provenga de un humano que a una emanada de un algoritmo.

No obstante, se podría argumentar de manera inversa con la misma validez, pues aunque es cierto que el sistema experto no humano también se equivoca, si éste está programado adecuadamente, jamás obrará con malicia, intereses en el asunto o prevaricación, cosa que habitualmente sí hacen los jueces humanos. De momento, como se ha explicado, parece que la aplicación de los sistemas expertos jurídicos va a ser a corto y medio plazo un sistema de ayuda o respaldo a jueces, y no el sistema autónomo al que se le confiere la potestad de decidir, situación que parece condenada a esperar muchas décadas todavía. No obstante, si ese momento llega, en opinión de los expertos reunidos en Asilomar, lo más aceptable sería establecer un sistema de auditoría humana a las decisiones finales de los sistemas inteligentes.

- 9- **Responsabilidad:** Los diseñadores y fabricantes de sistemas de inteligencia artificial son los garantes máximos de las implicaciones morales de su uso y mal uso, y tienen la capacidad de dar forma a los potenciales usos así como de restringir eventuales malos usos de los usuarios. Si analizamos este principio, aunque no lo dice explícitamente, parece que invita a establecer un sistema de responsabilidad civil en el cual se sitúe a diseñadores y fabricantes como los principales responsables.
- 10- **Alineación de valores:** Los sistemas de inteligencia artificial con un alto nivel de autonomía deben diseñarse y construirse para que sus comportamientos y metas se alineen, es decir, que sean compatibles con los valores y metas de los seres humanos.
- 11- **Valores humanos:** Los sistemas de inteligencia artificial deben ser diseñados y contruidos para ser compatibles con los valores superiores humanos de la dignidad humana, diversidad cultural y tolerancia racial.
- 12- **Privacidad personal:** Las personas deberían tener el derecho de acceso, control y gestión de los datos que generen debido a la capacidad de los sistemas de inteligencia artificial de analizar y utilizar dichos datos. En efecto, este principio viene a salvaguardar el derecho al honor y a la intimidad recogido en el artículo 18 de la Constitución Española, en nuestro ordenamiento garantizado por la Ley Orgánica 15/99, de 13 de diciembre, de Protección de Datos de Carácter Personal.

- 13- **Libertad:** La aplicación de los sistemas de inteligencia artificial no debería lesionar el bien jurídico de la libertad personal.
- 14- **Beneficio compartido:** Todas las tecnologías deberían beneficiar a todas las personas posibles. Este principio resulta un tanto utópico en un mundo dominado por el capitalismo salvaje, ya que históricamente ha sido posible hacer llegar las nuevas tecnologías a áreas subdesarrolladas y nunca se ha llevado a cabo. Lamentablemente, los estados que no tienen una economía poderosa tienen más dificultades para suministrar a su población los últimos avances tecnológicos de manera pública. Aun así, resulta plausible recoger un principio como este en una convención sobre ética de la tecnología.
- 15- **Prosperidad compartida:** La prosperidad económica generada por los sistemas de inteligencia artificial debería ser compartida ampliamente, para beneficiar a toda la humanidad. Este principio se encuentra en la línea de lo anterior, siendo del todo deseable, pero con la probabilidad de que acabe resultando una mera declaración de intenciones.
- 16- **Control humano:** Los humanos deberían poder elegir si delegar decisiones y cuándo delegarlas a sistemas de inteligencia artificial, y siempre con la finalidad de cumplir con objetivos humanos. Otro principio clave, sobre todo en lo que respecta a la aplicación de sistemas expertos en los distintos ámbitos de la sociedad. Esta titularidad de las decisiones debería ser siempre inajenable y sólo tener acceso a la toma de decisiones en cualquier campo un sistema artificial que haya sido habilitado para ello por una autoridad humana competente.
- 17- **No subversión:** El poder concedido controladamente a un sistema de inteligencia artificial debe respetar y contribuir a la mejora, y no subvertir los procesos cívicos y sociales de los cuales depende la prosperidad de la sociedad. En palabras más diáfanas, estos sistemas no deben alterar a la comunidad de forma agresiva, cambiando la forma de vivir de las personas, su ideología o sus creencias, sino que debe ser una adaptación progresiva, transparente y beneficiosa para las personas.
- 18- **Evitar una carrera armamentística basada en inteligencia artificial:** Los estados y los agentes institucionales de todo el mundo deberían abstenerse de empezar una carrera armamentística aprovechándose del potencial de la inteligencia artificial. Este principio consagra el carácter pacífico que han de tener los sistemas de inteligencia artificial, cuya responsabilidad recae en gran parte en los estados, que establecen las directrices políticas y financian proyectos. Se debería evitar el desarrollo de sistemas inteligentes para utilizarlos como armas o crear ejércitos, pero al igual que anteriormente, y teniendo en cuenta el agitado panorama político actual, se antoja poco menos que imposible que ciertos estados no aprovechen esta tecnología para fines ilícitos.

Problemas a largo plazo

- 19- Precaución con la capacidad de los sistemas inteligentes:** si no existe consenso, se deberían evitar algunas suposiciones optimistas acerca de esta tecnología. Este principio trata de no subestimar a la inteligencia artificial y dar por hecho que los sistemas funcionarán positivamente para la sociedad sin demasiada actuación de los humanos, ya que a largo plazo es posible que algunos sistemas adquieran habilidades para dañar a las personas y sus derechos, aunque no fuera al bien jurídico de la vida o la integridad, y ni siquiera voluntariamente – si es que un sistema de inteligencia artificial pudiera llegar a tener capacidad volitiva, cuyo espinoso debate se evitará aquí-.
- 20- Importancia:** La inteligencia artificial avanzada podría crear un cisma tecnológico profundo en la historia de la humanidad, por lo que debería controlarse y planearse con el mayor cuidado y la mayor parte de recursos posibles. Ello es así dado que nunca antes hasta estos tiempos el ser humano había dispuesto de la tecnología necesaria para crear algo que fuera capaz de pensar como él.
- 21- Riesgos:** Los riesgos que genera la inteligencia artificial, especialmente los riesgos de catástrofes o de la existencia misma de la humanidad, deben sujetarse a un plan para evitar o mitigar el impacto de dichos riesgos. Así pues, anima a las instituciones a crear, por ejemplo, planes de protección civil y ambiental acorde a posibles escenarios negativos a causa de la inteligencia artificial.
- 22- Auto-innovación recursiva:** Todos aquellos sistemas con capacidad suficiente para mejorarse o rediseñarse a sí mismos recursivamente, de manera que puedan incrementar su potencial o su cantidad- en efecto, se está refiriendo a sistemas capaces de realizar una clonación de sí mismos- deben estar sujetos a las normas y procedimientos de seguridad más estrictos.
- 23- Bien común:** La súper inteligencia debe desarrollarse únicamente al servicio del bienestar del mayor número posible de personas y para el beneficio de toda la humanidad, y no para el bien individual de un estado u organización.

Un principio que podría ser añadido a esta lista sería el de la armonización de las normas jurídicas concernientes a la inteligencia artificial. Esta directriz armonizadora permitiría un control y una transparencia acerca de lo que hacen los estados respecto a la inteligencia artificial. Sería conveniente, así, normas que armonizasen, por ejemplo, la financiación de proyectos sobre sistemas inteligentes, las normas de seguridad y control, o la responsabilidad por daños de estos sistemas. En ámbitos tan complejos de regular como la inteligencia artificial se antoja clave el hecho de disponer de normas marco unificadas, pues en caso contrario se pondría en peligro la seguridad jurídica de fabricantes, diseñadores y usuarios, que habrían de atender a normativas diferentes en función del estado en que se encuentren, cosa que en algo tan susceptible de ser globalizado masivamente no sería positivo. Por ello, en los próximos años sería interesante la celebración de conferencias que aglutinen al mayor número de estados

posibles y cuyo producto sean normas básicas en materia de inteligencia artificial y robótica suscritas por todos los estados.⁵⁰

4.7 Análisis del algoritmo COMPAS: Al límite de la ética

Se acaban de exponer una serie de problemas éticos teóricos o hipotéticos derivados de la implantación de la inteligencia artificial que podrían darse en un futuro si no llevamos a cabo las precauciones oportunas. Posteriormente se han expuesto una serie de principios que han de actuar como pauta para que dichos problemas, en la medida de lo posible, no nos afecten. Ahora se expondrá de manera ejemplificativa, a través de un caso real, la importancia práctica que pueden tener dichos problemas.⁵¹

La controversia que se plantea aquí sucedió en Wisconsin (EEUU). El objeto de la polémica reside en que en dicho lugar –así como en otros lugares de EEUU- se está ya utilizando un algoritmo que, teniendo en cuenta múltiples variables, determina, entre otras cosas, las posibilidades que tiene un reo de reincidir en el futuro y aconseja al juez a la hora de tomar la decisión acerca de la duración de la pena de prisión de un reo. En el caso del reo Eric L. Loomis, condenado a 6 años de prisión por cometer un delito de evasión policial en una persecución. El juez explicó que llegó a esa sentencia con la ayuda del mencionado algoritmo, llamado *Compas*, que está actualmente implantado en el sistema judicial de Wisconsin. Los abogados del señor Loomis cuestionaron seriamente la fiabilidad y los criterios utilizados por el algoritmo para determinar la probabilidad de reincidencia. No obstante, éstos se toparon con que los códigos para poder acceder al algoritmo y verificarlo son de carácter privado, ya que es la compañía Northpointe, Inc. la que tiene los derechos del algoritmo. El sistema funciona de manera que a través de los datos que se tienen del acusado, su historial, encuestas que se le realizan al mismo y otros parámetros, es capaz de puntuar del 1 al 10 diversas categorías acerca de los acusados. Entre esos otros parámetros el algoritmo utiliza variables como el sexo o la edad, o por su entorno económico-social –si viven en zonas pobres o conflictivas-. Así, el diario de investigación *ProPublica* realizó un análisis de los casos en los que había intervenido el algoritmo para determinar su fiabilidad y su objetividad –entendiendo objetividad como la no introducción de factores que ponderen de manera diferente a personas de diferente raza, por ejemplo-.⁵²

Centrándonos ya en dicho análisis, los periodistas de investigación se percataron de que COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*) atribuía a las personas de raza negra mayor probabilidad que a las personas blancas de

⁵⁰ Estas conferencias ya se han producido en otros ámbitos de primer orden –más allá de su aplicación y cumplimiento en la práctica- como por ejemplo en materia de cambio climático, mediante las Conferencias de Naciones Unidas sobre el Medio Ambiente y Desarrollo (Cumbres de la Tierra).

⁵¹ Smith, M. (22 de junio de 2016) "In Wisconsin, a Backlash Against Using Data to Foretell Defendants' Futures". *New York Times*. Recuperado de: <https://www.nytimes.com/2016/06/23/us/backlash-in-wisconsin-against-using-data-to-foretell-defendants-futures.html>

⁵² Larson, J., Mattu, S., Kirchner, L. y Angwin, J. (23 de mayo de 2016). "How we analyzed the COMPAS recidivism algorithm" *ProPublica*. Recuperado de: <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm> Para un análisis exhaustivo de todas las estadísticas se puede comprobar el estudio analítico.

ser juzgados incorrectamente como potenciales reincidentes de alto riesgo, mientras que las personas blancas tenían más probabilidades que las negras de ser juzgados incorrectamente como potenciales reincidentes de bajo riesgo. El estudio realizaba comparativas entre las estadísticas de reincidencia de los reos actuales con las predicciones del algoritmo de reincidencia, tras lo que se verificó que el sistema solo acertó un 60% de las veces, y tan sólo un 20% en las predicciones de reincidencia con violencia. De este completo y exhaustivo análisis estadístico se extrajeron conclusiones tan peligrosas como las dos siguientes:

- a) Los acusados negros fueron catalogados por el algoritmo con mayor riesgo de reincidencia del que realmente tenían.
- b) Los acusados blancos fueron catalogados por el algoritmo con menor riesgo de reincidencia del que realmente tenían.

De este análisis en un caso real se pueden extraer algunas conclusiones importantes. En primer lugar, se pone de manifiesto la excepcional relevancia de trasladar cuanto antes principios éticos como los propuestos en Asilomar a los ordenamientos jurídicos mundiales. En este caso, se observa que el principio de valores humanos y el principio de transparencia en los errores deberían ser convertidos en principios de configuración legal específicamente en la inteligencia artificial. Así, cualquier algoritmo que contraviniera los valores más importantes para la humanidad debería ser modificado o suprimido, y en relación con ello el principio de transparencia también debería de estar positivado en la ley, pues sólo así se haría posible un control humano en última instancia, algo que es básico, y más teniendo en cuenta que, de momento, los algoritmos no son infalibles. Esta afectación a los principios mencionados puede tener unas consecuencias jurídicas y sociales tales como las que se acaban de observar, y que pueden llevar a inocentes a la cárcel, o a culpables a condenas más altas de las que realmente merecen en términos legales.

Por otra parte, surge de nuevo la contraposición entre interés público e interés de mercado, que tanto daño hace a los derechos de las personas. En efecto, en este caso la compañía privada, creadora y poseedora de los derechos del algoritmo, tenía la exclusividad y secreto de su composición interna, por lo que no fue posible auditarlo y poder presentar una apelación digna para defender al acusado. Se hace obvio que las políticas públicas están priorizando los intereses capitalistas de la empresa propietaria del algoritmo antes que los intereses públicos. Es decir, se está favoreciendo la economía de las empresas tecnológicas aun sabiendo que los sistemas que crean pueden ser poco fiables y discriminatorios, afectando a los derechos de los ciudadanos. El mundo no se puede permitir panoramas tan oscuros y moralmente reprochables, ya que de lo contrario se podría acabar por desvirtuar y desproveer de su esencia un buen puñado de derechos humanos.

5. Nuevos problemas jurídicos a partir de la inteligencia artificial y la robótica

Desde una perspectiva general, y no sólo desde su aplicación al derecho, el potencial de expansión de la inteligencia artificial es prácticamente ilimitado. Su desarrollo conllevará, a medio y largo plazo, que los humanos abandonen muchos de los puestos de trabajo convencionales y hayan de crearse ocupaciones nuevas para el sostenimiento de la economía. La viabilidad de los empleos o del sistema de seguridad social, por ejemplo, son interrogantes todavía, siendo clave por tanto en las próximas décadas las tareas de economistas y juristas. Sin embargo, es incluso más importante disponer cuanto antes de un marco legal adecuado para canalizar el progreso de una manera sana, productiva y que respete siempre la dignidad humana, de acuerdo con lo que se ha expuesto en el apartado sobre ética.⁵³ En ese sentido, se tratará aquí de exponer lo que *de lege data* hoy día se encuentra regulado en materia de inteligencia artificial y robótica, además de las propuestas de regulación que hay sobre la mesa, incluyendo ciertas aportaciones propias al asunto.

5.1 Regulación actual sobre robótica e inteligencia artificial

Hay que partir de la base de que, actualmente, no existe en España ni en Europa legislación alguna acerca de la robótica o la inteligencia artificial. Para encontrar regulaciones legales sobre robótica e inteligencia artificial debemos irnos hasta Japón y Corea; Taiwan está investigando para establecer las bases de una futura normativa. Aparte de estas, las únicas normas relativas a la robótica que hay en la actualidad son las normativas ISO (International Organization for Standardization). Las normativas ISO más destacables son la “*ISO 10218-1:2011, Robots and robotic devices — Safety requirements for industrial robots*”, la “*ISO 13855, Safety of machinery — Positioning of safeguards with respect to the approach speeds of parts of the human body*” y la “*ISO/TS 15066:2016(en) Robots and robotic devices — Collaborative robots*”.

Las normativas ISO solamente regulan cuestiones industriales, y forman parte del denominado “*soft law*”. El “*soft law*”, término muy asociado al derecho internacional, suele englobar códigos de conducta, recomendaciones, principios, etc. A pesar de no ser en sentido estricto normas jurídicas, ello no quiere decir que no sean vinculantes y no puedan producir efectos jurídicos. De hecho, la jurisprudencia europea se ha pronunciado en varias ocasiones –por ejemplo, en la famosa “sentencia Grimaldi”– estableciendo que el “*soft law*” debe producir efectos jurídicos porque el juez puede, y en ocasiones debe, tener estas normas en cuenta para resolver un litigio. Las normativas ISO, en concreto, son certificaciones comerciales privadas que otorgan parámetros de calidad. Así pues, estas normativas atienden a aspectos como la calidad de fabricación, la seguridad industrial y salud en el trabajo, la previsión y prevención de incidentes por fallos mecánicos, los niveles de presión de la maquinaria respecto al operario en el caso que la manipule, con tal de no dañar al operario en caso de fallo mecánico o al usuario final.

⁵³ Supra p.26

Siguiendo con el tema de la regulación jurídica, es Corea del Sur la que dispone de la primera legislación en el mundo sobre robótica, donde se pone de manifiesto por primera vez la relación *maquina-hombre* únicamente desde la perspectiva de una regulación jurídica y vinculante a todos los sectores involucrados en el diseño, fabricación, utilización y desmantelamiento de “robots”. Se trata de la “*Korean Law on the Development and Distribution of Intelligent Robots (2005)*” y de la “*Legal Regulation of Autonomous Systems in South Korea (2012)*”. En Japón podemos encontrar las “*Guidelines to Secure the Safe Performance of Next Generation Robots*” y la “*New Robot Strategy - Japan’s Robot Strategy - Vision, Strategy, Action Plan*” y “*Headquarters for Japan’s Economic Revitalization*” de Febrero (2015). Por su parte, Taiwan está una investigación en este ámbito mediante un desarrollo legislativo múltiple.⁵⁴

Por lo que respecta al ámbito europeo, como se ha dicho, no hay normas sustantivas que regulen la robótica ni la inteligencia artificial. No obstante, existe el “*Proyecto Robolaw*”, denominación corta para el “*Regulating Emerging Robotic Technologies in Europe: Robotics facing Law and Ethics*”. Sin ánimo de realizar aquí un análisis exhaustivo de dicho proyecto, se expondrán las ideas del mismo de forma somera. Los autores de “*Robolaw*” consideran que no se puede dar una única definición de robot, pues jurídicamente no puede darse una definición homogénea sin que se presenten complicaciones en ciertos casos debido a la variedad de robots. Es por ello que en dicho proyecto se defiende un estudio casuístico de cada producto, es decir de cada sistema robótico o de inteligencia artificial. Por ejemplo, cuando la responsabilidad de un accidente de tráfico ya no sea de los conductores porque los coches autónomos sean ya una realidad, ¿quién será el responsable? ¿El fabricante o el desarrollador del 'software'? ¿Y si los humanos modifican el software? Jurídicamente, suele ser responsable la compañía que desarrolla el producto, pero ¿qué pasa cuando una compañía ha fabricado el robot pero otra ha programado las aplicaciones? Los autores de este proyecto defienden el establecimiento de unos principios básicos que regulen cada tipo de aplicación de la inteligencia artificial y robótica.

Siguiendo en el ámbito europeo, recientemente se emitió el Proyecto de Informe con recomendaciones destinadas a la Comisión, sobre normas de derecho civil sobre robótica, del Parlamento Europeo (2015/2103(INL)). Este informe, del cual se hará un análisis algo más en profundidad, establece las líneas básicas sobre las cuales la Unión Europea tiene la intención de legislar en un futuro.

El informe parte de una serie de Considerandos que ponen de manifiesto el imparable avance de la tecnología, que ha convertido en realidades lo que décadas atrás sólo era concebido en libros de ciencia-ficción, y los diferentes desafíos a los que se enfrenta la humanidad, como sería el caso de que eventualmente la inteligencia artificial superase la propia capacidad humana y esta última tuviera dificultades para controlar su propia creación. Por ello, es necesaria una regulación que permita a todos los sectores de la sociedad aprovecharse de los beneficios de la inteligencia artificial y de la robótica de

⁵⁴ Nisa Ávila, J.A. (30 de marzo de 2016) “Robotica e Inteligencia Artificial ¿Legislación social o nuevo ordenamiento jurídico?”. *El derecho*. Recuperado de: http://tecnologia.elderecho.com/tecnologia/internet_y_tecnologia/Robotica-Inteligencia-Artificial-legislacion-social-nuevo-ordenamiento_11_935305005.html

forma coherente, incidiendo sobre todo en los aspectos de responsabilidad por daños, de la seguridad y de los procesos de fabricación de dichos productos. Estas recomendaciones se basan en que hasta el día en que artilugios con inteligencia artificial sean capaces de tener conciencia de su propia existencia, si es que ese día llega alguna vez, ha de entenderse que las famosas leyes de Asimov van dirigidas a los fabricantes, diseñadores, y operadores. Por ese motivo, es necesario disponer de ciertas normas de responsabilidad y deontología que reflejen valores humanistas. Por ello, se va a analizar, según lo que propugna el texto del Parlamento Europeo, qué aspectos concretos han de ser regulados y porqué:

A) Responsabilidad y personalidad jurídica:

La autonomía de un robot o de un sistema de inteligencia artificial es su capacidad para tomar decisiones y aplicarlas en su entorno. A mayor grado de autonomía, menos se les puede considerar a estos sistemas meros objetos, como considera nuestro ordenamiento a efectos de responsabilidad, por ejemplo, a un animal doméstico. Por ello, si considerásemos que un robot puede tener una gran autonomía, sería conveniente plantearnos si, al menos parcialmente, puede ser él mismo y no los fabricantes, diseñadores u operadores, el responsable jurídicamente. Por ello, y aunque pueda parecer demasiado atrevido todavía a estas alturas, es conveniente también preguntarse si en cierto punto algunos sistemas de inteligencia artificial deberían de llegar a tener personalidad jurídica. No se trata, sin embargo, y por razones obvias, de equipararlos como persona jurídica a una persona física o a una sociedad, por ejemplo, sino de la creación de una nueva categoría jurídica diferente que, ante ciertos aspectos, el ordenamiento le confiriera capacidad de producir y de ser centro de imputación de efectos jurídicos. Nuestro ordenamiento actual, tanto el español como el europeo, según la Directiva 85/374/CEE del Consejo, de 25 de julio de 1985, prevé para fallos de robots que provoquen daños el régimen de la responsabilidad objetiva del fabricante—con sus característicos requisitos de daño real, causa real y relación de causalidad entre la causa y el daño—, pero este régimen puede resultar insuficiente en el futuro, pues si se va dotando a ciertos sistemas inteligentes de capacidad de aprendizaje y actuación en función de lo aprendido, esto les otorga un valor de imprevisibilidad en su actuación tal que haría imposible cualquier previsión para cubrir estos riesgos por parte de fabricantes, diseñadores o programadores, ya que dichos sistemas serían capaces de actuar conforme a su experiencia, con posibilidades infinitas, igual que los humanos, y ello les haría totalmente imprevisibles, cosa por la cual si estableciéramos igualmente un régimen de responsabilidad objetiva, esto podría desincentivar a los fabricantes a crear sistemas que podrían ser muy beneficiosos para la sociedad.

B) Principios generales

- a) Sería conveniente establecer una definición europea común tanto de “robot” como de “sistema dotado de inteligencia artificial” y de sus distintas subcategorías, en función del grado de sofisticación.
- b) Por otra parte, sería necesario crear un registro oficial de robots y sistemas de inteligencia artificial avanzados, con su correspondiente clasificación para poder mantener un control demográfico de estos sistemas.

- c) Además, se precisa cada vez más de la destinación de fondos para la investigación y desarrollo, tanto de los propios sistemas de inteligencia artificial, como del marco ético y normativo que debe regularlos, así como para el estudio de los riesgos físicos, psicológicos y sociológicos que podrían causar.

C) Principios éticos

- a) Es necesario preservar siempre en el desarrollo de sistemas de inteligencia artificial los bienes jurídicos de la seguridad humana, la intimidad, la integridad, la identidad, la dignidad, la autonomía y la propiedad de los datos.
- b) Creación de un código de conducta para los ingenieros en robótica, de un código deontológico destinado a la investigación y a la revisión de los protocolos de robótica, así como de licencias tipo para los diseñadores y los usuarios, -es decir, que se requiera una licencia para que los diseñadores y desarrolladores puedan llevar a cabo su labor y además se requiera otro tipo de licencia para que se permita a un usuario la utilización de cierto tipo de robots-. Los códigos de conducta y códigos deontológicos deben incluir en todo caso los principios de beneficencia, no maleficencia y autonomía, así como en los principios consagrados en la Carta de los derechos fundamentales de la Unión, como la dignidad humana y los demás derechos humanos, la igualdad, la justicia y la equidad, la no discriminación y no estigmatización, la autonomía y la responsabilidad individual, el consentimiento informado, la privacidad y la responsabilidad social, además de en los actuales códigos y prácticas éticas.⁵⁵

D) Creación de una agencia europea:

El Parlamento aboga por la creación de una agencia europea para la robótica y la inteligencia artificial, dotada de un presupuesto adecuado y que cuente con expertos jurídicos, expertos en inteligencia artificial y robótica y expertos en ética y deontología. Esta agencia debería controlar las aplicaciones que se le dan a la robótica y a la inteligencia artificial, así como recomendar medidas normativas y defender a los consumidores en este ámbito ante posibles contratiempos.

E) Derechos de propiedad intelectual y flujo de datos

La masividad que está adquiriendo el flujo de datos procedente de sistemas de inteligencia artificial y robótica provoca que la legislación actual no pueda resolver eficientemente los derechos de propiedad intelectual que se pueden dar. Por ello, se pretende que las instituciones europeas fomenten la creación y la innovación mediante criterios normativos de propiedad intelectual que también abarquen a las obras creadas por inteligencia artificial o robots, es decir, a todos nos interesa el progreso, por eso, hemos de proteger la innovación mediante un sistema de derechos de autor que también reconozca estos derechos a robots que puedan llegar a ser capaces de crear nuevas ideas, patentes e incluso obras artísticas o literarias.

⁵⁵ Proyecto de Informe con recomendaciones destinadas a la Comisión, sobre normas de derecho civil sobre robótica, del Parlamento Europeo (2015/2103(INL))

F) Vehículos autónomos

El sector automovilístico es el que necesita con más urgencia normas que regulen las aplicaciones de la inteligencia artificial, por ello urge una nueva regulación que dé cobertura a la fabricación, diseño y logística de vehículos autónomos, así como una normativa mejorada que incluya a estos vehículos en materia de responsabilidad por accidentes de tráfico. Me referiré con más profundidad a ellos posteriormente.⁵⁶

G) Médicos, rehabilitación y drones

Respecto a los robots asistenciales, el Parlamento europeo señala que el contacto humano es una característica esencial de la prestación de cuidados, de manera que la implementación de robots en este ámbito podría deshumanizar la prestación. Por lo que respecta a los robots médicos, se precisa que se establezcan legalmente o reglamentariamente los requisitos que han de cumplir los médicos para utilizarlos como asistencia en sus tareas, tales como una formación específica en la tecnología robótica-médica correspondiente y un mínimo de experiencia práctica. En lo que concierne a la rehabilitación de órganos dañados y el restablecimiento de funciones corporales reducidas mediante la robótica, se destacan las virtudes en este campo, pero se reclama la creación de comités de ética sobre robótica en hospitales. Respecto a los drones, el Parlamento europeo se pronunció con anterioridad a estas recomendaciones en su Resolución de 29 de octubre de 2015, sobre el uso de los sistemas de aeronaves pilotadas de forma remota (RPAS).

En el mismo terreno referente a los drones, los que no estén sujetos a la normativa europea se regirán por el *Proyecto Real Decreto por el que se regula la utilización civil de las aeronaves pilotadas por control remoto, y se modifica el Real Decreto 552/2014, de 27 de junio, por el que se desarrolla el Reglamento del aire y disposiciones operativas comunes para los Servicios y procedimientos de navegación aérea y se modifica el Real Decreto 57/2002, de 18 de enero, por el que se aprueba el Reglamento de Circulación Aérea*, que establecerá el marco jurídico aplicable a la utilización de drones en España. El régimen especial que disfrutará la policía en este Real Decreto puede ser un tanto polémico. Los drones pueden ir equipados con cámaras de alta definición y sensores capaces de detectar anomalías en grandes aglomeraciones de personas, ejecutar tareas de reconocimiento facial, rastrear redes de wifi y hackearlas. Además, existen muchos más usos como por ejemplo operaciones especiales en antiterrorismo, narcotráfico, tráfico, etc.

En resumen, el Proyecto de Real Decreto, que se aprobará probablemente este 2017, prevé un régimen específico para la policía que le otorga mucha libertad operativa. Aun así, en este caso el legislador se ha olvidado de regular el uso de drones por parte de otros operadores públicos que pueden requerir su uso, como los cuerpos de bomberos o los servicios sanitarios.

⁵⁶ Infra p.42

H) “Kill Switch” y otros principios básicos para diseñadores y usuarios

En el anexo del documento se incluye una recomendación a los diseñadores y fabricantes para que incluyan en los robots mecanismos que permitan interrumpir totalmente la conexión del robot en caso de peligro para la seguridad de los humanos – de ahí “kill switch” o botón de la muerte. Además, otra recomendación importante destinada a los fabricantes es que éstos, llegado el momento en que la tecnología sea tan avanzada que pueda confundir a simple vista a un robot con una persona, deban establecer algún mecanismo que permita diferenciar fácilmente a un robot de un humano. Es decir, los diseñadores y fabricantes deberán incluir “algo” que caracterice por completo a los robots, por obvias razones éticas.

I) Formación y empleo

Se pretende revisar por parte de las instituciones europeas en qué ámbitos se está destruyendo empleo y en cuáles se está creando a causa de la utilización de los robots, para a partir de ahí reestructurar el mercado laboral de una manera u otra. También propone exigir a las empresas que comuniquen qué porcentaje de sus ingresos corresponden a la inteligencia artificial y robótica. Además, en este ámbito se propone algo que ha dado mucho de qué hablar y sobre lo que se ha especulado durante los últimos años frente al avance imparable de la tecnología que destruye puestos de trabajo: la creación de una renta básica universal. En efecto, esta renta básica universal sería no una opción sino una necesidad en un escenario futuro muy posible en el que la mayoría de puestos de trabajo actuales queden substituidos por robots, en especial en el sector primario, que es el más susceptible de ser mecanizado. No obstante, hay que evaluar si la más que probable destrucción de puestos de trabajo no puede compensarse con la creación de nuevos puestos en otros ámbitos.

Finalmente, también han existido propuestas para que sean susceptibles de aplicárseles tributos a todas las tareas que llevan a cabo robots, y equiparar esa tributación a la que se aplica a las tareas llevadas a cabo por humanos. En efecto, personajes clave en el sector tecnológico, como Bill Gates,⁵⁷ son partidarios de establecer tributos para todos los trabajos llevados a cabo por robots y conseguir con esa recaudación que las personas que pierden sus puestos de trabajo puedan dedicarse a las tareas más sensibles de la sociedad –y que por ende consideramos el último reducto exclusivo de humanidad- como cuidar de las personas mayores o de niños –aunque incluso ya existan robots pensados para ello. Profundizaré en ello más adelante.⁵⁸

J) Transparencia de los algoritmos de la inteligencia artificial

En aras de que el impacto de la inteligencia artificial no distorsione la transparencia de los procesos sociales, los diseñadores de los robots deberán lograr que toda decisión que tome un sistema dotado de inteligencia artificial sea explicable y reconducible a su origen –los pasos lógicos que ha seguido el robot para llegar a esa decisión- y que todo

⁵⁷ (20 de febrero de 2017) “Bill Gates opina que los robots deben pagar impuestos.” *El Mundo*. Recuperado de: <http://www.elmundo.es/tecnologia/2017/02/20/58aab904ca4741657a8b45dd.html>

⁵⁸ Infra p.45

algoritmo y cálculos que el sistema utilice se puedan reducir a un lenguaje comprensible para los humanos. Esto se corresponde con el principio de transparencia en los errores formulado en Asilomar, y que anteriormente se ha mencionado.⁵⁹

Como se ha expuesto, no existe todavía una regulación concreta más allá de meras propuestas. Más allá de las acertadas Recomendaciones del Parlamento europeo, surgen algunas cuestiones conflictivas por la incertidumbre que proyectan en el futuro. Trataré más adelante de exponer algunas de ellas y de esbozar algunas propuestas que podrían ser de utilidad en dichas cuestiones.

5.2 Derecho internacional y armas autónomas

Uno de los muchos debates que surgen a raíz de los problemas legales del auge tecnológico presente y futuro versa sobre el uso armamentístico que se le pueda dar a la robótica y a la inteligencia artificial. Los estados, en el panorama político internacional siempre han tratado a lo largo de la historia de incentivar a sus científicos para desarrollar y construir armamento. Lo han hecho siempre con el pretexto, no exento de justificación, de que sólo mediante un poder militar y de defensa sólido y superior al del resto se puede desincentivar a otros estados u organizaciones terroristas a utilizar la fuerza para hacer el mal. En otras palabras, los estados, en mayor o menor medida, tratan de seguir siempre la máxima de *“si quieres la paz, prepárate para la guerra”*. En los últimos años hemos oído hablar de armas nucleares, bombas de hidrógeno y otros conceptos armamentísticos que, pese a no comprender con exactitud, hacen obvio su potencial destructivo. Los estados tratan de superarse a sí mismos para no permitir a los estados “enemigos” tener armas más poderosas.

Hoy en día, muchos estados están investigando en tecnología armamentística autónoma, que podría llegar a por ejemplo, seleccionar un objetivo en función de su comportamiento, y si éste es sospechoso, eliminarlo. Tales armas no existen aún, aunque se estima que pueden llegar a existir. Los peligros a nivel ético y legal son múltiples, y de hecho ya se han realizado varias convenciones de organismos internacionales, como la Asamblea General de la ONU, para discutir estos asuntos.

Entrando ya en materia, una “arma autónoma” por definición sería cualquier artillugio, sistema o mecanismo dotado de autosuficiencia y autogobierno, independiente o muy poco dependiente de un control externo, y capaz de causar daño y ser letal. Esas características están muy ligadas a la dotación a estas armas de inteligencia artificial, lo cual les permitiría hacer cosas como seleccionar y eliminar objetivos basándose en criterios preestablecidos por personas humanas al diseñarlos. Está claro que éste va a ser uno de los terrenos más difíciles para llegar a consensos internacionales y poder regular eficazmente, pero sí que desde ya podemos empezar a trazar algunas directrices básicas para que esos sistemas sirvan para protegernos y no para destruirnos. Por tanto, la clave residirá en establecer primeramente el marco ético y posteriormente confeccionar sobre ese marco el ordenamiento internacional público adecuado. No en vano, el poder de esta tecnología podría conseguir eliminar cualquier conflicto armado, pero también crear nuevos conflictos con armas aún más peligrosas y destructivas.

⁵⁹ Supra p.28

Será difícil, de momento, llegar a consensos en ésta materia, ya que es difícil ponerse de acuerdo en algo que aún no existe, y de lo que por tanto no podemos conocer su potencial. Además, el ordenamiento internacional suele incluir conceptos de textura abierta como “norma del jefe militar razonable” “sentido común” que complican más las cosas. En este campo, un error de interpretación de la máquina, puede suponer no ya la libertad o el patrimonio de una persona, sino la seguridad y la vida de civilizaciones enteras.

5.3 Derecho de daños y robótica

Se ha apuntado anteriormente la voluntad de la Unión Europea de crear un registro de robots y de poder atribuir a estos personalidad jurídica. No en vano, ante la inminente masividad de la robotización de la sociedad, nos hemos de preguntar: ¿Qué pasa si un robot daña un bien jurídico humano? Aquí cabe especificar que se está hablando en este apartado de responsabilidad civil, esto es, quién paga el coste económico de los daños que cause un robot. También es necesario matizar que en este apartado se trata el término “robot” en sentido estricto, es decir, aquellos sistemas robóticos con inteligencia artificial, y por tanto capacidad de aprendizaje y de llevar a cabo conductas muy similares a las humanas. Se excluyen por tanto los daños que puedan causar robots sencillos y monofuncionales –como una aspiradora-, ya que estos tienen ya una regulación claramente aplicable: en el ámbito europeo, la *Directiva 85/374/CEE del Consejo, de 25 de julio de 1985, relativa a la aproximación de las disposiciones legales, reglamentarias y administrativas de los Estados miembros en materia de responsabilidad por los daños causados por productos defectuosos*, mientras que en ámbito español tenemos la LGDCU (*Ley General para la Defensa de los Consumidores y Usuarios*) juntamente con los artículos 1902 y siguientes, y 1101 y siguientes de nuestro Código Civil.

Por lo que respecta, ahora sí, a los robots más desarrollados e inteligentes, el marco legal aplicable sería también el recién mencionado. No obstante, se hace patente la necesidad de reformarlo ante los diversos dilemas que la robótica avanzada puede causar en el derecho de daños. En efecto, a estos robots avanzados se les seguiría aplicando, en aras de determinar la existencia de responsabilidad civil, los requisitos de la existencia de defecto concreto en el producto –es decir, en los sistemas robóticos o con inteligencia artificial- daño efectivo, concreto e individualizado, y relación de causalidad entre el defecto del producto y el daño. Dicho esto, comienzan a surgir muchas trabas jurídicas a la hora de aplicar estos requisitos a la robótica sofisticada.

En primer lugar, se prevé que muchos robots incluyan opciones de personalización, esto es, que se le permita al usuario modificar las características de serie del robot, o incluso que determinados usuarios con los conocimientos adecuados puedan modificar el robot para programarlo con funciones para las cuales no estaban ideados. En esas circunstancias, se hace muy difícil determinar quién fue el causante del daño. Una posible solución a esto sería elaborar unas pautas de personalización y/o usos legítimos que se le pueden dar al robot, de manera que el usuario que lo utilice sepa que si

traspasa la línea que marcan esas pautas, quedará fuera de la cobertura legal de responsabilidad civil, sin perjuicio de la responsabilidad civil en la que pueda incurrir el mismo usuario por los daños que pueda causar el robot como consecuencia de sus modificaciones inadecuadas.

Otro de los problemas, probablemente el de mayor calibre, es el de la capacidad de aprender de la que gozan los sistemas de inteligencia artificial. Una capacidad que, como ya se ha apuntado, irá *in crescendo* hasta alcanzar niveles realmente asombrosos. Ante esa capacidad de aprender mediante la interacción con el entorno, un robot podría llegar a adoptar comportamientos inicialmente no previstos por sus programadores. Ante las situaciones que se pudieran dar como consecuencia de conductas inesperadas, ¿qué posición puede y/o debe adoptar la ley? Muchos consideran adecuado restringir la capacidad de aprender de los robots hasta un límite, es decir, establecer una línea roja para el potencial de los robots, de manera que únicamente puedan llegar a aprender tareas ya contempladas en su programación. Esta postura se conoce como “*code as law*”.⁶⁰ Desde mi punto de vista, ésta última parece una solución momentánea y preventiva, de manera que mientras la tecnología disponible en este campo se encuentre en una fase tan prematura lo más sensato sea ser realmente estrictos con las funcionalidades de los robots. No obstante, a medio y largo plazo seguir restringiendo el progreso no sería adecuado.

Por otra parte, también hay que plantearse los casos en que haya de establecerse un sistema de responsabilidad objetiva (art. 1903 CC). En efecto, parece sensato que, en ciertas situaciones exista un garante humano que deba responder por los daños del robot. Esta última afirmación puede llevar a confusiones, ya que en la responsabilidad en la que media dolo o culpa (art. 1902 CC), el daño material lo causa el robot pero se considera responsable civilmente del daño a otra persona –diseñador, fabricante, usuario, tercero,- no porque ésta tenga una obligación legal o contractual sino porque ante la imposibilidad – de momento- de hacer reparar a los propios robots sus daños, tiene que repararlos un humano. En cambio, lo que la responsabilidad objetiva regula son aquellos casos en los que existe una especial vinculación o relación entre el robot en este caso y la persona objetivamente responsable, ya sea esta relación de tipo contractual – por ejemplo, un empresario habría de responder por los daños que sus robots trabajadores causen en el ejercicio de sus tareas- o legal –por ejemplo, el dueño de un robot habrá de responder por los daños que éste cause a terceros-. En este régimen de responsabilidad objetiva, también sería lógico que se pudiera ejercer una acción de regreso por parte de los garantes, en este caso contra fabricantes, diseñadores o terceros, si el daño del cual los garantes son responsables objetivamente viene causado por dolo o negligencia de aquéllos.

En todo caso, de momento no existe una solución satisfactoria para regular eficazmente la responsabilidad civil en el ámbito de la robótica, y por lo tanto será la jurisprudencia la que adopte un papel relevante y empiece a establecer los criterios de determinación de la responsabilidad. Lo que parece claro es que el derecho de daños en este campo

⁶⁰ (4 de abril de 2016) “¿Qué ocurre cuando un robot causa daños?” Replicante Legal. Recuperado de: <http://replicantelegal.com/que-ocurre-cuando-un-robot-causa-danos/>

será tremendamente casuístico, y que cualquier matiz fáctico, por minúsculo que sea, puede influir en el fallo en uno u otro sentido.

5.3.1 Interrogantes jurídicos de los vehículos autónomos

Mención aparte merecen los vehículos autónomos. Para empezar a hablar de ellos y de los interrogantes jurídicos que implica su implementación, primeramente hemos de hablar de las previsiones de los expertos acerca de la misma. Las estimaciones más optimistas creen que a partir del año 2025 existirán ya vehículos capaces de realizar sin dependencia ni supervisión humana alguna el proceso de conducción, en cualquier trayecto. Si bien estas estimaciones son técnicamente posibles a pequeña escala, lo cierto es que una implantación a nivel global probablemente suponga un plazo de tiempo bastante más largo, debido a la necesidad de adaptar las infraestructuras existentes –carreteras, semáforos, señales de tráfico- a la circulación de estos coches. Lo más probable es que, tal y como ha manifestado ya nuestra Dirección General de Tráfico, la implantación sea progresiva, dividiéndola en niveles del 0 al 5, siendo 0 una conducción no automatizada en absoluto y 5 una conducción en la que el vehículo realice la totalidad de la conducción por sí solo. A medida que avance la tecnología de manera totalmente segura, se irá avanzando de nivel.⁶¹

Cuando llegue el momento en que el nivel de autonomía de los vehículos vaya aumentando, emergerán cuestiones ético-jurídicas de vital importancia. Por ejemplo, ¿qué prioridades deben tener los algoritmos de los vehículos en caso de accidente inevitable? Se está hablando aquí de una extrapolación a los vehículos autónomos del famoso dilema del tranvía, ideado por la filósofa británica Philippa Foot:

El dilema consiste en imaginar que estamos en un tranvía en marcha el cual no podemos detener de ninguna manera, y el cual circula en una vía en la que a pocos metros hay 5 personas. La única acción que tenemos disponible es accionar una palanca que cambiará el tranvía de vía, en la cual hay una sola persona. ¿Qué decisión es la más acertada? ¿Accionar la palanca y matar por acción a una persona para salvar a cinco? ¿O por el contrario, no hacer nada y matar por omisión a cinco personas para salvar a una? Desde que se planteará el dilema moral en 1967 han corrido ríos de tinta tratando de ponerle claridad, y se ha tratado de adoptar una u otra decisión en función de diversas corrientes de pensamiento. El dilema tiene muchas variantes. Una de ellas es la del túnel, y plantea una situación en la que nuestro vehículo se topa, sin tiempo para frenar, con un niño que ha tropezado a la entrada de un túnel muy estrecho. ¿Deberíamos girar bruscamente y colisionar, a riesgo de perder nuestra vida para salvar la del niño, o no deberíamos hacer nada para salvarnos a nosotros mismos?⁶² Esta última variante plantea una situación que se podría llegar a dar con los vehículos autónomos.

⁶¹ Instrucción 15/V-13 de la Dirección General de Tráfico. Tal y como se cita en: Sánchez Del Campo, A. (7 de noviembre de 2016). "Aspectos legales y éticos de los coches autónomos". *Fleet People*. Recuperado de: <https://fleetpeople.es/aspectos-legales-eticos-los-coches-autonomos/>

⁶² Rubio Hancock, J. (31 de marzo de 2017) "El dilema del tranvía: ¿Debo sacrificar una vida para salvar cinco?" *Verne El país*. Recuperado de: http://verne.elpais.com/verne/2017/03/27/articulo/1490625074_938459.html

En mi opinión, no existe una solución más plausible moralmente que otra en la práctica. Cada corriente filosófica aportará su visión teórica correspondiente, pero siendo pragmáticos y habiendo de aplicar una solución a la vida real, no parece que tengamos legitimidad para juzgar al que decida salvar su vida. Así pues, el dilema real consistirá en aquellas situaciones en las que estén en juego bienes jurídicos del mismo valor –al menos, del mismo valor ante la ley- y el algoritmo que lleve incorporado el vehículo autónomo deba decidir cuál de los dos sacrifica para salvar al otro. Al tener en la práctica, como he dicho, el mismo valor jurídico la vida de un anciano que la de un niño, parece que la decisión sobre cómo programar el algoritmo corresponde al fabricante. Es probable que la mayoría de ellos opten por programar sus algoritmos de manera que siempre que exista la más mínima posibilidad de supervivencia para el conductor, aquéllos ejecuten la acción que efectivamente pudiera llegar a salvar a éste. Esto último se debe a que, en abstracto, muchas personas responden de manera altruista al dilema del túnel, afirmando que girarían y se arriesgarían a morir para salvar al niño, pero a la hora de elegir un vehículo, la mayoría de los individuos preferiría uno que intentará salvarlos a ellos mismos por todos los medios. En pocas palabras, si los fabricantes, en situaciones de elección de iguales bienes jurídicos, tienen el poder para decidir, éstos optarán por fabricar lo que el público demanda: instinto de supervivencia, esto es, vehículos autónomos cuyo algoritmo priorice siempre salvar la vida de sus ocupantes. Otra opción disponible para el legislador sería la de imponer a los fabricantes la obligación de diseñar el algoritmo para que, en casos de dilemas como el anterior, los vehículos autónomos actuaran ejecutando la acción que matemáticamente tuviera más posibilidades de supervivencia.

Siguiendo en el terreno ético-jurídico de los vehículos con inteligencia artificial, el MIT (Massachusetts Institute of Technology) ha desarrollado una plataforma online mediante la cual los usuarios de la red pueden simular y dar su opinión acerca de la decisión más acertada moralmente en ciertas situaciones. La plataforma se denomina “Moral Machine” (máquina moral) y en ella cualquier persona puede visualizar escenarios hipotéticos en los cuales se plantean elecciones disyuntivas de alta complejidad moral, y simular la elección que esa persona tomaría. Así, la plataforma permite agrupar las opiniones de millones de personas respecto del tema, compararlas entre ellas e incluso crear nuevos escenarios. Estas opiniones pretenden ser utilizadas para contribuir al desarrollo de los algoritmos de los vehículos.⁶³

Para seguir profundizando en el tema de la adaptación de los ordenamientos jurídicos a los vehículos autónomos, se recomienda hacer una lectura del brillante análisis ético-jurídico acerca de este asunto de Colom Planas.⁶⁴ En él, se efectúa un repaso de la legislación vigente en materia de circulación vial, las especificidades técnicas de los vehículos autónomos y cómo éstas plantean ciertos problemas jurídicos, así como algunas posibles soluciones. Por ejemplo, cabe la posibilidad de atribuir personalidad jurídica a los software inteligentes que conduzcan los coches, dada la necesidad de

⁶³ Acceso a la plataforma “Moral Machine” en: <http://moralmachine.mit.edu/>

⁶⁴ Colom Planas, J.L. (15 de agosto de 2015) “Análisis jurídico de los vehículos autónomos: tecnología, ética y regulación”. *Aspectos profesionales: Protección de datos, cloud computing y sistemas de gestión*. Recuperado de: <http://www.aspectosprofesionales.info/2015/08/analisis-juridico-de-los-vehiculos.html>

encaje con la actual legislación – enmienda de 2014 a la Convención de Viena sobre circulación vial de 1978- que permite estos vehículos siempre y cuando haya presencia física de un conductor –de esta manera podríamos considerar al propio software como el conductor-. La otra solución sería simplemente redefinir el término conductor, dando cabida en él también a “cosas” (software).

Finalmente, sin apartarnos del derecho civil, cabe hablar de la relevancia que cobrará progresivamente el derecho de seguros. Esta disciplina también habrá de readaptarse al impacto de las nuevas tecnologías, pero en especial al de los vehículos autónomos. Actualmente, todo el derecho regulador de la circulación de vehículos y de los seguros de los mismos se basa en el control que ejerce una persona física sobre el vehículo. Como se ha apuntado antes, la implantación de los vehículos autónomos se llevará a cabo por fases o niveles. Mientras no se llegue al último nivel, es decir, al de conducción plenamente automatizada, el marco legal de los seguros vendrá a ser similar al actual. Así, la actual Ley de 35/2015, de 22 de septiembre, que modifica el texto refundido de la Ley sobre responsabilidad civil y seguro en la circulación de vehículos a motor, aprobado por el Real Decreto Legislativo 8/2004, de 29 de octubre coloca el peso de la responsabilidad civil sobre el conductor, o sea, la persona que en última instancia y pese a que el vehículo este tecnológicamente muy dotado, tiene el control final del mismo. Esto necesariamente habrá de cambiar, como se ha dicho, cuando el conductor sea el propio software del vehículo. Entonces, en un escenario de conducción ya totalmente automatizada en el que se produce un accidente, no se podría exigir en principio ningún tipo de responsabilidad civil a las personas físicas que viajen en el vehículo, puesto que ellas no tienen el control del vehículo. Habría que exceptuar los casos en que una persona física saboteara el software del vehículo o algún elemento externo necesario para la circulación eficiente y segura –señales, sensores, balizas, etc- aunque en ese caso no hablaríamos ya de responsabilidad civil, sino seguramente de responsabilidad penal.

Dicho esto, lo más probable es que la responsabilidad civil –y por ende la necesidad de contratar seguros- se acabe desplazando de la persona titular del vehículo hacia otras personas como fabricantes, diseñadores, constructores, contratistas y subcontratistas encargados de la construcción de las infraestructuras adaptadas a la circulación y personal de mantenimiento de dichas infraestructuras.

5.4 Derecho laboral y de la seguridad social y robótica

Tal y como se ha dicho anteriormente, el impacto de la robótica en la economía se espera que sea desmesuradamente fuerte. No es algo nuevo, sin embargo, el hecho de incorporar robots al proceso productivo. En lo que respecta al plano jurídico, los tribunales españoles ya han tenido en alguna ocasión la oportunidad de pronunciarse al respecto. Así, por ejemplo, ya existe jurisprudencia que considera una causa objetiva de despido el hecho de la incorporación de robots al proceso productivo, en detrimento de

un trabajador humano.⁶⁵ No obstante, antes de intentar arrojar algo de luz a la complejidad que presenta el fenómeno robótico en el campo laboral, es preciso resumir brevemente los pros y los contras del mismo.

Por un lado, los robots pueden en el futuro llegar a realizar tareas que históricamente y todavía hoy presentan riesgos para la integridad física y mental de las personas, e incluso para su dignidad. Imaginemos cuántas muertes en lugares como minas o andamios enormes podríamos evitar. Pensemos en todas las lesiones que no se producirían si no se realizaran determinados trabajos a mano. O quizá en un robot quirúrgico al que no le tiembla el pulso ni un milímetro y que tras las operaciones, los pacientes nunca hubieran de reclamar judicialmente las negligencias de otros humanos. E incluso toda la carga de estrés y presión que podríamos sustraer a millones de trabajadores en todo el mundo. Los beneficios de la robótica aplicada al trabajo son muy amplios.

Por otro lado, los inconvenientes potenciales también resultan claros: desempleo, sistemas de seguridad social y mayor peligro de que los robots puedan causar daño. Respecto al último de estos tres problemas planteados, no parece acertado frenar el auge de la aplicación robótica a la producción laboral solamente por el posible peligro de que estos robots puedan causar daños. Al contrario, está claro que si hay algún animal imperfecto, ése, es el ser humano. Así pues, siguiendo con el mismo ejemplo anterior del robot quirúrgico, estadísticamente tendríamos un porcentaje de negligencias muy inferior en robots que en humanos, por lo que no solamente no resulta temerario, en lo que a la seguridad e integridad de las personas respecta, el hecho de permitir la robotización de muchos campos laborales, sino que además resulta más fiable y seguro. En todo caso, como ya se ha expuesto en otro apartado, debe establecerse un sistema eficaz de responsabilidad civil para estos casos.

Respecto a los dos otros problemas –el desempleo y los sistemas de Seguridad Social– sí que plantean bastante más inconvenientes. Nuestro sistema actual de Seguridad Social es poco sostenible. Pese a los esfuerzos del poder ejecutivo para ocultar los defectos de dicho sistema, la realidad es que las cuentas cada vez salen menos. La población es cada año más longeva y los índices de natalidad, sobre todo después de los efectos de la reciente crisis económica, son demasiado bajos. La ecuación conlleva que no haya suficiente población en edad activa que cotice lo suficiente como para sostener el sistema de pensiones, prestaciones y subsidios. Para suavizar el riesgo de colapso del sistema se han llevado a cabo propuestas interesantes. Una de ellas es la que plantea que las empresas coticen por sus robots, entendiendo aquí el término robot en sentido amplio, de manera que cualquier máquina que realice trabajos mecánicos de producción y cuya implantación en la empresa haya comportado la pérdida de empleo de trabajadores, esté sujeta a cotización en número equivalente a los trabajadores que se requerirían para realizar su mismo trabajo. El debate, pues, se centra en la legitimidad de obligar a las empresas a cotizar por algo que no es sujeto pasivo de servicios sociales –no pide la baja, no tiene accidentes laborales, no queda en situación de desempleo-. Esta propuesta va en consonancia con la que permitiría la creación de un registro de

⁶⁵ (STSJ Castilla y León/Burgos, de 23-7-2009, JUR 361518. Tal y como se cita en: Mercader Uguina, J.R. (12 de marzo de 2017) “La robotización y el futuro del trabajo”. *Replicante legal*. Recuperado de: http://replicantelegal.com/la-robotizacion-y-el-futuro-del-trabajo/#_ftn8

robots –al menos de los más sofisticados- y que permitiría atribuir un cierto grado de personalidad jurídica a robots, al menos en lo que a obligaciones se refiere. Uno de los grandes defensores de esta idea es el magnate de la tecnología Bill Gates, quien sostiene que la destinación de los ingresos obtenidos por las cotizaciones de las empresas por sus robots debería destinarse a la creación de empleo en otras áreas para compensar la destrucción en las áreas en que se han implantado robots. Lo cierto es que, una vez más, huelga decir que es un problema de presente y no de futuro. De hecho, cada año avanza con más fuerza el proceso de automatización en muchas áreas, hasta el punto en que la OCDE estima que en los próximos años en España desaparecerán el 12% de las profesiones actuales.⁶⁶

No obstante, hay como mínimo las mismas razones para ser optimistas que pesimistas. De hecho, no sólo se puede afirmar que la robotización no tiene por qué afectar a las tasas de desempleo, sino que se puede decir incluso que ésta es positiva para la creación de empleo. Esta afirmación se basa en datos empíricos. Si atendemos a las cifras que presentan países como Corea, Japón, Singapur o Alemania, podemos observar cómo, siendo estos los países con más densidad robótica –es decir, más número de robots por cada trabajador humano- son en cambio países con índices de paro muy bajos.⁶⁷ En los Estados Unidos de América, por ejemplo, se instalaron más de 60.000 robots en la industria del automóvil entre 2010 y 2015, y durante ese mismo período el número de empleados en el mismo sector incrementó en 230.000.⁶⁸ Estos datos nos invitan a hacer una reflexión positiva acerca del asunto. Es cierto que la mayoría de puestos de trabajo tal y como los conocemos actualmente irán desapareciendo paulatinamente. Pero no es menos cierto que surgirán otros tipos de puestos de trabajo mucho más específicos, menos duros y mejor remunerados. La mayoría de estos nuevos puestos de trabajo tendrán que ver con el ocio, la creatividad, o tareas cognitivas delicadas. Las personas trabajarán menos horas, cobrando más y en empleos más específicos. Además, debido a la sustitución en muchos sectores de personas por robots, todo ese contingente de mano de obra podrá ser trasladado a sectores que actualmente necesitan especial atención, como el cuidado y vigilancia del medio ambiente, la educación o la cultura.

No obstante, para cosechar beneficios en vez de perjuicios es necesario adoptar políticas públicas proactivas y progresistas. En primer lugar, no es conveniente mantener a toda costa los puestos de trabajo fácilmente sustituibles por robots. Ello no es positivo ni para las empresas ni para la economía en general. En efecto, a primera impresión es fácil pensar que sería perjudicial, ya que cada sustitución conllevaría una persona más en el paro – en otras palabras, menos ingresos, menos consumo y, por tanto, decrecimiento económico-. No obstante, hemos de pensar que por cada trabajador que se sustituye la empresa reduce costes de producción, pudiendo llegar a ofrecer precios más competitivos, lo cual estimula la competencia empresarial –es decir, productos de más

⁶⁶ (28 de mayo de 2016) “La automatización pone en riesgo un 12% de empleos en España” *El País*. Recuperado de: http://economia.elpais.com/economia/2016/05/20/actualidad/1463769085_077235.html

⁶⁷ Garzón, E. (27 de febrero de 2017) “Los países que más robots tienen disfrutan de menos paro”. *Saque de esquina*. Recuperado de: <http://eduardogarzon.net/los-paises-que-mas-robots-tienen-disfrutan-de-menos-paro/>

⁶⁸ “Robots create Jobs”. *International Federation of Robotics*. Recuperado de: <https://ifr.org/robots-create-jobs>

calidad y más baratos, lo que conlleva más consumo y por tanto crecimiento económico. Ahora bien, las políticas públicas deberían ir en la línea de la investigación y estudio de nuevas profesiones que se necesiten en cada sector. Por decirlo de manera ejemplificativa, si sustituimos por robots a 50 temporeros que recogen aceitunas en un campo, hemos de invertir en aprender cómo recolocarlos en el mercado, por ejemplo, colocándoles en puestos de agricultor-chef – perfil laboral recientemente creado y consistente en el desarrollo y cultivo de productos agrícolas innovadores para a su vez innovar en el ámbito culinario-.⁶⁹

Por otro lado, también se plantea la creación de nuevos tributos relacionados con el impacto de la robótica. La naturaleza técnica de estos impuestos también está abierta a múltiples posibilidades, es decir, podríamos incluir un gravamen extra por la robotización en prácticamente cualquier tributo. Una opción sería incrementar el IVA de aquellos productos o servicios para los cuales se ha utilizado un elevado porcentaje de robots o gravar más mediante este impuesto la compra de robots o de tecnología con inteligencia artificial. Otra opción, compatible con la anterior, podría ser la de gravar tanto en el IRPF (Impuesto sobre la Renta de las Personas Físicas) como en el IS (Impuesto de Sociedades) aquellos ingresos obtenidos por personas físicas o jurídicas que provengan de una actividad en la que un robot haya tenido una participación relevante. De la misma manera que en las cotizaciones de la Seguridad Social por el uso de robots, parece que lo más sensato es destinar la recaudación por los tributos que gravan el impacto de la robótica a la creación de empleo en otras áreas. Las especificidades técnicas de la implantación de estas medidas han de ser estudiadas con cautela, pero es probable que aplicando dichas medidas se pueda compensar los dos grandes problemas que plantea la robotización del sistema productivo, tanto en España como en el resto del mundo.

5.5 Derecho penal y robótica

“El que matare a otro será castigado, como reo de homicidio, con la pena de prisión de diez a quince años”. Así reza literalmente el artículo 138 del Código Penal español. Cualquier ciudadano no tendría demasiadas dudas para interpretar este texto hoy en día, pero a raíz de la eclosión de la inteligencia artificial y el desarrollo de robots, es posible que hayamos de reinterpretar las expresiones que refieren a los sujetos pasivos de los delitos. Usando el art. 138 CP como ejemplo, se observa que las expresiones *“El que matare”* y *“a otro”*, pueden llegar a generar dudas en un futuro, cuando ciertos robots adquieran condiciones que, a efectos penales, no lo harían diferente de cualquier sujeto activo humano, es decir, que podría en un futuro darse perfectamente casos en que un robot sea capaz de cumplir los elementos volitivos y cognitivos (querer y conocer) de los tipos penales. Es obvio que, si como se ha dicho anteriormente, se cumplen estrictamente una serie de principios éticos y se aplica una regulación legal en el diseño, fabricación y comercialización de robots, difícilmente se podría llegar a una

⁶⁹ Otra cuestión sería la manera de configurar la cotización. Algunos expertos apuntan en la línea de una cotización fija, mientras que otros creen que una parte de la misma ha de cuantificarse en relación a los beneficios y a la productividad que genera cada robot.

situación en la que un robot voluntariamente cometa homicidio, pues el principio de no dañar a los seres humanos debería ser el principio básico.

Ahora bien, toda tecnología es susceptible de hacer el mal, y no es difícil imaginar situaciones en que la robótica pudiera ser utilizada para fines fraudulentos y malvados, como el terrorismo. En esa tesitura, parece que lo más plausible sería no considerar a un robot, por más semejante al humano que sea, como susceptible de ser sujeto activo de un delito. Esta afirmación es extrapolable también a todo el derecho sancionador en general (sanciones administrativas). Es complejo justificar la no atribución de responsabilidad criminal a sujetos que, ante las circunstancias de un ilícito, sean equiparables en cierto modo a los humanos, pero parece que las razones son más de carácter pragmático que moral. Es decir, para la humanidad es más sencillo y mucho menos caótico continuar con un modelo de sistemas penales antropocéntrico – aunque se haya previsto recientemente en nuestro ordenamiento y en otros la responsabilidad criminal de las personas jurídicas- que acabe reduciendo toda responsabilidad penal a un individuo humano. Eso no quiere decir que no pudiera funcionar un sistema de responsabilidad penal que incluyera a los robots como sujetos activos, sino que desconocemos las consecuencias que ello podría comportar, y no parece adecuado arriesgarse a afrontarlas. En conclusión, lo más acertado desde mi punto de vista sería no considerar a los robots como sujetos activos de delitos o sanciones administrativas, sino individualizar dicha responsabilidad en la persona humana garante de dicho robot, o responsable por los actos de éste.

Ahora bien, aunque la responsabilidad penal recayera sobre el garante humano del robot, sin lugar a dudas ante la comisión de delitos no podríamos permitir al robot seguir funcionando. Las razones parecen más pragmáticas que morales, es decir, deberíamos eliminar o desactivar permanentemente el robot simplemente por evitar una posible reincidencia. No obstante, llegado un determinado momento de sofisticación intelectual e incluso moral de los robots, podríamos justificar la punición de igual manera que con los humanos, es decir, también por razones disuasorias –evitar que otros robots sigan ese comportamiento- o incluso por razones de ideología retribucionista –todo el que hace un mal debe pagar un castigo-. En cualquier caso, la razón más importante para ejecutar la pena de desactivación o eliminación del robot sería, como se ha dicho, la de evitar que el delito vuelva a suceder. En estos casos se aplicaría el anteriormente mencionado “kill switch” o botón de la muerte. Los delitos susceptibles de comportar dicha pena tendrían que ser todos los que prevé el código penal, pues incluso delitos menores como el hurto o las lesiones leves causarían un gran caos social si pudieran ser cometidos con facilidad por robots, lo cual podrían aprovechar los propietarios de los mismos para tratar de utilizarlos con fines maliciosos.

Respecto a los robots como sujetos pasivos, podemos argumentar de manera idéntica a la anterior, ya que, si antes en el apartado ético hemos dudado acerca de la atribución de un estatus moral a los robots, en realidad también podemos plantearnos proteger a los mismos penalmente. Ahora bien, igual que en el caso anterior, otorgar una protección penal a los robots de la misma manera que la que se otorga a las personas puede comportar consecuencias ético-jurídicas negativas como sería la de desvirtuar valores como la dignidad humana. Además, probablemente sea una idea que no tendría

el consenso de la sociedad. No obstante, es una cuestión que también podría llegar a plantearse, eso sí, en un plazo de tiempo bastante largo.

Habiendo dejado claro que de momento no es factible plantearse la atribución de responsabilidad penal como sujeto activo a un robot, hemos de plantearnos el uso de los mismos como instrumento del delito y la responsabilidad de la persona humana que haya detrás como autora mediata del delito. Si tenemos en cuenta el escalofriante dato que apuntan los expertos en inteligencia artificial de la *National Crime Agency* y de *Europol European Cybercrime Center*, quienes prevén que para el año 2040 los robots cometerán más crímenes que los humanos; amén de las incidencias de los coches sin conductor y los *drones*, hemos de establecer un sistema eficaz que transfiera la responsabilidad penal a una persona humana. Todo apunta a que, en un futuro a corto y medio plazo este sistema debería convertirse en una suerte de responsabilidad penal objetiva, similar al sistema de responsabilidad civil de nuestro artículo 1903 CC, el cual establece: *“La obligación... es exigible, no sólo por los actos u omisiones propios, sino por los de aquellas personas de quienes se debe responder. Los padres son responsables de los daños causados por los hijos que se encuentren bajo su guarda...”* De esta manera, y pese a que responder penalmente por los actos de otro no parece muy conforme a nuestra manera de concebir el derecho penal, esta solución parece la más adecuada, teniendo en cuenta que a este sistema habría que añadirle, al igual que en el sistema civil (art. 1904 CC), una especie de acciones de regreso. Es decir, si nos situamos en un caso en el que un robot ha causado lesiones a una persona, ha de haber otra persona, según lo que conste en un eventual- e indispensable- registro de robots, que sea titular de dicho robot, con lo cual parece que en principio sería ésta –o sea, la persona titular del robot- la que respondería penalmente, y aquí entrarían las acciones de regreso, mediante las cuales dicha persona podría desplazar la responsabilidad penal hacia la que, según la cadena de responsabilidad, sería culpable. En otras, palabras, tendríamos una presunción *iuris tantum* de responsabilidad criminal del titular del robot, que podría ser desvirtuada probando que el fallo del robot no se debió a su uso doloso o imprudente, sino a fallos en su diseño o montaje, con lo cual el primer garante de la cadena de responsabilidad podría desplazar dicha responsabilidad al siguiente garante de la cadena –fabricantes, diseñadores, comerciales o incluso terceros-.

El principal problema al que se enfrenta esta responsabilidad penal objetiva en los delitos cometidos por robots sin duda es el encaje con la presunción de inocencia. En efecto, parece que considerar a alguien responsable criminalmente por un hecho que el no cometió vulnera totalmente dicha presunción. Ahora bien, para que el tipo penal encajara con dicho principio, podría catalogarse el delito como omisión imprudente. Es decir, la persona titular del robot, sabiendo de su especial situación de responsabilidad hacia éste, comete una imprudencia muy grave al permitir que el robot cometa el delito, y precisamente éste sería el hecho típico: omitir imprudentemente los cuidados necesarios para evitar que el robot propiedad de uno mismo cometa ilícitos penales. No obstante, parece que, de establecerse un ilícito penal de este tipo, lo más habitual fuera, como se ha dicho, poder repercutir esa responsabilidad penal –siempre y cuando se pruebe que el acusado puso la diligencia esperada para evitar el hecho típico- en fabricantes, diseñadores, distribuidores o terceros. Dicha repercusión de la responsabilidad sería más factible hoy con la posibilidad de que las personas jurídicas

puedan ser responsables penalmente. Se podría añadir que en caso de que una persona fuera condenada por omisión imprudente de vigilancia y cuidado de un robot de su propiedad que ha cometido un delito, la pena nunca pudiera ser de privación de libertad, sino únicamente de multa. Otra posible solución sería considerar el acto del robot efectivamente como un delito, y por tanto, sustanciar el caso ante un juez de lo penal, pero juzgar a los garantes solamente como responsables civiles –responsabilidad civil derivada del delito- pudiendo únicamente condenarlos a obligaciones de dar (cantidades pecuniarias), hacer (destruir el robot) o no hacer (retirada de la capacidad para ser titular de más robots en el futuro).

En conclusión, puede resultar un tanto frívolo y macabro el hecho de que se abra la posibilidad de pasarse la patata caliente de la responsabilidad criminal del mismo modo que se hace, por ejemplo, con la responsabilidad civil cuando se derrumba una casa, ya que aquí no se está discutiendo solamente el patrimonio de los garantes, sino también su libertad –si se trata de una persona física-. Aun con sus obvias razones para criticar este sistema, es una posibilidad que queda abierta a partir de ahora. Sea como fuere, esto es tan sólo la larva o la semilla de un debate que a buen seguro se hará extenso entre los juristas cuando llegue el momento, aunque haríamos bien de comenzar a abrirlo para no tener que buscar soluciones legales *ex post*.

6. Conclusiones

Tras sumirnos en una visión panorámica acerca de las cuestiones éticas y morales de la inteligencia artificial más relevantes y que más inquietudes proyectan hacia el futuro, se pueden extraer ciertas conclusiones. A continuación, expondré lo que en mi opinión es más relevante de todo lo que cuentan las líneas anteriores. Se trata de una visión crítica y con perspectiva de las cuestiones que se han tratado, no de ninguna de ellas en concreto, sino del fenómeno de la inteligencia artificial y de la robótica en general y todo lo que jurídica, política y socialmente podemos y debemos llevar a cabo para sacarle el máximo partido.

Apocalípticos e integrados.⁷⁰ Ha llovido bastante desde que el italiano Umberto Eco publicara un libro así titulado. La contraposición que Eco planteaba en dicha obra parece darse también hoy entre los individuos de las distintas sociedades que conforman el panorama internacional. Así, los “apocalípticos” de Eco serían hoy en día, si extrapolamos su significado al fenómeno del auge tecnológico en general, y por lo que a este trabajo respecta, a la inteligencia artificial y la robótica, aquellas personas que tienen una posición escéptica y de rechazo a todo desarrollo tecnológico que implique grandes cambios. No son sólo considerados como cambios por los apocalípticos los productos innovadores que se ofrecen en el mercado –por ejemplo, un sofisticado robot capaz de cuidar de personas mayores- sino también en las variaciones sustanciales que

⁷⁰ “*Apocalípticos e integrados*” (1964) es una obra del italiano y recientemente fallecido Umberto Eco, uno de los grandes escritores y filósofos del siglo XX, y quien fue autor de conocidas novelas y ensayos éticos, filosóficos y lingüísticos. En este libro, el autor nos contrapone dos modelos o posiciones opuestas ante la cultura, dando argumentos a favor y en contra de cada una de ellas: los apocalípticos y los integrados.

el mencionado fenómeno tecnológico puede producir en el modo en el que los individuos se relacionan con el medio físico –el planeta o el universo- entre ellos, o incluso con otras especies terrestres - flora y fauna- y, por qué no, con otras civilizaciones extraterrestres. Así pues, los apocalípticos, valga la redundancia, considerarían como una suerte de apocalipsis un escenario en el que el ser humano desarrolle su tecnología hasta llegar a límites que sobrepasan nuestra lógica y entendimiento, e incluso los más apocalípticos rechazarían cualquier cambio sin necesidad de llegar a tales límites.

Por otra parte, nos encontraríamos con los integrados, los cuáles serían individuos con tendencia a realizar interpretaciones demasiado benévolas, poco prudentes e incluso temerarias ante el cisma tecnológico venidero. Para un integrado cualquier innovación que pueda modificar, mediante la sofisticación, nuestra forma de vivir, relacionarnos y entender el mundo sería positiva. El integrado, yéndose al polo opuesto que el apocalíptico, peca de falta de rigor y de prudencia a la hora de tratar el fenómeno tecnológico. Una ciudadanía demasiado “integrada” corre el riesgo de convertir la tecnología en el centro alrededor del cual giran todas sus metas, objetivos y aspiraciones.

Estas dos posiciones tan generales, tendrían su especificación o concreción en diversas teorías que varían según el ámbito científico-tecnológico del que se hable. En este trabajo se han planteado, por ejemplo, dos posturas totalmente opuestas acerca de las pautas éticas en relación con el fenómeno tecnológico, como son el bioconservadurismo – posición apocalíptica en este caso- y el transhumanismo –posición integradora. Ambas posiciones tienen sus argumentos a favor y en contra, pero esencialmente ambas pecan de acercarse peligrosamente a los extremos. Los argumentos a favor de cada una de ellas han quedado bien resumidos con el bioconservadurismo y el transhumanismo. En síntesis, podríamos afirmar que de una postura apocalíptica podemos extraer, como inputs positivos para un futuro más próspero y sostenible, valores como la prudencia, la moderación o la preservación prioritaria de la vida humana y su dignidad y libertad de acción. Por el contrario, esta postura tendría el inconveniente de ser demasiado cerrada ante el progreso, ponderando los valores anteriores con un peso demasiado grande como para que se pueda avanzar.

Por la otra parte, de la postura integradora podríamos destacar como rasgos positivos el espíritu de superación humano, capaz de generar tecnología que incluso puede llegar a superarle en inteligencia y todo el potencial de bienestar que ello puede suponer. Ahora bien, los peligros de esta posición son innumerables, pues el hecho de abordar el progreso tecnológico con excesiva euforia puede llevarnos a no adecuar las medidas regulatorias necesarias para no convertir el mencionado potencial de bienestar en potencial destructivo. Un paralelismo que resume bien esta contraposición serían los estándares de prueba en el derecho penal.

¿Qué ocurre si se baja el nivel de exigencia probatoria en sede penal, de manera que sea fácil en exceso tener por probado falsamente un delito? Pues que podríamos dar lugar a una sociedad reprimida, en la que las personas no desarrollaran su potencial de vida por temor a ser sancionados por unos estándares demasiado laxos. Además de

encontrar inocentes en prisión, la sociedad viviría reprimida. Por el contrario, ¿qué ocurriría si elevásemos el nivel de exigencia probatoria de manera desproporcionada? Pues que probar los delitos sería extremadamente difícil, y a parte de encontrar personas culpables en libertad, daríamos lugar a una sociedad caótica, delictiva, violenta y corrupta, debido a que no sería posible en casi ningún caso demostrar en sede penal la comisión de un delito, con lo cual no se conseguiría disuadir a los criminales ni evitar el crimen. La alusión a los estándares del derecho penal no es sino una analogía de lo que sucede con la tecnología en general, y por lo que aquí atañe, con la inteligencia artificial, en relación a la cual surge un nuevo reto para el ser humano, consistente en aunar las virtudes de las dos posiciones extremas anteriores y construir con ellas los cimientos de la base ideológica, ética y jurídica de dicha tecnología.

El cambio ya ha empezado. Se hace evidente, al escudriñar con un poco de empeño todo lo que nos puede ofrecer la tecnología, en el presente y en el futuro, que no podemos mirar hacia otro lado. De ello se desprende, que las bases éticas han de ser lo suficientemente sólidas como para soportar el peso de la estructura jurídica que pretenda adaptar la inteligencia artificial y la robótica a nuestra sociedad. El término “sólidas” no quiere decir necesariamente “restrictivas”, más bien se refiere a que las pautas o estándares éticos mínimos han de ser coherentes, y en la línea de lo anterior, equilibrados. A partir de aquí, está claro que los juristas hemos de evolucionar, y además a un ritmo vertiginoso. Tradicionalmente el derecho se crea o modifica para regular situaciones a medida que van sucediendo. Los juristas suelen ir tarde. No obstante, en temas que conciernen a una tecnología tan potente sería realmente peligroso un exceso de relajación. No debemos pensar que las cuestiones que se tratan en este es un trabajo son futuristas. Son cuestiones de rabiosa actualidad. En ese sentido, parece que las instituciones europeas están comenzando a tomar medidas. No obstante, al menos en el estado español, las muestras de voluntad de adaptarse al cambio de paradigma son escasas. Las cuestiones que aquí se han tratado deberían de empezar a aparecer constantemente en los debates parlamentarios, tanto estatales como regionales.

Desde un punto de vista sociológico, parece realmente importante comenzar a concienciar a los ciudadanos de lo que viene, y de cómo tendríamos que adaptarnos a ello. En ese aspecto, me parece realmente necesario desde ya comenzar a introducir en las aulas de estudio estas cuestiones. La educación para la ciudadanía o ética que sustituye a la religión en los centros públicos debería comenzar a incluir la adaptación y el trato con las nuevas tecnologías para que las personas las conozcan y sepan hacer un buen uso de ellas con tal de extraer beneficios y evitar perjuicios. En lo que atañe a los estudios universitarios, en especial a las facultades de derecho, es también una necesidad comenzar a abandonar ciertas disciplinas un tanto arcaicas u obsoletas y con poco peso en la actualidad y empezar a formar a los juristas del futuro en cuestiones de nuevas tecnologías. No en vano, es un hecho que tanto abogados como jueces, fiscales, notarios, registradores y todas las personas que conforman la retahíla de profesiones derivadas del derecho llevarán a cabo en pocos lustros funciones considerablemente diferente que las que ahora vienen realizando.

Ello no significa que la esencia de las profesiones jurídicas haya de desaparecer, al menos en un futuro a corto y medio plazo. En efecto, el juez seguirá juzgando, el fiscal seguirá

ejerciendo la acusación y el abogado seguirá asesorando o defendiendo al cliente. Sin embargo, y como hemos podido observar al inicio de este trabajo, con la aparición de los sistemas expertos jurídicos y su aplicación a los diversos ámbitos del derecho, las funciones de los profesionales jurídicos variarán sustancialmente. Parece que estas funciones se verán restringidas a tareas de selección de información ya procesada de forma muy eficaz por los sistemas y una posterior interpretación coherente de esa información –que a su vez puede ya venir interpretada por los sistemas expertos-. En otras palabras, los sistemas expertos optimizarán el proceso de los juristas de tal manera que lo que constituya el factor diferenciador entre un lego y un experto en derecho no será tanto el conocimiento legal, sino la experiencia. Parece claro que hay ciertas características intangibles de los expertos jurídicos –creatividad en un abogado o sentido común en un juez, por ejemplo- que van a tardar todavía muchísimos años en ser sustituidas, si es que permitimos que puedan llegar a serlo. Por todo ello, cabe reiterar la urgencia en comenzar a instruir a los juristas para que sean más competentes en cuestiones no solamente relacionadas con la inteligencia artificial y los sistemas expertos sino con todo el alcance de la tecnología en general.

Invertir en educación y talento. Durante los últimos tiempos, estamos ya acostumbrados a escuchar cómo miles de jóvenes españoles, o formados en España emigran a otros países en pos de oportunidades que no encuentran aquí. Nuestro tejido académico tiene primera materia más que suficiente para liderar proyectos del más alto nivel. Nuestras universidades rebosan de mentes capaces de cambiar el mundo. Sin embargo, es habitual escuchar cómo talentos locales que aquí estaban condenados a la cola del INEM triunfan en el extranjero. Vivimos una época excepcional, en la que en un año se producen tantos descubrimientos como en una década antaño. Ello hace que los plazos tradicionales se vuelvan relativos en esta materia. Dichos descubrimientos, pues, con los recursos adecuados, tienen un potencial inmenso para generar bienestar. Así pues, todo se reduce a una cuestión de voluntad política. Desgraciadamente, si no hay dinero de por medio no se pueden financiar proyectos. Un ejemplo, estrechamente relacionado con la inteligencia artificial es el del almeriense Ramon González Sánchez, quien después de no encontrar oportunidades en España triunfa hoy en el *Massachusetts Institute of Technology* trabajando en proyectos de inteligencia artificial y robótica para la NASA o la OTAN.⁷¹ Por ello, se ruega encarecidamente a gobiernos estatales y autonómicos que sean valientes y que doten de recursos al sector de la investigación en todos sus ámbitos. Cabe recalcar que el progreso es una cadena conformada por muchísimos eslabones. Los rostros que suelen aparecer en los medios son los de los ingenieros que aportan la innovación técnica, o la de los políticos que se atribuyen los méritos, aunque sin embargo detrás de ello hay muchas otras personas detrás que, en su campo, trabajan por adaptar las otras disciplinas a la vanguardia de la técnica. Los robots no se implantarán eficazmente sin juristas. Tampoco sin economistas, ni sin sociólogos, filósofos, epistemólogos, psicólogos o antropólogos. Puede que ésta parezca una reflexión obvia, pero su redundancia parece ser insuficiente para espolear a una clase política que de momento juega a verlas venir, a contemplar como los demás producen e innovan más y mejor, para más inri con mano de obra formada y costada por nosotros. Hay dos maneras de invertir para tratar de crear bienestar: se puede invertir

⁷¹ Para saber más sobre este autor, se recomienda la lectura de: González Sanchez R. “Innovar o morir: esa es la cuestión”

en aviones de combate o se puede invertir en libros, en seguir enriqueciendo a los oligopolios o en vehículos eléctricos, en soldados o en profesores, en definitiva en represión o en progreso.⁷² Hace tiempo que la humanidad está inmersa en una lucha contra sí misma. Una batalla que, no obstante, no se puede ganar con armas, sino con conocimiento.

Inteligencia artificial: una tecnología para todos. Por último, y siguiendo en la línea de lo anterior, hemos de reivindicar una igualdad social en la implantación de la inteligencia artificial. Como se ha visto, esta idea aparece reflejada en los principios de Asilomar (principio de prosperidad compartida). Lo que este principio significa y representa es la no utilización de esta tecnología con intereses concretos y determinados. Es obvio que, por mucho que los ciudadanos se muestren en contra, la élite utilizará un poder como este para seguir dirigiendo desde las sombras la orquesta del planeta – la cual sigue tocando resignada mientras el mismo se hunde-, pero más allá de ello, los ciudadanos han de ser conscientes que han de cambiar lo poco que está en sus manos. La inteligencia artificial debería ayudar a salvar vidas –en la carretera, en los hospitales,- a eliminar vilezas como la corrupción, la trata de seres humanos, el esclavismo encubierto –o no tan encubierto- , a formar personas más sabias, con más recursos para pensar por sí mismas y defenderse, y a eliminar las desigualdades económicas, sociales y culturales. Todo ello suena a utopía, pero el primer eslabón de la cadena nos corresponde los ciudadanos, y consiste en reclamar a los gobernantes lo que nos pertenece: una igualdad efectiva en todos los aspectos, plasmada en la realidad, y no solamente en textos jurídicos tan idealizados como ineficaces.

⁷² (12 de marzo de 2017) “España estudia comprar a EEUU el cazabombardero más caro” *El Mundo*. Recuperado de : <http://www.elmundo.es/espana/2017/03/12/58c460fbe2704ee6678b45fe.html>