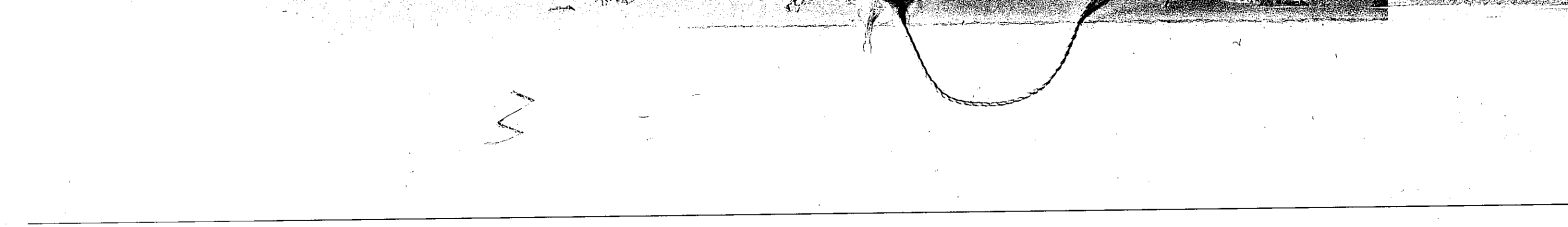
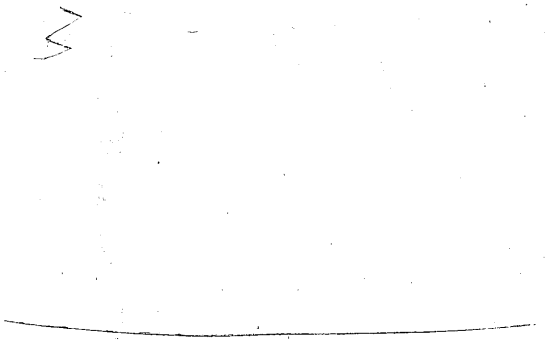
10

JOSÉ LUIS VERDEGAY GALDEANO

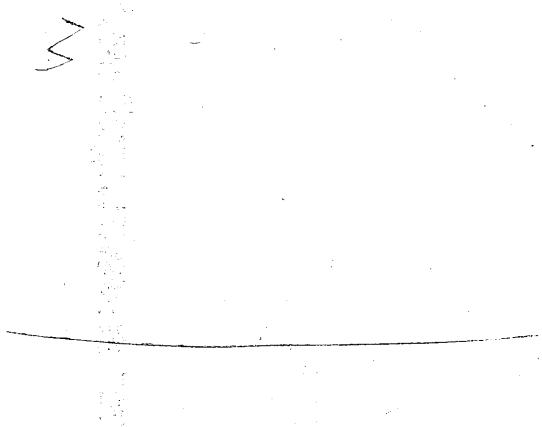
¿EN QUÉ PIENSAN LOS ALGORITMOS?

DISCURSO DE APERTURA
UNIVERSIDAD DE GRANADA
CURSO ACADÉMICO 2019 - 2020





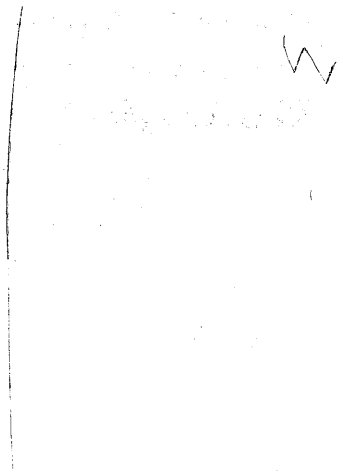
¿EN QUÉ PIENSAN LOS ALGORITMOS?



JOSÉ LUIS VERDEGAY GALDEANO
Catedrático de Ciencias de la Computación e Inteligencia Artificial (I.A.)
Universidad de Granada

¿EN QUÉ PIENSAN LOS ALGORITMOS?

DISCURSO DE APERTURA
UNIVERSIDAD DE GRANADA
CURSO ACADÉMICO 2019 - 2020



CATEDRÁTICO DE CIENCIAS DE LA COMPUTACIÓN
E INTELIGENCIA ARTIFICIAL (I.A.)
LECCIÓN INAUGURAL. APERTURA CURSO ACADÉMICO 2019-2020.

Edita: Secretaría General de la Universidad de Granada.

Imprime: Imprenta Comercial. Motril.

Depósito Legal: GR 1181-2019

Printed in Spain *Impreso en España*

Sra. Rectora Magnífica de la Universidad de Granada.
Excelentísimas e Ilustrísimas Autoridades.
Miembros de la Comunidad Universitaria.
Señoras y señores.
Amigas y amigos.

Mis primeras palabras han de ser las del agradecimiento sincero a nuestra Rectora por su invitación para intervenir en este solemne acto de inauguración del Curso Académico 2019-2020 en nuestra Universidad. Tengo que confesarles que la llamada de la Rectora, encomendándome esta tarea, supuso para mí una gran sorpresa al no ser actualmente el catedrático más antiguo de la Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación, e impartir esta lección no constaba en mi lista de débitos. Pero claro, como no podía ser de otra forma, acepté inmediatamente su ofrecimiento: entre ilusionado y emocionado, por el honor que supone tal encargo, y también en

cierto grado inseguro, diría que un punto preocupado, por la responsabilidad que conlleva asumir un compromiso de semejante calado.

Se trata, cumpliendo con un rito universitario ya secular, de ser partícipe de una tradición que muy pocas veces un profesor en la Universidad tiene el privilegio de protagonizar. Actuando para ello en el emblemático marco del Crucero Bajo de este Hospital, un sitio que impone, y no solo por el lugar y la audiencia, sino porque en el caso de nuestra Universidad, la singular ceremonia que nos convoca mantiene su esencia académica, científica y protocolaria, lo cual invita a contribuir para que así siga siendo en pro del prestigio académico de nuestra institución.

Por la simple y bien engrasada "rueda de los oficios" de la obligada rotación protocolaria entre los Centros de nuestra Universidad, se dio en el pasado la circunstancia que el actual catedrático más antiguo de la E.T.S. de Ingenierías Informática y de Telecomunicación intervino desde este mismo estrado en la entonces apertura del curso 1996-97; y en consecuencia hoy, por mero escalafón, tengo la oportunidad de poder dirigirme a todos ustedes en este acto inaugural. Lo hago desde el recuerdo agradecido y sincero hacia todas las personas que han formado y forman parte de mi vida, a quienes

constituyen mi mundo universitario y muy especialmente a mi familia.

****

Uno de los temas que tuvo más notoriedad en el mundo científico en aquel curso 96-97 fue que Deep Blue, un computador fabricado por IBM, le ganó en torneo de ajedrez al archiconocido Garri Kasparov, campeón del mundo en aquellos años. Fue como llegar a una cima de difícil conquista. Por fin parecía que se podía contestar, de forma contundente, a la clásica pregunta sobre si alguna vez las máquinas superarían a las personas. En la actualidad aquel debate no tiene mayor interés, pues conviviendo con nuestros portátiles, "tablets", teléfonos móviles, asistentes virtuales, etc., las preguntas son otras, y en palabras de Alan Turing [22]: «Sólo podemos ver un poco del futuro, pero lo suficiente para darnos cuenta de que hay mucho que hacer».

Parece que nuestro día a día lo controla la I.A., los algoritmos, y que nuestro futuro va a depender de ellos. Están presentes a diario en los medios, en nuestras conversaciones o en las redes sociales; y la verdad es que las noticias que se refieren a ellos nos preocupan y nos atemorizan hasta provocar

nuestra desconfianza. Solemos hablar de su presencia y sus consecuencias con rotundidad, pero muchas veces con gran desconocimiento, habilitando escenarios irrealizables o de difícil justificación científica, que parecen ampararse en aquellos versos del *Monte de Perfección* de San Juan de la Cruz que rezan: «Para venir a lo que no sabes, has de ir por donde no sabes» y que justifican que nos planteemos desde un punto de vista científico si los algoritmos piensan.

No hace mucho tiempo eran muy frecuentes los titulares del tipo: «Un error informático obliga a las universidades a solicitar las becas por correo ordinario» (Diario EL PAÍS, 2003); o bien «Un error informático de la DGT impide a los ciudadanos pagar sus multas durante dos días» (Periódico ABC, 2014); «Un error informático lleva a varias administraciones de la ilusión del Gordo a la decepción» (Cinco Días, 2012); «Un error informático...»... como si la culpa del fallo o del accidente la hubiera tenido el programa informático del que se tratara y no su responsable (recordemos también aquí el famoso "Efecto 2000" o el "caso Volkswagen" de manipulación del software de control de la emisión de gases contaminantes). Hoy día esos titulares casi han desaparecido, pero son constantes otros del mismo tenor, ahora hablando de los algoritmos, si bien estos últimos más que referirse

a hechos pasados, pretenden adivinar cuál será el futuro, generalmente con tintes alarmistas, previniéndonos sobre la falta de trabajo que provocarán, de las consecuencias de sus decisiones, de la irresponsabilidad de sus acciones, etc. Y es curioso que, aunque en el mundo tecnológico generalmente se asume que cualquier tiempo pasado fue peor, cuando nos referimos al futuro que nos espera de la mano de la I.A., de los algoritmos, el panorama se torna incierto, sombrío, y en definitiva pesimista hasta el punto de preferir el presente, aunque ello suponga la incoherencia de tener que defender que cualquier tiempo pasado fue mejor.

La verdad, la propia palabra "algoritmo" puede resultar inquietante. ¿A quién se le ocurrió semejante nombre? El verdadero origen de la palabra algoritmo (o la llamada *algorismo* durante mucho tiempo) no se conoció hasta el siglo XIX, aunque desde tiempos de Leibniz era de uso corriente. Se pensaba en una degeneración del término logaritmo, y se daba por seguro que estaba relacionada con la palabra griega "arithmos". Fueron los historiadores quienes explicaron el acertijo. En efecto, uno de los más grandes matemáticos árabes del siglo IX de nuestra era, Abu 'Abd Allah Muhhamad ibn Musa al-Khwarizmi (literalmente Padre de Abdullah, Mohamed, hijo de Moisés, nativo de Khwarizm, hoy Khiva en Uzbekistán) con su

Compendio de cálculo por reintegración y comparación (en árabe "Kitab al-jabr wa'l-muqabala"), ayudó a difundir las matemáticas árabes por el mundo occidental, hasta tal grado que del título de su obra se ha desprendido el término álgebra (al-jabr). Con el paso del tiempo, por defectos de pronunciación, su nombre se difundió simplemente como Al-Juarismi y de éste surgieron los términos guarismo y algorismo (acuñados para referirse a cualquier método de cómputo usando la notación arábiga de numeración). A su vez, el término algorismo también fue corrompido en su pronunciación hasta derivar en uno más difundido en latín como algoritmus, que fue empleado desde el siglo XVII por los matemáticos para referirse a procedimientos de cálculo, a raíz de la traducción al señalado latín de otro de los libros de nuestro protagonista y que se tituló *Algoritmi de Numero Indorum*. Finalmente, la palabra que conocemos no apareció en un diccionario hasta la edición en el año de 1957 del *Webster´s New World Dictionary*.

Aclarado pues ese origen, de modo más formal podemos definir al concepto de algoritmo que nos interesa como una secuencia ordenada de pasos, exentos de ambigüedad, que al llevarse a cabo con fidelidad dará como resultado que se realice la tarea para la que se ha diseñado (se obtenga la solución del problema planteado) en un tiempo finito.

Desde esta perspectiva, para resolver un problema con un ordenador es necesario, en principio, diseñar un algoritmo que describa la forma en que debe efectuarse el proceso de hallar la solución y posteriormente expresar cada uno de esos pasos de la forma adecuada para que la máquina lo pueda llevar a cabo. Esto es, se deberá expresar el algoritmo como un programa en un lenguaje de programación y, por último, lograr que el ordenador ejecute el programa correctamente. El algoritmo es, por tanto, un concepto central de la Ciencia de la Computación y naturalmente de la I.A., que no tiene sentido sin ellos.

La actual controversia sobre los algoritmos ni es nueva, ni menos aún superficial, bien al contrario es de una gran profundidad teórica en el área de la Ciencia de la Computación. El objetivo de desarrollar algoritmos para resolver problemas, ha sido considerado siempre por matemáticos e informáticos especialmente importante. A través de los siglos, los intentos por desarrollar algoritmos capaces de resolver cualquier problema no han tenido éxito. Debido principalmente a Hilbert, durante muchos años se creyó que si cualquier problema podía plantearse de manera precisa, entonces, con suficiente esfuerzo, sería posible encontrar una solución con el tiempo (o tal vez podría proporcionarse en el transcurso del propio tiempo una

prueba de que no existía solución). En otras palabras, se creía que no había problema que fuera tan intrínsecamente difícil que en principio nunca pudiera resolverse. El problema no se resolvió hasta setenta años después, pero en sentido negativo: en 1970 Yuri Matiyasévich culminó más de veinte años de trabajo de varios matemáticos, con la demostración de la inexistencia de un algoritmo del tipo que buscaba Hilbert.

A partir de la formalización del concepto de algoritmo, se planteó la cuestión de clasificar los problemas según que siempre se pueda encontrar la solución por medio de un algoritmo (problemas computables) o que no se pueda asegurar que existan algoritmos que siempre produzcan una solución (problemas no computables). Al principio de la década de los 50, la necesidad de resolver problemas prácticos, en su mayoría problemas computables de tamaño cada vez mayor, junto con el comienzo del desarrollo de los computadores, y la posibilidad de utilizar estos para la resolución de los problemas, cambió el interés principal, el del estudio de la computabilidad o no computabilidad de los problemas, por el análisis de la complejidad de los problemas computables, es decir al estudio de cómo encontrar algoritmos que no solo hallen una solución, sino que la encuentren en el mínimo número de operaciones y utilizando el mínimo

espacio posible; en otras palabras, buscar algoritmos más eficaces. Posteriormente se clasificaron los problemas computables en dos tipos: aquellos para los que existe un algoritmo de complejidad polinómica (en el tamaño de la entrada del problema) y aquellos problemas para los que parece que los mejores algoritmos tienen una complejidad exponencial. Al principio de los años 70 del pasado siglo dichos tipos de problemas cristalizaron en las clases P y NP y el problema abierto sobre estas clases, por cuya solución el "Clay Mathematics Institute" (Cambridge, Massachusetts, EE.UU.) ofrece un millón de dólares, consiste en saber si la clase P es igual a la NP, es decir, si todo problema se puede resolver en tiempo polinómico.

El sentido moderno de algoritmo es algo bastante similar a receta, proceso, método, técnica, procedimiento, rutina, con la excepción de que la misma palabra algoritmo connota algo un poco diferente de todo eso. Además, de ser un conjunto finito de reglas que constituyen una secuencia de operaciones encaminadas a resolver un tipo específico de problema, un algoritmo tiene cinco características primordiales.

a) Finitud, de modo que un algoritmo debe terminar siempre tras un número finito de etapas.

b) Especificidad, en el sentido que cada etapa debe estar precisamente definida; las acciones que hay que llevar a cabo deben estar rigurosamente especificadas para cada caso.

c) El input, es decir los valores que se le dan inicialmente antes de que el algoritmo comience. Estos inputs se toman de conjuntos de objetos pre-especificados.

d) El output, es decir, resultados que son cantidades específicamente relacionadas con los inputs.

e) Efectividad. Se espera generalmente que un algoritmo sea efectivo, lo que significa que todas las operaciones que hay que realizar en el algoritmo deben ser lo suficientemente básicas como para que, en principio, se hagan exactamente y en un periodo finito de tiempo por una persona que solo use lápiz y papel.

Los algoritmos no son por tanto como las recetas, que pueden tener reglas imprecisas. Son [6] procesos iterativos que generan una sucesión de puntos, conforme a un conjunto dado de instrucciones y un criterio de parada, y como tales no están sujetos a restricciones tecnológicas de tipo alguno, es decir, son absolutamente independien-

tes del equipamiento tecnológico disponible para resolver el problema que afronten. Es el programa en el que se escriba el algoritmo el que depende de la tecnología disponible, es decir, un algoritmo que pueda resolver un problema complicadísimo no sirve de nada si no puede ejecutarse, si no puede desarrollar sus cálculos, en una máquina apropiada, lo que como se ha dicho es misión del programa en el que se codifique el algoritmo y del computador que se use.

En muchos casos esto ha sido motivo de una importante ralentización del progreso científico en diferentes ámbitos del saber, como puede ser el caso de las Heurísticas, de la Programación Dinámica, de la Prospectiva Meteorológica o del Análisis Exploratorio de Datos. Conocíamos los algoritmos que podían dar respuesta a problemas trascendentales para la sociedad (reconocimiento de secuencias de ADN, itinerarios óptimos, alarmas de catástrofes, reglas de conducta, etc.) pero no podíamos programarlos y ejecutarlos porque la tecnología disponible no lo permitía, es decir, los computadores disponibles no podían calcular las soluciones buscadas en tiempos razonables.

Sin embargo, el abaratamiento de los computadores, el aumento de su velocidad y el tremendo incremento de sus capacidades, que se produce de

forma incesante desde finales del siglo veinte, ha tenido dos consecuencias importantes. La primera el hecho de que hoy día no quede prácticamente ninguna parcela que tenga un problema que no se pueda abordar con un computador. Este es el caso precisamente de la I.A., que durante años vivió al amparo de la ciencia ficción, seguramente proyectando una imagen tan falsa como ridícula, y en la actualidad es omnipresente en cualquier actividad real de nuestro día a día. La segunda consecuencia es que estemos siendo testigos de la cuarta revolución industrial. En efecto, si la tercera revolución fue la conocida revolución digital que se desarrolló desde mediados del siglo XX, vinculada después con el término Sociedad de la Información y amparada por las nuevas Tecnologías de la Información y la Comunicación (TIC) y las Energías Renovables, esta cuarta revolución que estamos viviendo está marcada por avances tecnológicos emergentes en una serie de campos (Robótica, Computación Cuántica, Biotecnología, Internet de las Cosas, etc.) de entre los que sobresale la I.A., los algoritmos, que seguramente por desconocimiento y por la mala información que se da sobre ella, causa preocupación en la Sociedad, cuando sin embargo su fin último es conseguir una sociedad mejor para todos.

La I.A. juega un papel clave en esa cuarta revolución industrial debido principalmente a cinco características, que de manera singular le confieren una naturaleza especial: la transparencia, porque no solemos detectarla cuando interactuamos con algún sistema dotado de I.A.; su dificultad, porque la referencia esencial para su trabajo la constituyen ni más ni menos que las personas humanas; su adaptabilidad, puesto que es contexto dependiente y por tanto puede resultar polimórfica; la transversalidad, porque en la actualidad no hay ámbito que esté al margen de sus aplicaciones; y por último su necesaria y permanente renovación y mejora, como es inherente al área de las TIC, que conducen a una necesidad de inmediatez en las respuestas como en ningún otro campo.

Lo que pasa con los coches autónomos, es decir, con los automóviles sin conductor, es un buen ejemplo de lo sensible que está la sociedad con la I.A. Se plantea como imperioso, antes de proseguir con la fabricación de automóviles autónomos, que estos estén preparados para reaccionar correctamente ante situaciones que implican atropellos a peatones en diferentes circunstancias y otros tipos de accidentes. Siendo razonable el punto de partida, no es de recibo la exigencia porque se trata de un problema bien conocido, que no tiene una solución universal. En efecto, el problema en cues-

tión es el conocido Dilema del Tranvía, formulado por la filósofa Philippa Foot en un artículo de 1967 [13], del siguiente modo: Se supone que observamos un tranvía que se dirige, fuera de control y sin frenos, hacia cinco personas que están trabajando en la vía. No podemos avisarles y tampoco podemos parar el tren. Pero si podemos accionar una palanca que desviará el vagón hacia otra vía. Lo que pasa es que en esa otra vía hay otra persona. ¿Deberíamos mover la palanca? La discusión ética sobre como resolver este problema está abierta desde 1967, y probablemente así seguirá durante muchos años. Sin embargo hoy día se exige que los coches autónomos sepan reaccionar ante situaciones parecidas a estas, y que además reaccionen como cada uno de nosotros reaccionaría.

De este modo, planteando a la sociedad situaciones como esta o parecidas en otros contextos, como pueden ser las sentencias penales, las intervenciones quirúrgicas, la fabricación de armas, etc., es razonable que aparezcan dudas sobre cuál será el futuro que vendrá de la mano de la I.A. Pero es que además, después de haber estado muchos años, seguramente desde el nacimiento de la I.A. en la Conferencia de Dartmouth (1956), exclusivamente concentrados en responder a la pregunta de si las máquinas superarían a las personas, de pronto los avances que se producen casi a diario

nos resultan inquietantes al dejarnos entrever conductas autónomas, desde luego basadas en algoritmos, que por ser quizás impredecibles, pueden condicionar nuestra forma de vida.

Esa inquietud está más que justificada por la creciente aceleración con la que surgen en los últimos años nuevos resultados que refuerzan la autonomía de los Sistemas basados en I.A.:

1994 Dos vehículos autónomos (Mercedes 500 SEL) recorren 1.000 km de autopista en Paris.

1996 Nacen lo que se denominan "agentes" inteligentes, que perciben el entorno.

1997 Deep Blue gana a Kasparov al ajedrez.

2004 Vehículos autónomos, sin intervención humana alguna, recorren en competición 100 km en el desierto de El Mojave.

2008 Google lanza la primera App que reconoce la voz.

2011 Watson (IBM) gana a los concursantes humanos en el concurso estadounidense de televisión Jeopardy. ¡Todo se acelera!

2013 Boston Dynamics construye Atlas, un robot humanoide bípedo destinado a ayudar a los servicios de emergencia en

las operaciones de búsqueda y rescate, realizando tareas como cerrar válvulas, abrir puertas y operar equipos motorizados en entornos donde los humanos no podrían sobrevivir.

2014 El programa Eugene Goostman pasa el Test de Turing. Se trata de un programa diseñado para mantener conversaciones, que se hace pasar por un adolescente ucraniano de 13 años. En la prueba que se organizó en la Royal Society de Londres, consiguió engañar a un 33 por ciento de los 150 jueces humanos que tenían que decidir si conversaban con una persona o no, haciéndoles creer que en efecto era un chaval de 13 años.

2015 Los avances en reconocimiento de imágenes son incuestionables, pudiendo decir sin pudor que las máquinas "ven" mejor que los humanos.

2016 AlphaGo, desarrollado por Google DeepMind, gana a Lee Sedol, campeón mundial de go El Massachusetts Institute of Technology (MIT) pone en circulación comercial el primer taxi autónomo en Singapur.

2017 El programa Libratus (creado por Tuomas Sandholm y Noam Brown en la Universidad Carnegie Mellon, EE.UU.) vence al póker contra cuatro personas.

DeepMind se lleva lo aprendido en un juego a otro.

El ginoide Sophia (desarrollado por la compañía Hanson Robótics con sede en Hong Kong) se convierte en ciudadana Saudí.

En los Emiratos Árabes Unidos, se nombra a Omar bin Sultan Al Olama ministro de I.A.

2018 AlphaZero (una versión más generalizada de AlphaGo Zero, que aprendió por sí mismo a jugar al Go) aprende también por sí mismo a jugar al ajedrez.

Atlas aprende a hacer parkour (un deporte de origen francés, muy físico, en el que las personas que lo practican utilizan su capacidad motriz para superar los obstáculos urbanos que se encuentran a su paso realizando acrobacias).

Los profesionales del estado de Kerala (India) solicitan al Gobierno de Nueva Delhi la creación de un Ministerio de I.A.

Un robot humanoide llamado Michihito Matsuda es candidato a alcalde en el distrito de Tama en Tokio. Lo notable es que prometiendo justicia para todos y el fin de la corrupción, logró quedar en tercer lugar a escasos 400 votos del segundo candidato.

2019 Según la consultora Gartner [16] las 10 tecnologías estratégicas más disruptoras para los próximos cinco años son: los dispositivos autónomos, las analíticas aumentadas, los desarrollos reforzados con I.A., los gemelos digitales, *el edge computing* potenciado, las tecnologías inmersivas, el *blockchain*, los espacios inteligentes, la computación cuántica y la privacidad y la ética digital, que va a jugar un papel trascendental.

Con toda seguridad este curso asistiremos a nuevos y espectaculares avances que, con independencia del carácter de sus contribuciones, deberían servir para empezar a dar solución a los diferentes desafíos que se nos plantean en prácticamente todos los escenarios posibles:

- el laboral, para garantizar desde las instancias públicas el empleo sostenible, estable y de calidad para las personas, porque cam-

biará la estructura social y el mercado de trabajo;

- el económico, justificado por el continuo crecimiento del mercado (desde robots y coches sin conductor con un gran nivel de sofisticación, hasta un amplio abanico de técnicas "ocultas" que los utilizan). Nótese que según un informe de la empresa de estudios Tractica [4], es probable que ese volumen de negocios continúe creciendo desde los 5.400 millones de dólares netos de beneficios de 2017, hasta los 105.800 millones en 2025, con cálculos más que conservadores. De ahí que el propio sector empresarial tenga que acometer el desarrollo de iniciativas, fundamental pero no exclusivamente, formativas encaminadas a lograr una transformación acorde a las nuevas tendencias y ajustada a la realidad de su entorno geográfico y competitivo;
- el educativo, ya que para minimizar los efectos negativos de esta cuarta revolución será necesario fomentar y potenciar las actividades formativas en todos los niveles, favoreciendo por un lado que las personas sean conscientes de la importancia que tienen sus datos y la seguridad de los mismos,

y por otro ofreciendo líneas formativas y de investigación prometedoras, que no queden obsoletas en poco tiempo. En este sentido interesa señalar que en China se ha puesto en marcha un programa para formar en I.A. a sus estudiantes desde preescolar (han publicado una colección de 33 libros de texto que, partiendo de la etapa de la guardería, cubren la totalidad de la educación obligatoria).

el social, porque hay que establecer las reglas de comportamiento que habrá que exigir a los sistemas basados en I.A. para que no resulten perjudiciales para la sociedad.

Las sospechas y prevenciones que muestra la sociedad ante los sistemas basados en I.A. están fundamentadas, porque está comprobado que cuando estos sistemas se gestionan con técnicas de automatización inteligente pueden aumentar, y en algunos casos sustituir por medio de programas completamente autónomos, la capacidad de actuar y de tomar decisiones de las personas humanas. En lo que sigue, para abreviar, a estos sistemas los denominaremos aquí Sistemas Autónomos de Decisión (SAD).

El hecho crucial como hemos dicho es que esa "sustitución" de funciones podría producir, más

pronto que tarde, cambios estructurales en la sociedad en su conjunto; pérdidas masivas de puestos de trabajo y por tanto la minusvaloración y descalificación de las personas que los desempeñaban; efectos desconocidos en los sistemas que puedan gestionar o situaciones de ingobernabilidad no deseadas. Por todo ello resulta patente que:

a) las cuestiones éticas relativas al comportamiento de los SAD deben incluirse en su diseño tecnológico, para facilitar que más que como amenazas, se entiendan como promotores de la innovación, y

b) en caso de producirse esa sustitución de funciones, que no produzca disfunciones, es decir, que el correspondiente sistema actúe exactamente igual que lo haría el decisor humano de turno, reproduciendo y mejorando su conducta y procurando evitar los ineludibles e imprevisibles fallos que las personas podemos tener a la hora de tomar decisiones, sobre todo cuando estas han de tomarse en ambientes desconocidos [18].

Conscientes de estas dos exigencias sobre el carácter ético y el correcto funcionamiento de los SAD, múltiples organizaciones de todo el mundo han comenzado a debatir las condiciones bajo las

que deberían desempeñarse dichos sistemas, así como las premisas que deberían guiar su diseño, construcción y ubicación material. Las agrupamos y describimos a continuación considerando tres perspectivas diferentes: la profesional y socio-política europeas y la científica.

Así, a finales del año pasado, Informatics Europe, una organización privada que representa ante la Unión Europea a la comunidad investigadora y académica en Informática de Europa, el *ACM Europe Council,* que persigue incrementar en Europa el nivel y visibilidad de las actividades de la *Association for Computing Machinery* y el *ACM Europe Policy Committee* publicaron un informe que recoge las medidas que se habrán de tener en cuenta para lograr un desarrollo equilibrado y eficaz de los SAD en la Sociedad [17].

Dado que en la práctica sería peligroso responsabilizar de las cuestiones éticas relacionadas con estos sistemas a comités de expertos o a la industria implicada en cada caso de uso, ya que principalmente lo que se requiere es una profunda comprensión e incorporación de la ética en todo el diseño de la tecnología, los valores sociales y morales no deberían ser vistos como simples "factores de riesgo" o restricciones, sino como los principales impulsores y moldeadores de la

innovación, y por tanto deberían estar incorporados en los SAD desde el primer momento de su concepción.

Para ello se hacen una serie de recomendaciones de tipo técnico, ético, legal, social, económico y educativo, que de forma resumida exponemos a continuación.

a) Recomendaciones de tipo técnico

1. Establecer medios, medidas y estándares para asegurar que los SAD sean objetivos. Todos los actores clave (instituciones gubernamentales, academia, industria, instituciones internacionales, O.N.G. y ciudadanos) deben estar involucrados en la formulación de normas y prácticas que aseguren el bien público como primer criterio que ha de conducir el diseño y la construcción de los SAD. Estas normas deben estar formuladas de manera flexible para que perduren ante la rápida evolución de la tecnología y de las aplicaciones industriales de los SAD. Para facilitar este objetivo, es necesario incentivar la investigación en I.A. de cara a desarrollar una base teórica sólida sobre la Toma de Decisiones automatizada.

b) Recomendaciones de tipo ético

2. Asegurar que la ética se mantenga a la vanguardia del desarrollo y la implementación de SAD y que sea parte integral de esta. Al igual que con la Salud y la Biología, los países miembros y la UE deberían desarrollar comités de ética para asesorar a las organizaciones sociales, políticas, académicas y legales sobre las consecuencias positivas y negativas de las iniciativas, herramientas y sistemas relativos a los SAD. También, como garante del interés público, debería crearse una nueva Agencia europea que supervisara el desarrollo y despliegue de los SAD en toda Europa.

3. Promover el diseño de SAD sensibles a valores. En educación superior, deberían diseñarse programas especiales sobre técnicas sensibles a los valores, resaltando que los valores sociales y las prioridades éticas de los usuarios deben tenerse en cuenta en todos los aspectos y elementos asociados a un SAD.

c) Recomendaciones de tipo legal

4. Definir claramente la responsabilidad legal del uso e impacto de los SAD. Los principios básicos que gobiernan actualmente el desarrollo de SAD desde el punto de vista

informático-profesional, deben ser la base de un amplio debate entre expertos legales y técnicos, medios de comunicación y sociedad en busca de nuevas normas legales para gobernar la implementación masiva de SAD. En particular, hay que volver a considerar el descargo general de responsabilidades que tiene casi todo el software actual, y revisarse o rechazarse si, como parece, no es aplicable a muchos de los actuales o futuros usos de los SAD. La Agencia europea propuesta en la Recomendación 2 debería fomentar y facilitar este debate, así como proponer una legislación adecuada.

d) Recomendaciones de tipo económico

5. Garantizar que las consecuencias económicas de la adopción de SAD sean siempre consideradas en su totalidad. Entre sus primeras iniciativas oficiales, con el fin de emitir directrices y reglamentos apropiados, la nueva Agencia propuesta anteriormente debería comenzar emitiendo informes sobre una serie de problemas concretos de índole económica y socioeconómica, a los que el desarrollo y acelerada aplicación de los SAD probablemente den lugar. Se debe reconocer explícitamente que la

misión de la referida Agencia siempre se orientará a dos objetivos inherentemente interrelacionados: fomentar la evolución y el uso responsable de los SAD y minimizar sus posibles disrupciones de tipo personal, social y económico sobre los individuos y las naciones.

e) Recomendaciones de tipo social

6. Imponer legalmente que se informe claramente a los usuarios de los SAD sobre todas las prácticas de privacidad y adquisición de datos de sus implementadores. El aprendizaje automático funciona a partir de datos y por tanto cuando y donde se recopile la información, lo que se recopile y los usos que se le dará, deben ser descritos obligatoriamente por el proveedor de los datos de manera concisa y clara.

7. Aumentar significativamente la financiación pública para la investigación no comercial relacionada con SAD. Es necesario incentivar la investigación dirigida a comprender mejor el aprendizaje automático y su uso en sistemas que puedan influir en el comportamiento humano. Quedan por investigar muchas cuestiones fundamentales, pero el conocimiento público y riguroso de

los resultados logrados por estas técnicas, no exclusivamente dependientes de la industria, ha de ser un requisito previo para un posterior debate sobre su aceptabilidad y adopción efectiva por parte de las empresas europeas.

f) Recomendaciones de tipo educativo

8. Fomentar la formación técnica universitaria relacionada con los SAD. Todos los estudiantes universitarios deberían recibir formación sobre los aspectos prácticos y el potencial del aprendizaje automático. Los estudiantes de todas las disciplinas deben ser conscientes del impacto que esta tecnología tendrá en su campo y en su futuro trabajo.

9. Complementar la formación técnica con formación socio humanista del mismo nivel. Debido al impacto cada vez mayor que la tecnología tendrá en la sociedad, los currículos técnicos también deberían formar a los estudiantes para enfrentar escenarios complejos al complementar las habilidades técnicas con el desarrollo del pensamiento crítico, la formación digital y el juicio ético. Los planes de estudio de educación superior deberían fomentar los estudios

interdisciplinares, basados en el patrimonio cultural europeo, tanto en las disciplinas científicas como en las artes liberales. También debería incluirse en los planes de estudio de secundaria una introducción accesible a los SAD y a los problemas que plantean.

10. Aumentar la conciencia y comprensión del público sobre los SAD y sus impactos. Existe una clara necesidad de formar a la sociedad sobre esta tecnología, ya que se está introduciendo rápidamente y nos afectará prácticamente a todos en nuestras vidas profesionales y privadas. Como la mayoría de las personas no siguen cursos adicionales después de completar su formación, los medios de comunicación públicos representan los medios "de facto" más adecuados para formar a la población en general. En consecuencia, los profesionales informáticos y los responsables de las políticas tecnológicas deben actuar coordinadamente con la prensa para transmitir la información a la que se refieren las recomendaciones recogidas aquí. Se debe prestar la debida atención al uso preocupante de las técnicas de I.A. para influir en la opinión pública.

Por su lado, la UE preocupada también por este tema y convencida de que los SAD pueden aportar prosperidad, contribuir al bienestar y ayudar a mantener en un buen nivel los objetivos morales y socioeconómicos europeos si esas disciplinas se diseñan y despliegan con inteligencia, ha emitido una Declaración a través del "European Group on Ethics in Science and New Technologies", exigiendo la apertura de un proceso que allane el camino hacia un marco ético y legal común, reconocido internacionalmente, para el diseño, producción, uso y gobierno de los SAD. También propone un conjunto de principios básicos, sobre los valores establecidos en los Tratados y Carta de Derechos Fundamentales de la Unión, cuya versión íntegra puede consultarse en [11].

a) Principio de dignidad humana, entendido como el reconocimiento de que el estado humano inherente a ser digno de respeto, no debe ser violado por las tecnologías autónomas o automáticas. Esto supone poner límites a las determinaciones y clasificaciones relativas a las personas, hechas sobre la base de algoritmos y SAD, especialmente cuando los afectados no están informados sobre ello. También implicaría que tiene que haber límites (legales) a las formas en que se puede hacer creer a las personas que

están tratando con seres humanos mientras que, de hecho, están tratando con algoritmos y máquinas. Una concepción relacional de la dignidad humana que se caracterice por nuestras relaciones sociales, requiere que seamos conscientes de si interactuamos con una máquina o con otro ser humano, a fin de poder asignar unas tareas al ser humano y otras a la máquina.

b) Principio de autonomía, que afecta a la libertad del ser humano y se traduce en la responsabilidad humana sobre los SAD, y por lo tanto de su control y conocimiento, ya que dichos sistemas no deben impedir la libertad de los seres humanos para establecer sus propios estándares y normas de vida. Todas las tecnologías autónomas o automáticas deben por lo tanto respetar la capacidad humana para elegir si, cuándo y cómo delegar decisiones y acciones en ellas. Esto también implica la transparencia y la previsibilidad de los SAD, sin las que los usuarios podrían renunciar a su uso si lo consideraran moralmente necesario.

c) Principio de responsabilidad, fundamental para la investigación y aplicación de los SAD, que solo deberían desarrollarse y uti-

lizarse de modo que sirvan al bien social y ambiental común, según lo que democráticamente se haya acordado. Esto implica que deben diseñarse de modo que sus efectos se alineen con los valores y derechos humanos fundamentales. Dado que el posible uso indebido de estos sistemas plantea un gran desafío, hay que tomar conciencia de ello y por tanto sus aplicaciones ni deben suponer riesgos inaceptables de daño para los seres humanos, ni deben comprometer la libertad y la autonomía humanas al reducir ilegítimamente y de forma subrepticia las opciones y el conocimiento de los ciudadanos. En cambio deberían estar orientados a su desarrollo y uso para incrementar el acceso al conocimiento y aumentar las oportunidades que puedan tener las personas. La investigación, diseño y desarrollo de SAD deben guiarse por una auténtica preocupación por la ética de la investigación, la responsabilidad social de los desarrolladores y la cooperación académica global, para proteger los derechos y valores fundamentales y apuntar a diseñar tecnologías que los respalden y no les resten valor.

d) Principio de justicia, igualdad y solidaridad, según el cual los SAD deben contri-

buir a la justicia global y al acceso equitativo a los beneficios y ventajas que puedan aportar. Los sesgos discriminatorios en los conjuntos de datos que se utilizan para entrenar y aplicar estos sistemas deben poder prevenirse, detectarse, alertarse y neutralizarse cuanto antes sea posible.

Necesitamos hacer un esfuerzo global para lograr la igualdad de acceso a los SAD, así como la distribución justa de los beneficios y la igualdad de oportunidades en la sociedad. Esto incluye la formulación de nuevos modelos de distribución equitativa y la participación en los beneficios de los que se pueda disponer para responder a las transformaciones económicas causadas por la automatización, la digitalización y los SAD en sí mismos, asegurar el acceso a las principales tecnologías basadas en I.A. y facilitar y potenciar la capacitación en Ciencias, Tecnologías, Ingenierías y Matemáticas (STEM) y otras disciplinas digitales, en particular en las regiones y grupos sociales más desfavorecidos. Así mismo hay que cuidar las acumulaciones masivas de datos de los individuos, para tratar de evitar presiones que limiten la solidaridad, por ejemplo en sistemas de asistencia mutua como los seguros sociales y sanitarios, ya que estos procesos pueden socavar la cohesión social y dar lugar a individualismos radicales.

e) Principio de democracia, estableciendo que las decisiones clave para regular el desarrollo y la aplicación de los SAD deben ser el resultado de un debate democrático con un compromiso público. La cooperación global y el diálogo público garantizarán que esas decisiones se tomen de manera inclusiva, informada y con visión de futuro. El derecho a recibir educación y a acceder a la información sobre las nuevas tecnologías y sus implicaciones éticas facilitarán que todos comprendamos los riesgos y las oportunidades y estemos facultados para participar en los procesos de decisión que determinan de manera crucial nuestro futuro.

Los Principios de dignidad y autonomía humana involucran centralmente el derecho de las personas a la autodeterminación por medios democráticos. Los principios de pluralidad de valores, de diversidad de pensamiento y de libertad sobre la forma de vida de los ciudadanos son de crucial importancia para nuestros sistemas sociales, por lo que las nuevas tecnologías nunca deben ponerlos en peligro, subvertirlos o equipararlos a otros similares proporcionados por aquellas tecnologías que influyan en la toma de decisiones políticas e infrinjan la libertad de expresión y el derecho a

recibir e impartir información sin interferencias. Las tecnologías digitales deberían utilizarse más bien para aprovechar la inteligencia y el apoyo colectivos y mejorar los procesos cívicos de los que dependen las sociedades democráticas.

f) Principio de estado de derecho y rendición de cuentas, concretando que el estado de derecho, el acceso a la justicia y el derecho a la reparación y a un juicio justo proporcionan el marco necesario para garantizar el cumplimiento de los derechos humanos y los posibles reglamentos específicos de los SAD. Esto incluye la protección contra los riesgos derivados de esos sistemas que pudieran infringir los derechos humanos, como la seguridad y la privacidad. La gama completa de desafíos legales que surgen en este campo debe abordarse con una inversión oportuna en el desarrollo de soluciones sólidas que proporcionen una asignación justa y clara de responsabilidades y mecanismos eficientes de leyes vinculantes.

En este sentido, los gobiernos y las organizaciones internacionales deberían aumentar sus esfuerzos para aclarar en quién recaen las responsabilidades por los daños causados por el comportamiento

no deseado de los SAD. Además, deberían existir sistemas eficaces para minimizar los de daños.

g) Principio de seguridad, protección e integridad física y mental, que establece que la seguridad y la protección de los SAD se materializan en tres formas:

- garantizando la seguridad externa para su entorno y usuarios,
- aportando confiabilidad y solidez interna (por ejemplo contra la piratería) y
- proporcionando seguridad emocional con respecto a la interacción hombre-máquina.

Los desarrolladores de SAD deben tener en cuenta todas las dimensiones de la seguridad y, antes de cualquier puesta en servicio, superar las suficientes pruebas como para garantizar que los sistemas no infrinjan el derecho humano a la integridad física y mental y a disponer de un entorno seguro. En todo caso se debe prestar especial atención a las personas que se encuentran en una posición vulnerable, así como al potencial uso dual armamentista de la I.A., como por ejemplo puede pasar en entornos de ciberseguridad, finanzas, infraestructuras y conflictos armados.

h) Principio de protección de datos y privacidad, que tiene un valor fundamental porque en una época de captura masiva y generalizada de datos, el derecho a la protección de la información personal y el derecho al respeto a la privacidad son un desafío crucial. Tanto los robots físicos con I.A. como parte de la Internet de las Cosas, como los "softbots" de I.A. que operan a través de la web, deben cumplir con las leyes sobre protección de datos y no recopilarlos, ni difundirlos, ni ejecutarlos sobre conjuntos de datos para cuyo uso y difusión no se haya dado consentimiento. Los SAD no deben interferir con el derecho a la vida privada, que comprende el derecho a estar libre de tecnologías que influyan en el desarrollo y las opiniones personales, el derecho a establecer y desarrollar relaciones con otros seres humanos y el derecho a estar libres de vigilancia. También en este sentido, se deben definir criterios exactos y establecer mecanismos que aseguren el desarrollo ético y la aplicación ética de estos sistemas.

A la vista de las actuales preocupaciones sobre las implicaciones de los SAD en la vida privada y la privacidad, parece oportuno abrir el debate sobre la conveniencia de introducir nuevos dere-

chos, como son el derecho a un contacto humano significativo y el derecho a no usar los perfiles personales ni a ser medido, analizado, entrenado o persuadido.

i) Principio de sostenibilidad, según el cual la tecnología asociada a los SAD debe estar en línea con la responsabilidad humana de garantizar las condiciones previas básicas para la vida en nuestro planeta, continuar prosperando para la humanidad y preservar un buen ambiente para las generaciones futuras. Las estrategias para evitar que las tecnologías futuras afecten de manera perjudicial a la vida y la naturaleza humana se deben basar en políticas que aseguren la prioridad de la protección del medio ambiente y la sostenibilidad.

El "Future of Life Institute" es una institución que tiene como misión catalizar y apoyar la investigación y las iniciativas encaminadas a salvaguardar la vida, así como favorecer corrientes de opinión optimistas sobre el futuro que faciliten el que la humanidad siga su propio curso con la integración en el mismo de las nuevas tecnologías y los desafíos que conllevan. Entre los temas de su interés viene priorizando el impulso y estudio de proyectos que fortalezcan los efectos beneficiosos

de la I.A. En este contexto en enero de 2017 se celebró en California (EE.UU.) la "Asilomar Conference on Beneficial A.I.". La conferencia reunió docenas de expertos en Robótica, Física, Economía, Filosofía, I.A. y otras áreas que mantuvieron debates sobre la seguridad de la I.A., el impacto económico en los trabajadores humanos y sobre la ética de la programación, entre otros aspectos.

Como fruto de las discusiones que allí se celebraron, surgieron los denominados "Principios de I.A. de Asilomar" [5]. La lista, consensuada por el 90% de los expertos convocados, y en la que hasta ahora han participado más de 1273 investigadores en I.A. y otros 2541 expertos en distintos temas (entre ellos Elon Musk, CEO de Tesla, o Stephen Hawking) está constituida por 23 principios referidos a distintos aspectos, que van desde estrategias de investigación hasta derechos sobre los datos y otros problemas futuros, incluidas las posibles superinteligencias.

Aunque la lista de principios no se considera cerrada o acabada, si refleja una serie de puntos que los expertos consideran que han de tenerse en cuenta, dado que en la actualidad hay una especie de "conducta predeterminada" en torno a muchos temas relevantes, que bien podría violar los principios que la mayoría de los participantes acordaron

que era importante mantener. Por otro lado, aunque algunos principios tienen un menor respaldo que otros, como es el caso de los de transparencia o investigación compartida por compañías competidoras, es indudable que el respeto por el cumplimiento del conjunto de los 23 principios en todo lo que se pueda, producirá importantes mejoras en los aspectos que más preocupan de la aplicación de los SAD.

A continuación recogemos los 23 Principios consensuados sobre Investigación, Ética y Valores y los contemplados a largo plazo:

a) Principios relativos a la investigación

1. De los objetivos: El objetivo de la investigación en I.A. no puede ser producir SAD sin más, sino crear inteligencia que beneficie a la humanidad.

2. De la financiación: Las inversiones en investigación en I.A. deben ir acompañadas de fondos que aseguren el uso beneficioso de la misma, incluyendo todos aquellos aspectos que puedan resultar complejos en Informática, Economía, Derecho, Ética o estudios sociales.

3. De la relación entre Política y Ciencia: Mantener un intercambio sólido, claro y cons-

tructivo entre los responsables de la política científica y los investigadores en I.A.

4. De las formas de la investigación: Fomentar una cultura de cooperación, confianza y transparencia entre investigadores y desarrolladores de SAD.

5. Del límite en la competición: Los equipos desarrollando SAD han de cooperar activamente evitando los atajos en el cumplimiento de los estándares de seguridad.

b) Principios referidos a Ética y Valores

6. De la seguridad: Siempre y cuando sea aplicable y factible, los SAD tienen que ser seguros y fiables durante toda su vida operativa.

7. De la transparencia de los fallos: Si un SAD causara perjuicios, ha de ser posible saber cuáles fueron las causas.

8. De la transparencia judicial: Cualquier intervención de un SAD que pueda conllevar una decisión judicial se tiene que poder explicar, de forma satisfactoria y auditable, por parte de las autoridades humanas.

9. De la responsabilidad: Las implicaciones morales derivadas del buen o mal uso y de las acciones que puedan cometer los SAD

son responsabilidad de los diseñadores y constructores de los mismos, que tienen la oportunidad de darle forma a las mismas.

10. De la alineación de valores: Los SAD altamente automáticos han de diseñarse de modo que mientras estén operativos, sus metas y conductas permanezcan alineadas con los valores humanos.

11. De los valores humanos: Los SAD han de estar diseñados y operar de forma compatible con los ideales humanos de dignidad, derechos, libertades y diversidad cultural.

12. De la intimidad personal: Las personas tienen derecho a acceder, gestionar y controlar los datos que generan, dando a los SAD la capacidad de utilizar y analizar esos datos.

13. De la intimidad y libertad: La aplicación de los SAD a datos personales no puede restringir injustificadamente la libertad real o percibida de las personas.

14. Del beneficio compartido: Las tecnologías asociadas a los SAD han de beneficiar y favorecer a tantas personas como sea posible.

15. De la prosperidad compartida: La prosperidad económica que creen los SAD tiene

que ser compartida ampliamente para el beneficio de toda la humanidad.

16. Del control humano: Las personas han de poder elegir como y cuando delegar sus decisiones en los SAD, a fin de poder cumplir con los objetivos que hubieran seleccionado con anterioridad.

17. De no inversión: El poder que puedan tener los SAD, más que invertir tiene que respetar y mejorar los procesos cívicos y sociales de los que depende el bienestar social.

18. De la carrera armamentista: Evitar participar en cualquier tipo de carrera armamentista en la que los SAD puedan llegar a intervenir como armas letales.

b) Principios a tener en cuenta a largo plazo

19. De la capacidad de precaución: Cuando no haya consenso, evitar hipótesis sobre cuales han de ser los límites superiores de las capacidades futuras de los SAD.

20. De la importancia: Los SAD pueden representar un cambio profundo en la historia de la vida en la Tierra, por lo que se deben planificar y gestionar cuidadosamente y siempre con los recursos adecuados.

21. Del riesgo: Los riesgos que conllevan los SAD, especialmente los de tipo catastrófico o existencial, estarán sujetos a medidas de planificación y mitigación acordes con el impacto esperado.

22. De la auto-mejora recursiva: Los SAD diseñados para auto-mejorarse o auto-replicarse recursivamente, pudiendo producir un aumento rápido de su calidad o cantidad, tienen que estar sujetos a estrictas medidas de seguridad y control.

23. Del bien común: Las superinteligencias solo se desarrollarán al servicio de ideales éticos ampliamente compartidos que beneficien a toda la humanidad más que a alguna organización de carácter regional.

Todas estas recomendaciones, estrategias y principios que se propone que se pongan en práctica, de nada servirán si uno de los actores principales, los SAD y por tanto los algoritmos, pueden actuar por libre, es decir, si tienen capacidad de pensar de manera autónoma, sin supervisión.

Pero los algoritmos no piensan. Tampoco lo hacen los programas mediante los que se aplican. Es verdad que los resultados que a veces presentan esos programas parece que son el resultado de una forma de pensamiento o de razonamiento. Pero no

hay pensamiento ni razonamiento volitivo. No es más que la ilusión que experimentamos al ver que las soluciones que nos aportan (sin conciencia de ningún tipo, es decir, sin tener conciencia de lo que están obteniendo) son muy buenas y nunca antes se nos habían ocurrido a nosotros mismos.

Ahí, justo ahí, es donde aparecen las dudas, las desconfianzas y los temores, porque sospechamos que esos programas, la I.A. como un todo o si queremos los algoritmos, también pueden dar soluciones que no nos gusten, no conozcamos o nos perjudiquen. Pero esos algoritmos, esos programas, esos sistemas basados en I.A. están diseñados, programados e implementados por nosotros mismos, quienes somos en definitiva los responsables de su funcionamiento, ya sea perjudicial, beneficioso o erróneo.

En cualquier caso, como es patente, los principales avances en I.A. de la última década han manifestado su capacidad como tecnología de propósito general, impulsando la innovación en múltiples áreas, como movilidad, salud, robótica doméstica y de servicios, educación o seguridad cibernética, y consiguiendo generar importantes beneficios, no solo para los individuos sino también para la sociedad en general, tendiendo a que el abanico de aplicaciones vaya en aumento en los próximos años, al

igual que ocurrirá con los beneficios que nos aportarán a la hora de abordar y resolver desafíos trascendentales, como es el caso del cambio climático o el diagnóstico de enfermedades raras y el bienestar mundial. Pero al mismo tiempo, su aplicación y uso conlleva riesgos y desafíos asociados con los derechos humanos fundamentales y la Ética.

Todo ello explica que un gran número de países [1, 3, 8, 12, 14, 19, 20, 21] hayan elaborado planes estratégicos para maximizar los beneficios del uso de los SAD y minimizar sus riesgos. También en esa línea se encuentra España [10], donde aparte de la presentación que realizó el gobierno en Granada, el pasado mes de marzo, de la Estrategia Española de I+D+I en Inteligencia Artificial, y que establece una serie de prioridades que serán enmarcadas en la nueva Estrategia Española de Ciencia, Tecnología e Innovación 2021-2028, hay otras organizaciones, como por ejemplo el Real Instituto Elcano o la Cátedra Privacidad y Transformación Digital Microsoft-Universidad de Valencia, que están realizando importantes contribuciones [2,7].

Tenemos derecho a decidir cómo queremos que sea nuestro futuro. En la misma medida tenemos la obligación de tratar de evitar los perjuicios que pueda causar la aplicación de los algoritmos en el ámbito de la I.A.

Yo, como José Luis Aranguren [9] «No me veo legitimado para dar ningún recado a nadie», pero si para subrayar que para contribuir a minimizar esos posibles efectos negativos, muchas de las anteriores propuestas y recomendaciones pueden impulsarse desde la Universidad, con muy bajo costo y en muy corto plazo, sin más que activar medidas como por ejemplo las siguientes:

1. Incentivar las acciones formativas sobre I.A. en todos los ámbitos sociales y niveles educativos, como garantía de conocimiento de riesgos y oportunidades y como seguro ante la disrupción que pueda provocar la I.A. en el mercado de trabajo y la economía.

2. Reforzar el apoyo a los ecosistemas locales del ámbito de la I.A. fomentando la experimentación, la captación y retención del talento y la I + D, para facilitar el mantenimiento y la creación de puestos de trabajo en la economía local.

3. Crear una Comisión de Garantías Éticas sobre el uso de la I.A. y sus aplicaciones, que también vele por asegurar el principio de transparencia de los algoritmos, que en todos los casos han de ser auditables, y las políticas de "open data".

4. Garantizar el derecho a la intimidad y seguridad de las personas sobre los datos, y en especial sobre sus datos, mediante reglamentos propios adecuados a la normativa legal, cuya referencia podría incorporarse a los convenios científico-académicos que corresponda.

5. Favorecer y explorar nuevas acciones a favor de la inclusión y la igualdad y contra la violencia y la brecha de género a partir de algoritmos basados en I.A.

Una última reflexión para terminar. Los algoritmos, los computadores o los sistemas basados en I.A. no piensan en nada. Pueden razonar sin pensar porque podemos dotarlos de mecanismos de razonamiento, pero no de pensamiento. Sin embargo las personas razonamos porque pensamos, es decir, mientras que los computadores pueden razonar deductivamente por medio de algoritmos, las personas podemos hacerlo también de forma inductiva y no solo algorítmica. Desde este punto de vista, consciente de que la aplicación de la I.A. conllevará una automatización de trabajos y tareas que transformará la economía y el actual modelo social en los próximos 15 o 20 años, que esa transformación sea suave y cause los mínimos perjuicios posibles solo nos concierne a nosotros, que

deberíamos ocuparnos y preocuparnos de capacitar a las personas de cualquier nivel social para que pudieran alcanzar el máximo nivel formativo posible en su entorno socio-laboral.

Como dijo el universal Federico García Lorca [15]: «Bien está que todos los hombres coman, pero que todos los hombres sepan. Que gocen todos los frutos del espíritu humano porque lo contrario es convertirlos en máquinas al servicio del Estado, es convertirlos en esclavos de una terrible organización social».

Muchas gracias por su atención.

Almuñécar, agosto de 2019

REFERENCIAS

[1] AI Sector Deal. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/702810/180425_BEIS_AI_Sector_Deal__4_.pdf

[2] F. Arteaga y A. Ortega (2019): Hacia un ecosistema español de Inteligencia Artificial: una propuesta. Real Instituto Elcano (2019).

[3] Artificial Intelligence at the service of the citizen. https://ai-white-paper.readthedocs.io/en/ latest/

[4] Artificial Intelligence Market Forecasts https://www.tractica.com/research/artificial-intelligence-market-forecasts/

[5] Asilomar Principles. https://futureoflife.org/ai-principles

[6] M.S. Bazaraa, H.D. Sherali and C. M. Shetty: Nonlinear Programming: Theory and Algorithms. Wiley (2006)

[7] Cátedra Privacidad y Transformación Digital Microsoft-UV. https://www.uv.es/catedra-microsoft/es/catedra-privacidad-transformacion-digital-microsoft-uv.html

[8] A New Generation of Artificial Intelligence Development Plan. https://flia.org/wp-content/uploads/2017/07/A-New-Generation-of-Artificial-Intelligence-Development-Plan-1.pdf

[9] EL PAÍS, Cultura, 18 de abril de 1996. https://elpais.com/diario/1996/04/18/cultura/829778401_850215.html

[10] Estrategia Española de I+D+I en Inteligencia Artificial. Ministerio de Ciencia, Innovación y Universidades. Marzo (2019).

[11] European Group on Ethics in Science and New Technologies: Statement on A.I., Robotics and Autonomous Systems. European Commission. Directorate-General for Research and Innovation Unit RTD.01 (2018).

[12] Finland's Age of Artificial Intelligence. https://tem.fi/en/artificial-intelligence-programme

[13] P. Foot: The Problem of Abortion and the Doctrine of the Double Effect. Oxford Rev. 5 (1967).

[14] For a Meaningful Artificial Intelligence: Towards a French and European Strategy. https://www.aiforhumanity.fr/pdfs/Mission-Villani_Report_ENG-VF.pdf

[15] F. García Lorca: Medio pan y un libro. Discurso pronunciado por Federico García Lorca en la inauguración de la biblioteca de Fuente Vaqueros, Granada (1931). https://algundiaenalgunaparte. com/2016/06/09/medio-pan-y-un-libro-de-federico-garcia-lorca

[16] Gartner Top 10 Strategic Technology Trends for 2019 https://www.gartner.com/ smarter-withgartner/gartner-top-10-strategic-technology-trends-for-2019/

[17] Informatics Europe: When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making. Informatics Europe & EUACM (2018). https://www.informatics-europe.org/publications.html

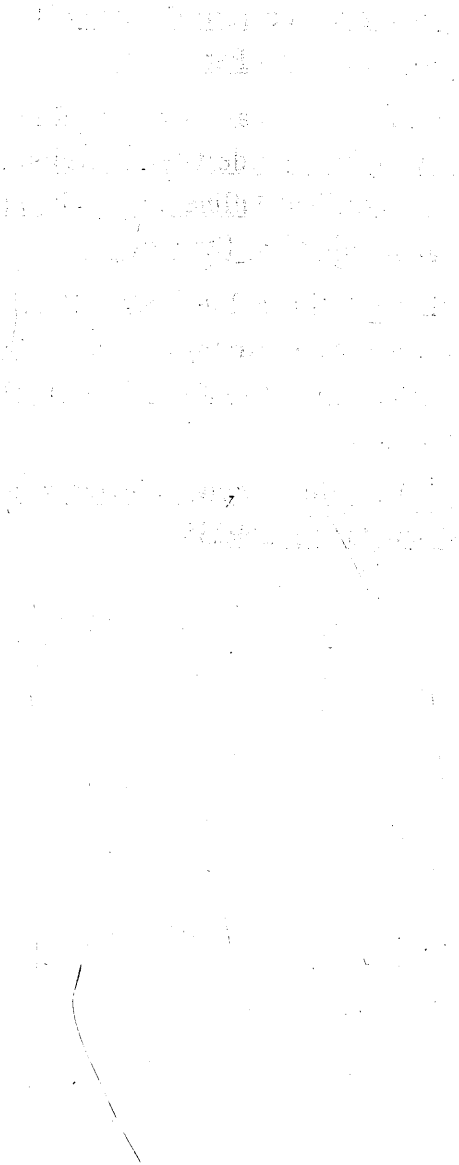
[18] M.T. Lamata, D.A. Pelta y J.L. Verdegay: Fuzzy Information and Contexts for Designing Automatic Decision-Making Systems. Advances in Artificial Intelligence. Lecture Notes in Artificial Intelligence 11160. Springer Cham 174-183 (2018).

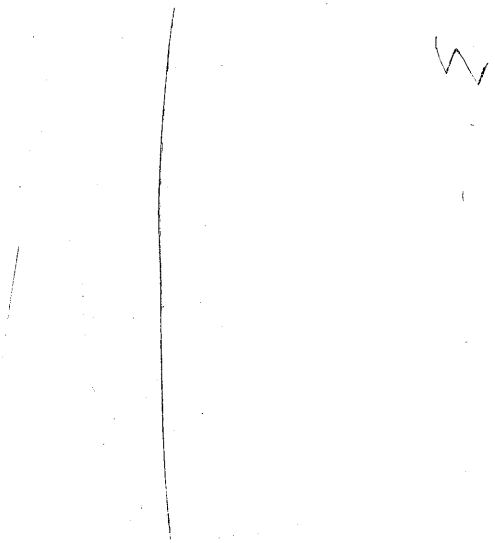
[19] National Strategy for Artificial Intelligence. http://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf

[20] Outline for a German Strategy for A.I. https://www.ip.mpg.de/fileadmin/ipmpg/content/ aktuelles/Outline_for_a_German_Artificial_Intelligence_Strategy.pdf

[21] The National Artificial Intelligence Research and Development Strategic Plan. https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf

[22] Turing, A.: https://www.brainyquote.com/quotes/alan_turing_269238





UNIVERSIDAD
DE GRANADA