



ACTA RESUMEN Y CONCLUSIONES DE #ROBOTIURIS17

16 DE NOVIEMBRE DE 2017

El pasado 16 de noviembre se celebró en Fide la II Jornada sobre los aspectos legales de la Robótica, **#Robotiuris17**. A continuación, recogemos los resúmenes de algunas sesiones.

Fundación para la Investigación sobre el Derecho y la Empresa

c/Serrano 26, 4º Dcha, 28001 Madrid.
www.fidefundacion.es
fidefundacion@fidefundacion.es
[@fide_fundacion](https://twitter.com/fide_fundacion)

Patrocinado por:



Microsoft



LEFEBVRE
EL DERECHO

FUNDACIÓN
GARRIGUES

Contenido

¿Cómo aprenden las máquinas? Entendiendo la inteligencia artificial y las distintas tecnologías que la componen (deep learning, cognitive computing, machine learning, neural networks, etc.).....	2
El sesgo de los algoritmos y del big data ¿Cómo nos afecta en la práctica?	5
Las máquinas inteligentes como creadoras y autoras de obras protegidas por propiedad intelectual.....	9
El papel de la educación en la sociedad digital	11
Ética y responsabilidad en el diseño de robots e inteligencia artificial	14
The future of Artificial Intelligence.....	16
Killer robots y sistemas de armas autónomos en el marco del derecho internacional. Implicaciones, desafíos y riesgos	17

¿Cómo aprenden las máquinas? Entendiendo la inteligencia artificial y las distintas tecnologías que la componen (deep learning, cognitive computing, machine learning, neural networks, etc.)

Ponente: *Ramón López de Mántaras Badía. Profesor de Investigación del Consejo Superior de Investigaciones Científicas (CSIC). Fundador y Director del Instituto de Investigación en Inteligencia Artificial del CSIC. Miembro del Consejo Académico de Fide.*

Ramón López de Mántaras compartió con los asistentes a esta sesión los **principales progresos y los problemas a los que se enfrenta la Inteligencia Artificial en la actualidad.**

Comenzó señalando que el concepto mismo de Inteligencia Artificial (IA) forma parte de la informática, aunque con el paso del tiempo esta rama ha alcanzado un cariz más **interdisciplinar** debido a las grandes implicaciones que ha tenido la IA en todos los campos del conocimiento humano. A día de hoy allá dónde miremos podemos encontrar ejemplos de IA en nuestra vida cotidiana como es el caso del transporte, de la logística o de los diagnósticos médicos por lo que no es exagerado afirmar que la misma se ha hecho omnipresente en nuestra rutina diaria.

Es necesario apuntar que la IA no deja de ser diferentes **modelos matemáticos** que tratan de reproducir un modelo humano, en este sentido se utilizan sistemas en una combinación de software más hardware a gran velocidad que tienen como resultado la creación de diversas competencias en las máquinas que se asemejan y se asocian a la inteligencia humana como pueden ser comunicar, percibir, hablar...

Tenemos que tener en cuenta que existen **palabras “maleta”** como inteligencia o aprendizaje y que cuando aplicamos estas habilidades a las máquinas no hablamos de las mismas capacidades que las que poseen los humanos. En esta línea, se puso como ejemplo la descripción realizada por un robot de una escena en la que aparece una jugadora de tenis frente a la descripción que haría una persona.

El ejemplo antes descrito explicó la diferencia entre la **capacidad y la inteligencia humana** que es versátil ya que tiene el poder de una comprensión profunda de una escena o de un lenguaje que no se encuentra explícito a primera vista en comparación con las capacidades de las inteligencias artificiales que a día de hoy son muy específicas. Así pues, los humanos gozan de conocimientos de sentido común creados por la pertenencia a diversos pueblos que tienen una cultura común frente a la IA que carece de un conocimiento general de lo que le rodea.

Otro ejemplo que se puso en relación con este contexto fue el referido al software Alpha Go que es un programa que juega al ajedrez a nivel de Gran Maestro ganando 20 partidas frente a una que es capaz de ganar el ser humano. Al ver estos resultados los

humanos tendemos a extrapolar estos datos a cualquier acción sin darnos cuenta de que la IA que es capaz de jugar al ajedrez es completamente incapaz de jugar a las damas a pesar de ser un juego mucho más sencillo. Sin embargo, las capacidades humanas son infinitas ya que cualquier jugador de ajedrez puede aprovechar sus conocimientos sobre este juego para, en cuestión de segundos, jugar a las damas perfectamente.

En cuanto al aprendizaje automático de la IA se analizó el ejemplo del proyecto The Next Rembrandt. Aquí lo que se comprobó es si una máquina analizando una infinidad de cuadros de un pintor, en este caso Rembrandt, podía copiar y reflejar fielmente el estilo completo del pintor.

Los resultados de este experimento fueron increíbles puesto que los críticos de arte no fueron capaces de detectar que la obra creada por la IA no era una obra original del pintor holandés. Esto lo que demuestra es que los avances en cuanto a las técnicas de aprendizaje de la IA están siendo enormes y se están dando a pasos agigantados.

Un campo en el que también cada día interviene más la IA es en el de la medicina y en particular en el diagnóstico médico. Así pues, son innumerables los ejemplos existentes a día de hoy en cuanto a la ayuda proporcionada por la IA para detectar enfermedades o para predecir las posibilidades de sufrir problemas graves de salud en el futuro tanto cercano como lejano.

Todos estos ejemplos de IA suponen el éxito de las máquinas basándose en la disponibilidad del acceso a una enorme cantidad de datos en forma electrónica además de una capacidad de cálculo grande.

Llegados a este punto y después de poner de manifiesto estos ejemplos de IA fue muy necesario interrogarnos sobre estas cuestiones: ¿Puede hacer la IA cualquier cosa? ¿Puede la IA resolver cualquier tipo de acciones? A juicio del ponente y a día de hoy la respuesta es un no rotundo ya que **la IA necesita de más planificación, comunicación y raciocinio** que es lo que otorga robustez a la máquina ya que la misma no es versátil. En esta línea también se señaló que las máquinas no tienen la capacidad necesaria para llenar los huecos en el conocimiento lo que genera a efectos prácticos enormes limitaciones.

Estas cortapisas en el desarrollo de las IA ayudaron al ponente a poner otros ejemplos en los que pudimos observar claramente los fallos que tienen las máquinas. En este contexto, se hizo referencia al programa Google Translate cuyo fallo radica en que la IA hace un análisis superficial de las frases y no semántico por lo que la máquina no entiende el alcance del significado de lo que se quiere traducir.

Otro ejemplo en el que se pone de manifiesto la capacidad limitada de la IA es aquella generada por problemas con la percepción visual tanto en el análisis como en la posterior descripción de fotos a partir de otras similares. Aquí el problema reside en la incapacidad de la máquina para hacer un análisis profundo de lo que ve.

Los riesgos actuales de la IA pivotan en torno a la **privacidad de datos sensibles**, la autonomía y el desempleo.

En cuanto a la autonomía debemos afrontar la problemática de si es conveniente o no dotar a la IA de una gran independencia y por tanto sacar la decisión humana de la máquina.

A día de hoy hay diferentes prototipos de **coches autónomos**, pero no se ha creado todavía el prototipo de un coche que se conduzca de manera independiente sin ninguna posibilidad de intervención humana (nivel 5 de conducción autónoma) ya que para que la conducción se produzca de manera autónoma se necesita de una gran robustez de la IA que a día de hoy es impensable. Este hecho se produce porque una conducción completamente autónoma se tiene que basar en un sistema integrado el cual tendría que ser caracterizado de todos los componentes de una inteligencia humana como son la percepción, planificación, razonamiento, aprendizaje, comunicación...

Por último, se abordó por el ponente la cuestión de si los humanos debemos temer que una máquina llegue algún día a pensar como nosotros. Su respuesta fue contundente ya que dejó claro que a su juicio los robots no son la mayor amenaza para el ser humano en un periodo de 20-30 años ya que **la inteligencia de las máquinas no aparecerá de forma súbita como tampoco su consciencia.**

El sesgo de los algoritmos y del big data ¿Cómo nos afecta en la práctica?

Ponentes: *Marlon Molina*, Director de ComputerWorld University y *Antonio Muñoz*, Senior Legal Counsel en Telefónica Digital.

Marlon Molina y Antonio Muñoz compartieron con los asistentes a esta sesión **la problemática de los sesgos a la hora de la programación, así como del extracto posterior de datos.**

Marlon Molina comenzó su exposición haciendo hincapié en el hecho de que la tecnología es un fenómeno nuevo ya que sólo tiene un recorrido de cuatro décadas por lo que todo lo que estamos viviendo es novedoso y muchas veces no tenemos las herramientas necesarias para enfrentarnos a ello.

En su opinión, el *big data* se puede definir como aquellas técnicas que nos ayudan a tomar decisiones a través de datos que se encuentran conectados, esto se produce a través del análisis sistemático de grandes datos a mucha velocidad. Dicho en otras palabras, el *big data* es capaz de construir datos de diferentes procedencias pero que están ligados para sacar un aprendizaje y poder llegar hasta la sabiduría. En este contexto es muy importante señalar y no perder de vista que el objetivo de la Inteligencia Artificial (IA) es la **toma de decisiones** a una velocidad vertiginosa, en tiempo real.

En otro punto de la sesión se puso de relieve que el *big data* hoy en día está inmerso en nuestra sociedad y no somos verdaderamente capaces de darnos cuenta. Así por ejemplo el análisis de datos a gran velocidad puede ser clave a la hora de que un piloto resulte victorioso en una carrera ya que el estudio pormenorizado de los datos puede asistirle para que éste pueda tomar mejores decisiones basadas en datos reales, o un piloto estándar mantenerse competitivo. Gracias a este ejemplo debemos ser conscientes de que si yo no hago un uso de la tecnología y el resto de mis rivales sí que lo hace no voy a poder ser competitivo y me voy a ir quedando atrás.

Otro ejemplo es la utilización del *big data* en las **financieras y bancos** cuyo principal objetivo es llegar hasta ciertos perfiles de clientes según parámetros previamente delimitados. En este sentido, se trata de separar personas en función de a qué grupos me quiero dirigir. Esto implica que discriminación relativa al grupo en el que se pertenezca, un algoritmo determina si se es capaz de acceder a un producto o no.

Como se puede observar la utilización del *big data* tiene un gran impacto en nuestros días y es primordial hacer un buen uso de los datos para no caer en malas prácticas. Así pues, y según expertos políticos a nivel mundial, el *big data* fue clave para las victorias presidenciales de Obama como en la de su sucesor Donald Trump al igual que el *big data* tiene un gran impacto para el Gobierno chino como forma de conocer y condicionar la fidelidad de su población al partido comunista.

Para finalizar el conferenciante quiso poner de manifiesto que hay cuatro dominios de impacto del *big data* como son:

- **Tecnología:** Así pues, el sentido de la tecnología tiene que tener siempre un trasfondo económico que sustente las nuevas creaciones.
- **Sociedad:** Como hemos podido observar en estas cuatro décadas en las que la tecnología ha irrumpido en nuestras vidas los humanos y la sociedad en su conjunto están ávidos de nuevas tecnologías y aprenden muy rápido.
- **Política:** Es necesario que los estamentos jurídicos tomen un rol activo en cuanto a las nuevas tecnologías para legislar los ámbitos de acción de las mismas dando un marco responsable y coherente a todo el sistema.
- **Economía:** Es muy importante tener en cuenta que gracias al análisis de *big data* basado en datos de compradores a escala mundial podemos incidir en sus deseos de una manera brutal ya que podemos guiar sus deseos tendentes a adquirir un determinado producto o a disfrutar de un servicio. Esto es muy relevante y es necesario por tanto de una responsabilidad de la industria.

Por su parte, **Antonio Muñoz** habló sobre que uno de los inconvenientes más recurrentes a la hora de tener en cuenta la IA es el provocado por el sesgo. Éste se debe entender como el error sistemático relacionado con los datos provocando un desdoblamiento entre la realidad y el resultado obtenido.

Una apreciación importante que quiso realizar el conferenciante fue la relativa a que **cuando hay sesgo puede haber discriminación** pero que hay veces que la misma se da sin existir un sesgo previo. Hay mucha desconfianza por parte de los humanos en general entre el sesgo humano y el sesgo del *big data*, de este modo por ejemplo en EEUU en 2014 se manifestó que los sesgos se pueden paliar con una mayor analítica y con medidas que mitiguen esas malas consecuencias. Los sesgos y por tanto los errores consiguientes pueden provenir de:

- **la fuente de datos:** es necesario advertir que sin datos de calidad no se pueden hacer predicciones ni mucho menos tomar decisiones de manera fiable.
- **el algoritmo:** es el propio analista o programador de la máquina quién genera ese sesgo que después la máquina continúa.
- **la interpretación** de las conclusiones por el usuario.

Del mismo modo también existen **sesgos sociales** que nos ayudan a anticipar situaciones; lingüísticos (ejemplo traductor de Google que cambia el género de la persona según la determinación cultural de una profesión masculina o femenina: doctora en castellano y doctor en turco) o cognitivos (ejemplo sesgo de la disponibilidad que afecta tanto al analista como al programador o jurista).

¿Qué **medidas** se pueden tomar para evitar los sesgos? Mejorar las fuentes otorgando más representatividad de las mismas en la analítica, controlar al analista, formar un comité ético que valore los resultados, tener un sentido crítico del control de las fuentes, así como tener un equipo multidisciplinar a la hora de tomar decisiones sobre los datos arrojados por una IA.

Tecnología y salud. Retos de ciborgs, prótesis biónicas, exoesqueletos o nanobots

Ponente: Pedro García Barreno, Doctor en Medicina. Catedrático emérito de la Universidad Complutense. Miembro del Consejo Académico de Fide y José Luis Pons, Director del Grupo de Neurorehabilitación, Departamento de Neurociencia Traslacional. Instituto Cajal. CSIC.

Pedro García Barreno y José Luis Pons compartieron con los asistentes a esta sesión **los retos a los cuáles se tienen que enfrentar los profesionales sanitarios por el uso de la tecnología en su área de acción.**

Pedro García Barreno comenzó señalando que en la sociedad en la que vivimos desde que nos levantamos hasta que nos acostamos estamos inmersos en un mundo de ciencia, tecnología e ingeniería. Asimismo, afirmó que desde el punto de vista ético en medicina no hay restricciones en el uso de la tecnología en el cuerpo humano, aunque se han puesto límites antroposomáticos tanto en el cuerpo como en la cognición.

El **uso de la tecnología en la medicina** va mucho más avanzado que los humanos por ejemplo así en los años 80 se crearon las primeras píldoras abortivas, en los años 90 se produce la clonación a partir de células especializadas o más recientemente hemos asistido a la creación y perfeccionamiento de técnicas para reproducción asistida.

El problema de la medicina avanzada es el error médico como venía ocurriendo en el pasado. Un ejemplo de esto fue la implantación de unos electrodos en el cerebro pero que no estaban realizando su función ayudando al tratamiento del paciente debido a que los mismos no fueron colocados en una posición correcta.

Otro problema con el que se encuentra la medicina avanzada es la existencia de una **brecha** entre diferentes agentes y estamentos de la sociedad en relación con el uso de las nuevas tecnologías aplicadas a la medicina ya que esta nueva medicina tiene unos costes que no todo el mundo puede asumir.

Esta nueva medicina más avanzada gracias al uso de la Inteligencia Artificial (IA) abre la puerta a la **mejora de las capacidades físicas y cognitivas** de los hombres lo que a su vez pone de relieve el tema de las manipulaciones genéticas no sólo en personas adultas sino también en menores o incluso nonatos lo que va a generar no pocos problemas en la sociedad. En este sentido, los médicos piensan que la evolución ya no dependerá del azar sino de la mutación genética realizada con diferentes técnicas.

En palabras de Pedro García Barreno, es importante que **sepamos trazar una línea entre la medicina y el transhumanismo**. Así pues, el transhumanismo es un movimiento intelectual que tiene como objetivo principal transformar la propia naturaleza humana mediante el desarrollo y fabricación de tecnología, que mejoren las capacidades humanas, tanto a nivel físico como psicológico o intelectual mientras que la medicina es la prevención, la predicción, el diagnóstico, el pronóstico, el tratamiento

y la rehabilitación de enfermedades utilizando entre otros medios los fármacos, cirugías, prótesis, fisioterapia y otras medidas complementarias.

José Luis Pons basó su exposición en las nuevas tecnologías que se están usando en el ámbito de la medicina y más en concreto en la **neurorrehabilitación**. La motivación última para el uso de la Inteligencia Artificial (IA) en el campo de la neurorrehabilitación es clave porque puede ayudar a superar desórdenes motores que dificultan en gran medida la vida de quienes los padecen.

La presentación se basó en la explicación tanto de los *wearable robots* también llamados **robots vestibles** como de los **exoesqueletos**.

Con respecto a las nuevas tecnologías que se están utilizando en el campo de la medicina puso el énfasis en la idea de que gracias a la utilización de robots en nuestro cuerpo humano hay una interacción que une estas máquinas con los humanos tanto de una manera física como cognitiva para promover cambios en la mejora del ser humano ya sea recuperando una función perdida o ayudando a realizar una acción que no se podía llevar a cabo hasta ese momento. De esta manera y gracias a la utilización de la tecnología en nuestro cuerpo se ayuda a un cambio en el sistema nervioso central del ser humano por lo que la misma es capaz de cambiar o modular el comportamiento humano.

Los exoesqueletos son robots que disponen de motores y se visten en el cuerpo humano para movilizar y asistir a los humanos en los movimientos que se quieran desarrollar. En este sentido, un exoesqueleto sirve para promover un reaprendizaje y una asistencia que permiten al humano sustituir una función perdida en su cuerpo o su mejora funcional.

La tecnología abre posibilidades tremendas cuando la misma se aplica al cuerpo humano ya que es capaz de predecir lo que queremos hacer y asistirnos para alcanzar nuestro objetivo.

La Comisión Europea se encuentra preocupada por el uso de la tecnología en la medicina ya que se podrían generar problemas en ámbitos como la **protección de datos** de aquellos enfermos que utilicen I.A en su vida cotidiana, así como por aquellos avances médicos que supongan mutaciones o alteraciones enormes en las capacidades físicas o mentales de los hombres.

Las máquinas inteligentes como creadoras y autoras de obras protegidas por propiedad intelectual

Ponentes: *Juan Carrasco Linares, Socio responsable del área de TMT de Santiago Mediano Abogados, SLP y Joaquín Muñoz Rodríguez, Director internacional del área de Nuevas Tecnologías y Propiedad Intelectual de Ontier.*

Juan Carrasco Linares y Joaquín Muñoz Rodríguez compartieron con los asistentes a esta sesión de cuáles son los **principales retos en el seno de la propiedad intelectual a la hora de determinar una posible protección de las obras creadas por robots.**

Comenzaron señalando que el **machine learning** tiende a la toma de decisiones por parte de la Inteligencia Artificial (IA) lo que sin duda genera una libertad de la máquina en la que, por tanto, se tiende a que la intervención humana sea mínima. Al incorporar IA a las máquinas se ayuda a que las mismas tomen decisiones por ellas mismas y de manera automática.

En cuanto a la propiedad intelectual deviene por tanto fundamental conocer el **proceso creativo** interno de la máquina en el caso de no existir la intervención humana.

Los ordenadores son herramientas para llevar a cabo la obra intelectual pero cada vez debemos ser más conscientes de que el input del programador es menor por lo que es clave saber cuál es el grado de intervención humana y en qué momento se produce la misma.

Las preguntas fundamentales que cabe hacerse en el seno de la propiedad intelectual interrelacionada con la IA son: ¿se puede reconocer protección por derechos de autor? y ¿sobre quién recae la consideración de autor?

En la legislación española hay una equiparación entre el autor y la persona física/natural incorporándose más tarde la posibilidad o ficción jurídica por previsión ex lege de otorgar la autoría a personas jurídicas. En este sentido, las personas que están a favor de que una máquina sea autor argumentan que las personas jurídicas son entes capaces de ser sujetos de derechos y obligaciones a pesar de no poder crear ni desarrollar actividades intelectuales por lo que las máquinas que creasen obras con una cierta altura creativa deberían ser consideradas entes capaces de ostentar derechos de autor.

En el seno de la Unión Europea existe una recomendación denominada **Report with recommendations to the Commission on Civil Law Rules on Robotics** (2015/2103 (INL)) en la que en su literal 18 se recoge con respecto a la propiedad intelectual que: *“no hay ninguna disposición jurídica que se aplique específicamente a la robótica, pero que las doctrinas y los regímenes jurídicos actuales pueden aplicarse fácilmente a esta, aunque algunos aspectos requieran especial consideración; pide a la Comisión que apoye un enfoque horizontal y de neutralidad tecnológica para la propiedad intelectual*

en los distintos sectores en que se pueda utilizar la robótica". Esta recomendación lo que pone de manifiesto es la necesidad de una futura regulación en materia de IA en relación con la propiedad intelectual.

Hoy en día existen **dos posiciones contrapuestas** entre pensadores y juristas de la propiedad intelectual que piensan que una máquina podría ser considerada autor y aquellos que rechazan esta idea de plano. Entre los primeros sus argumentos a favor giran en torno a la idea de que algún día una IA pueda crear una obra original sin depender del input del programador es decir que sea capaz de realizar una creación ex novo. Asimismo, basan su perspectiva en que a día de hoy existen legislaciones que no exigen que el autor sea una persona física como por ejemplo en la *Copyright, Design and Patents Act* del Reino Unido.

En cuanto a aquellas posiciones reacias a otorgar derechos de autor a una máquina basan su enfoque en que las máquinas a día de hoy carecen de personalidad jurídica por lo que son entes que no son sujeto de derechos. Habría por tanto que plantearse primero la creación de un concepto diferente de persona después de la física y la jurídica que podría ser la robótica como paso previo para que las IA tuviesen legitimación y por tanto en etapas consiguientes poder ejercitar sus derechos.

Es en este contexto dual, en el que nos encontramos, en el que surgen numerosos interrogantes ya que debemos indagar en problemas relativos a qué es lo que ocurre cuando la IA de una máquina es suficientemente autónoma ¿a quién podrían atribuirse los derechos de autor de sus creaciones? Algunos autores sostienen que los derechos sobre el software original que ha derivado en una IA son del programador que para eso ha sido quien ha creado la máquina y ha invertido e incentivado una industria emergente, pero esto genera también dudas e interrogantes en el seno de una posible venta de la máquina. Así mismo, y como ya se ha podido apuntar antes podrían aparecer **problemas derivados de los derechos morales y la legitimación de las máquinas.**

A modo de conclusión los ponentes quisieron hacer hincapié en el hecho de que el **trinomio entre máquinas, proceso creativo y propiedad intelectual** es una cuestión cada vez más compleja y que por tanto requiere de una labor colectiva para decidir qué tipo de protección se querrá ofrecer a las creaciones surgidas de algoritmos inteligentes con una mínima intervención humana o sin ella ya que en un futuro no muy lejano no será fácil distinguir creaciones intelectuales del hombre frente a creaciones de máquinas inteligentes.

El papel de la educación en la sociedad digital

Ponente: José Antonio Marina, Filósofo y pedagogo.

José Antonio Marina compartió con los asistentes a esta sesión de cuál es **el papel de la educación en la era digital en la que vivimos**.

Comenzó señalando que las sociedades se guían por la velocidad por aprender, de este modo podemos considerar que la sociedad en la que habitamos está en un constante estado de emergencia ya que los hombres de hoy en día están hambrientos de nuevos conocimientos a través del uso de las nuevas tecnologías.

A nadie le cabe la menor duda de que los niños de nuestra sociedad son **nativos digitales**, es decir, son personas que desde su nacimiento han convivido con las nuevas tecnologías y que para ellos su uso es algo cotidiano. Este hecho, sin embargo, nos lleva a interrogarnos sobre si el uso que se hace de las nuevas tecnologías en el ámbito de la educación es algo positivo o por el contrario pernicioso ya que muchos de estos niños repiten como un mantra la frase de: “para que voy a aprender si lo puedo buscar”.

Tal vez muchos de los errores educativos que hemos cometido proceden de haber intentado educar la inteligencia cognitiva, proporcionando conocimientos a los niños, y la inteligencia emocional, intentando fomentar sus sentimientos agradables, pero descuidando la educación de la **inteligencia ejecutiva**. No debemos de olvidar que la memoria humana se basa en **bloques cognitivos, afectivos y musculares** por lo que la educación tiene que tender a ampliar estos contenidos mientras que los sistemas de evocación tendrán que ser los encargados del desarrollo de la inteligencia ejecutiva que es la que tiene el gran objetivo de dirigir la acción mental y física aprovechando nuestros conocimientos y emociones.

Es importante señalar que **la memoria es la inteligencia** por lo que el aprendizaje parte de esto y si cercioramos esta parte estamos poniendo cortapisas al conocimiento fructífero. Es por todo esto que las **nuevas tecnologías** se deben utilizar no como fuente de información, sino que las mismas se tienen que **integrar en el propio proceso de aprendizaje**.

Durante mucho tiempo, la Inteligencia Artificial (IA) se encalló por no ser capaz de reconocer patrones complejos y durante mucho tiempo no se llevó a cabo ningún proyecto interesante en este sentido. Esta situación cambia en 1986 con el procesamiento en paralelo también llamado redes neuronales artificiales que suponen una revolución en el campo de la IA ya que estas máquinas no sólo son capaces de aprender de la experiencia sin grandes complicaciones de software o hardware, sino que pueden resolver de una manera muy sencilla problemas que constituyen dificultades para máquinas tradicionales tales como el reconocimiento de patrones complejos, el procesamiento de imágenes o incluso la generación e interpretación del lenguaje natural.

Podemos afirmar por tanto que el gran salto exponencial se produce gracias al deep learning de las IA que confiere a las mismas las capacidades necesarias para aprender así pues y a modo de ejemplo, se puede poner de relieve que el big data lo que hace es copiar al cerebro humano para ser capaz de reconocer patrones.

La función de la IA según José Antonio Marina no es la de conocer, sino que por el contrario es la de **tomar buenas decisiones y actuar en cuestión.**

Ahora bien, vistas las capacidades enormes que tienen las máquinas y su evolución deberíamos interrogarnos sobre la cuestión de ¿qué pasará con los humanos? De la contestación de esta pregunta han salido defensores a ultranza de la IA como herramienta capital para los humanos y detractores de la misma. Entre los primeros, podemos apuntar la Universidad de la Singularidad creada en 2008 por Ray Kurzweil y en la que se busca reunir, educar, e inspirar a un grupo de dirigentes que se esfuercen por comprender y facilitar el desarrollo exponencial de las tecnologías y promover, aplicar, orientar y guiar estas herramientas para resolver los grandes desafíos a los que se enfrenta la humanidad en su conjunto.

En el otro lado de la balanza, podemos encontrar las posiciones de los detractores entre los que señalamos John Searle y el experimento denominado habitación china. Gracias a este experimento John Searle trata de rebatir la validez del test de Turing y la creencia de que una máquina pueda llegar a pensar. En consecuencia, a su juicio las máquinas no saben qué es lo que hacen ya que no son capaces de crear significados al usar solamente significantes.

Searle lo que se cuestiona es la IA como modelo epistemológico o psicológico para la **explicación de la mente humana y el funcionamiento del cerebro**, en este contexto el filósofo ataca directamente a una versión radical de la IA, según la cual “el cerebro es simplemente un ordenador digital y la mente es simplemente un programa de ordenador”.

Otro detractor del uso de la tecnología es Nicholas Carr quien afirma que por estas nuevas tecnologías estamos perdiendo el sentido crítico y profundo del ser humano convirtiéndonos en seres superficiales. Así pues, plantea que el uso constante de la tecnología podría estar afectando de manera profunda a nuestra biología cerebral y alterando la forma en que los humanos pensamos dicho lo cual no es una cuestión de tirar por tierra las bondades de la IA, sino que se trata de poner de relieve que junto con esas múltiples ventajas tenemos como contrapartida el triunfo de la superficialidad y la distracción.

Jason Lanier es otro pensador que también habla de la interacción entre las personas y las máquinas. Según este autor, cuando un ser humano empieza a usar un programa el propio individuo comienza a pensar de la misma forma que el programa, siguiendo un patrón determinado.

Lanier pone de relieve la deriva hacia la que internet nos está llevando y por eso para él es fundamental una reflexión crítica acerca de por qué el negocio está en la publicidad en lugar de en la calidad de los contenidos o en generar información llamativa en lugar de profundizar en la misma o buscar nuevas formas narrativas. En su opinión, el

totalitarismo digital es la nueva amenaza que acecha el orden social por lo que la única solución es reinventar internet y sus aplicaciones de tal forma que se piense mucho más en el individuo que en la máquina, en lo útil más que en lo rentable y en el progreso del ser humano en su conjunto más que en el resultado de manera inmediata.

Otro filósofo que también tiene sus reparos en el uso de la IA es Nick Brostrom quien en sus trabajos ha analizado el peligro de la superinteligencia artificial. Brostrom sostiene que no se trata de demonizar a las máquinas sino de realizar una reflexión profunda sobre la relación que tienen que tener tanto máquinas como hombres.

En opinión de José Antonio Marina es necesario que los humanos cambiemos el paradigma que teníamos hasta hoy en día sobre la inteligencia ya que esto va a ser clave para años venideros. Asimismo, es importante centrar la educación en cómo organizar la **memoria informática personal** y en cómo tomar las mejores decisiones basadas en esa memoria por lo que habrá de hacerse una selección previa sobre qué conocimientos son necesarios tenerlos en nuestra memoria y cuáles pueden estar en la memoria de un ordenador.

Ética y responsabilidad en el diseño de robots e inteligencia artificial

Ponentes: *Javier Torre de Silva*, Socio del Departamento de Derecho Público y Sectores Regulados de CMS-Albiñana y Suárez de Lezo. Letrado del Consejo de Estado (en excedencia). Premio Nacional de Licenciatura en Filosofía, **Amparo Grau**, Profesora Titular de Universidad, acreditada como Catedrática. Departamento de Derecho Financiero y Tributario. Facultad de Derecho. Universidad Complutense de Madrid y **Eduard Fosch-Villaronga**, Investigador Posdoctoral en la Queen Mary University of London, Miembro del Grupo de Estandarización de la SPARC y de la Foundation for Responsible Robotics.

Javier Torre de Silva, Amparo Grau y Eduard Fosch compartieron con los asistentes a esta sesión **la perspectiva ética en el diseño y responsabilidad de los robots**.

Javier Torre de Silva comenzó señalando que la ética en los robots es una necesidad y no una opción. En la medida en que delegamos la toma de decisiones en sistemas de inteligencia artificial, las **exigencias éticas que aplicamos a nuestras propias decisiones deben extenderse a las que tomen dichos sistemas**. Así pues, la ética que debemos inculcar a los sistemas de Inteligencia Artificial (IA) tiene que estar en línea con nuestros propios sistemas de valores. El problema que subyace con respecto a la ética y la IA es ¿cómo se enseña esto a una máquina? La ética no se basa sólo en un conjunto de normas morales y su aplicación consiguiente. Cualquier aproximación que se limite a una mera aplicación de normas generales resultará insuficiente.

Existe una Carta sobre robótica elaborada con la asistencia de la Unidad de Prospectiva Científica (STOA) de la *DG European Parliament Research Service* en el seno del Consejo de Europa en la que se insiste en la necesidad de que las máquinas gocen de parámetros éticos, pero estas reglas, aunque inspiradoras, no solucionan el problema.

Es difícil encajar o traducir a código fuente por ejemplo la ética de Aristóteles o Kant, que no se reducen a elencos de normas. Además, la ética aplicada a un ordenador no es la misma que la ética de un humano. No se persigue el libre desarrollo de la personalidad de la máquina, sino el del humano que interactúa con ella. Lo verdaderamente fundamental es que la máquina permita al hombre ser ético no que la máquina sea ética en sí misma.

La única forma en la que un sistema de IA puede interiorizar valoraciones éticas es mediante una **combinación de métodos deductivos** (basados en reglas) **e inductivos** (basados en la experiencia y en el falseamiento de errores), de forma similar a como aprenden los niños.

Probablemente, si logramos inculcar nuestros valores a los sistemas de IA, nos sorprenderán los resultados, pues las decisiones de tales sistemas serán probablemente mucho más éticas que las que tomamos nosotros en la vida real.

Amparo Grau comenzó su ponencia poniendo de relieve el hecho de que algunos de los problemas que plantea la robótica no son nuevos, pero el de las **relaciones entre los robots y las personas** sí.

Es innegable que la ciencia y la tecnología avanzan porque las personas quieren que se sigan desarrollando. Es necesaria una robótica socialmente responsable que contribuya al logro de los Objetivos de Desarrollo Sostenible, fijados en la Agenda 2030 de las Naciones Unidas. En el panorama actual en el que nos encontramos, de plena incertidumbre con respecto a muchos aspectos de la IA, necesitaremos que sean los legisladores y los poderes públicos quienes establezcan sistemas de gobernanza que sean inclusivos y justos. Para Amparo Grau con respecto a los robots cabe preguntarse si los sistemas de IA tienen que ser éticos o vale sólo con que lo parezcan, así como si los robots son considerados un fin o un medio.

El rasgo esencial en la conducta moral de los hombres es querer libremente lo que se hace, pero una máquina no se puede guiar por estos parámetros puesto que lleva a cabo la tarea que le ha sido encomendada por un programador a través de unos códigos de programación. Hoy en día ya existen robots que están planteando problemas y dilemas de carácter ético, tales como los *sex robots*, los robots de asistencia o los drones militares. Es importante fijar un marco seguro para la IA, pero no hay acuerdo todavía sobre cómo regularlo y hasta dónde, dejando un margen para la responsabilidad social empresarial.

A juicio de **Eduard Fosch-Villaronga**, interactuamos con los robots de formas muy distintas, entre ellas, física y cognitivamente; y así es como debería ser reflejado en los instrumentos jurídicos que gobiernan a los robots. No reconocer la **interconectividad** entre el diseño de los robots, la regulación y los humanos que interactuarán con ellas, conllevará a la insuficiencia reguladora. Esto va a ganar importancia sobre todo con el aumento de la robótica de servicio, la cual está dirigida a la población general.

Eduard Fosch-Villaronga señala que, cuando se habla de “ética y responsabilidad en el diseño de los robots”, se hace referencia al comportamiento de las máquinas, de sus diseñadores y a la interacción que las mismas van a tener con los usuarios. A ello, la Unión Europea aboga por la **creación de un código de conducta para los ingenieros**, y el establecimiento de políticas sobre la utilización de los robots desde el punto de vista de los usuarios, y así poder ayudar en la transición de una robótica centrada en lo industrial a otra de servicio.

Fosch piensa que a los principios planteados por las instituciones europeas les falta una base técnica, pues conceptos como “reversibilidad” van a carecer de sentido en robots con cierta apariencia física (*embodiment*). Por ejemplo, si una persona lleva un exoesqueleto atado al cuerpo y cae, dicha acción difícilmente va a poder ser revertida. Eduard argumenta que si las futuras políticas europeas que gobiernen sistemas ciberfísicos no tienen en cuenta la naturaleza dual de los mismos (ciber y física) y cómo afecta una parte a la otra, corren el riesgo de ser completamente inaplicables.

The future of Artificial Intelligence

Ponente: James Whittaker, Distinguido Ingeniero y Evangelista Técnico en Microsoft.

James Whittaker compartió con los asistentes a esta sesión cuál es el **futuro de la Inteligencia Artificial**.

Comenzó señalando que en nuestros días todo está regido por los datos ya que los mismos se encuentran en el corazón de todas las cosas por lo que cada día se generan **más datos** y se tiene más acceso a ellos. El cambio de paradigma se ha producido por el poder enorme que ha supuesto poder copiar y almacenar esos datos mediante el sistema **cloud** lo que ha provocado que se puedan extraer los mismos en un contexto determinado y poder relacionarlos, algo impensable no hace mucho tiempo.

Las encargadas de poder extractar y acceder a toda esa fuente de datos son las máquinas lo que supone en la práctica que poco a poco la tecnología se vaya incorporando en todos los ámbitos de nuestra vida cotidiana.

El **algoritmo es la organización de los datos** y eso es crucial para el desarrollo de cualquier Inteligencia Artificial (IA).

En base a todo lo mencionado hasta aquí, James Whittaker recomienda a los **abogados formarse en el ámbito de la ciencia de datos** y estar integrados y formar parte de los equipos que se encarguen de diseñar esas IA ya que el problema radica en conocer y entender los datos para ser más competitivo.

Killer robots y sistemas de armas autónomos en el marco del derecho internacional. Implicaciones, desafíos y riesgos

Ponentes: *Milton Meza Rivas*, Investigador en derecho internacional público y relaciones internacionales de la Universidad de Barcelona. Experto y asesor del Grupo de Trabajo sobre sistemas de armas autónomas letales de la ONU. **Carlos Espósito**, Catedrático de derecho internacional público en la Facultad de Derecho de la Universidad Autónoma de Madrid. Miembro del Consejo Académico de Fide.

Milton Meza Rivas y Carlos Espósito compartieron con los asistentes a esta sesión cuáles son las **principales implicaciones, desafíos y riesgos de los *killer robots***.

Carlos Espósito y Milton Meza Rivas decidieron que sus consideraciones versaran sobre tres ejes centrales de la discusión sobre los *killer robots*: legalidad, autonomía y responsabilidad.

Milton Meza Rivas comenzó su exposición aclarando que las palabras vertidas en la sesión no podían ser interpretadas como posiciones del Grupo de Trabajo sobre sistemas de armas autónomas letales de la ONU, sino que las mismas se encuadran en opiniones personales.

Los *killer robots* se encuadran en el hecho fundamental que, si bien la tecnología ha sido desarrollada con **fines creativos y beneficiosos** para la calidad de vida, también ésta puede ser usada con **fines destructivos y/o nocivos**, por lo que la tecnología y por tanto la Inteligencia Artificial (IA) tiene un carácter dual y ambivalente.

En el año 2010 en la Primera Resolución del Relator Especial para las Ejecuciones Extrajudiciales, Sumarias o Arbitrarias de la ONU se hablaba ya de drones, del binomio inteligencia y de robótica, así como de la autonomía de la máquina para seleccionar y/o atacar objetivos. En 2013 en la Segunda Resolución del Relator Especial para las Ejecuciones Extrajudiciales, Sumarias o Arbitrarias de la ONU también se puso el acento en las nuevas tecnologías como medios para producir ataques en contiendas militares lo que provocó que existieran reuniones informales de expertos en el seno de la CCW (Convención sobre ciertas armas convencionales) y se llegase a la conclusión de la necesidad de crear un Grupo Especial en el seno de la ONU encargado de velar por la aplicación de IA en conflictos bélicos.

Este Grupo Especial de la ONU tiene un **enfoque interdisciplinar y prospectivo** dado que el mismo está integrado por expertos en diferentes ramas de los Estados Miembros de la CCW, así como de academias especializadas.

En cuanto a la legalidad es necesario señalar que los *killer robots* tienen que cumplir con el **targeting law** es decir tienen que saber qué tipo de target es al que tienen que atacar por el principio de distinción en el marco de un conflicto armado. Como

consecuencia de esto surge la primera pregunta ¿la IA va a ser capaz de targetizar un objetivo y distinguir?

Por otro lado, el uso de la IA en un conflicto bélico tiene que cumplir también con el denominado **weapon law** el cual establece que el uso de las armas no puede suponer un sufrimiento innecesario.

En cuanto a la **ejecución del ataque** con mecanismos de IA se debe dar una precaución y previsibilidad ya que no se sabe a priori qué puede pasar con la propia máquina por lo que hay que tener todo muy controlado desde un momento inicial para que los *killer robots* cumplan con los presupuestos anteriormente señalados.

En cuanto al problema central de la legalidad de los *killer robots* hay que ser capaces de **valorar con antelación cuáles van a ser los daños colaterales** que se puedan producir, así como la necesidad de la realización de un **cálculo del riesgo** al utilizar una máquina de IA.

Las alternativas en el seno de la legalidad serían las de promulgar leyes internacionales que prohibiesen, limitasen, controlasen o permitiesen máquinas que incorporan mecanismos de IA. En este sentido, surgen problemas que deberemos solucionar basados en el **doble uso de las tecnologías** (una IA dependiendo del uso que se le dé puede ayudar al hombre o provocar destrucción) así como en el **binomio desarrollo-seguridad en el contexto de problemas con terrorismo y seguridad nacional**.

Otro punto clave en los *killer robots* es el basado en su autonomía ya que aquí nos tenemos que hacer la siguiente pregunta **¿hasta qué punto se desconecta el humano de una máquina de guerra?** Es importante el abordaje de esta cuestión ya que dependiendo de la respuesta que demos dependerá quién va a decidir sobre si debemos morir o vivir en un conflicto bélico. Por otro lado, también es importante conocer cuál es la relación entre el humano y la máquina ya que antes la misma era conceptuada como una herramienta de ayuda, hoy en día como un compañero, pero en el futuro no sabemos que evolución puede tener esa relación.

En cuanto al problema de la responsabilidad de los *killer robots* hay diferentes opiniones al respecto que pasamos a resumir:

- **Responsabilidad jerárquica:** Aquí la responsabilidad se computa en función de la cadena de mando y por tanto la rendición de cuentas por las acciones de la IA en un conflicto bélico se harían conforme a los manuales militares y estratégicos de seguridad y defensa (activador/ jefe de misiones/ comandante...)
- **Responsabilidad derivada de relación causal:** En este sentido se haría responsable de las acciones cometidas por la IA a quién la fabricó, diseñó o programó. No debemos de olvidar que en el diseño de las máquinas siempre tiene que haber un nivel significativo de intervención humana.
- **Responsabilidad objetiva:** En este tipo de responsabilidad se produciría una analogía entre los *killer robots* con la responsabilidad por el uso de desechos nucleares u objetos espaciales.

- **Responsabilidad de la máquina/sistema/tecnología:** Aquí se plantea el hecho de que sea la propia IA quién se responsabilice de sus actos. Así pues, los autores y expertos que abogan por esto plantean el *taxing robots*, la *e-personality* (Unión Europea) o la personalidad jurídica de los robots (Estonia).

Carlos Espósito hizo mención a que, aunque ahora o en un futuro se utilice la IA en guerras y conflictos bélicos “no habrá nada nuevo desde la perspectiva del derecho internacional hasta que la tecnología permita que los drones sean totalmente autónomos, es decir, que su programación le permita disparar sin esperar una orden o autorización de un operador humano.”

En cuanto a la autonomía el conferenciante quiso poner de manifiesto que es importante saber lo que implica en este contexto ya que es fundamental tratar de limitar la confusión existente sobre este concepto por lo que tenemos que ver a qué tipo de autonomía nos estamos refiriendo. Así pues, en el binomio humano/máquina hay diferentes grados de control:

- **Human in the loop:** El humano tiene el control total y absoluto de las acciones llevadas a cabo.
- **Human on the loop:** El humano tiene el control necesario para poder parar la máquina en un momento determinado.
- **Human out the loop:** El humano no puede parar el objetivo que se ha puesto la máquina.

En cuanto al problema que se genera con la legalidad de los killer robots Carlos Espósito expuso que es difícil que una IA pueda cumplir con el principio de decisión entre personas civiles y militares ya que no es sencillo saber quién forma parte de la guerra y en qué grado. También argumentó que **el problema es ético** puesto que ¿vamos a darle poder a un robot para matar si a priori no tiene capacidad empática? Esto está conectado con la determinación de la responsabilidad, ya que la atribución de responsabilidad y la reparación de los daños y perjuicios son problemas que se resuelven jurídicamente; en cambio, la cuestión relativa a si es justo delegar la decisión de matar en una máquina es un problema ético.

#Robotiuris17Fide, enero 2018.

Crónicas redactadas por la abogada Marta Pérez Cañón