

ARTÍCULO DE REFLEXIÓN

Consideraciones éticas sobre el diseño, desarrollo y aplicación de los Sistemas Autónomos de Decisión

*Ethical Aspects about Design, Development and Engage
of Autonomous Decision Systems*

José Luis Verdegay

verdegay@ugr.es • <https://orcid.org/0000-0003-2487-942X>

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN E INTELIGENCIA ARTIFICIAL,
UNIVERSIDAD DE GRANADA, ESPAÑA

Recibido: 2019-04-16 • Aceptado: 2020-03-31

RESUMEN

En estos momentos todos estamos debatiendo sobre la cuarta revolución industrial, en la que las tecnologías digitales coexisten y colaboran con las Ciencias clásicas en la automatización de procesos y servicios en todas las áreas de la vida humana. Esta cooperación tiene importantes consecuencias en muchos contextos y entornos sociales. Además, la vertiginosa velocidad a la que llegan las nuevas tecnologías y su asunción social, con mayor frecuencia que la deseada, impiden la asimilación de los cambios por parte de la Sociedad y, lo que es más importante, produce una imposibilidad práctica de ofrecer respuestas éticas, legales y sociales correctamente adaptadas a los ritmos y al alcance del proceso de transformación. Esta disrupción tecnológica genera nuevos problemas y conflictos, que incluso pueden dañar los derechos humanos. Por esta razón, desde muchas instituciones, áreas y organismos profesionales, se han iniciado acciones destinadas a analizar las repercusiones que los sistemas autónomos basados en aprendizaje computacional pueden tener en nuestra forma de vida actual. Este artículo está dedicado a presentar algunas de las iniciativas diseñadas desde la perspectiva profesional, sociopolítica o científica, con especial énfasis en aquellas orientadas a los aspectos éticos asociados con estos sistemas, con especial referencia a los casos europeo y español, mostrando algunos sistemas autónomos para la toma de decisiones en dos áreas muy importantes: la sanitaria y la jurídica.

PALABRAS CLAVE: inteligencia artificial; ética; sistemas autónomos de decisión.

ABSTRACT

We are just debating about the fourth industrial revolution, in which digital technologies coexist and collaborate with classical Sciences in the automation of processes and services in all areas of human life. This cooperation has important consequences in many contexts and social environments. Besides, the vertiginous speed at which new technologies are arriving and their social assumption, more often than the desired one, prevent the assimilation of changes by society and, what is more important, it produces a practical impossibility to offer ethical, legal and social answers right adapted to the rhythms and scope of the transformation process. This technological disruption generates new problems and conflicts, which can even injure human rights. For this reason, from many institutions, areas and professional bodies, actions have been initiated aimed at analyzing the repercussions that machine-learned automated systems can have on our current way of life. This paper is dedicated to present some of the initiatives designed from the professional, socio-political or scientific perspective, with special emphasis on those oriented to the ethical aspects associated with these systems and particular reference to the European and Spanish instances and showing some automatic systems for decision making in two key areas: health and legal.

KEYWORDS: artificial intelligence; ethics; autonomous decision systems.

INTRODUCCIÓN

Las Tecnologías de la Información y la Comunicación (TIC) están transformándolo todo. El impacto sobre la Sociedad de las nuevas tecnologías es trascendental, pues significan nuevas formas de comunicación, de relación, de supervivencia y de progreso. Este cambio afecta a las personas individualmente y a la Sociedad en general, a los procesos de negocio en las cadenas de valor, a las formas conocidas de trabajo, al trabajo en sí mismo o a las universidades entre muchos otros ámbitos. Es justo por esto por lo que hoy hablamos de “Transformación digital”, entendida como el cambio asociado con la aplicación de la tecnología digital en todos los aspectos de la sociedad humana. De todos esos aspectos, de sus repercusiones, exigencias, métricas o sostenibilidad, en lo que sigue nos concentraremos en explicar su influencia en la Sociedad a través de los desarrollos provenientes del área de la Inteligencia artificial (I.A.).

En nuestro día a día nos enfrentamos de forma constante a la toma de decisiones: desde revisar nuestros teléfonos celulares, a decidir si cruzamos o no una calle. Así, mientras se sabe que realizamos una media de 35.000 decisiones diarias, un reciente informe de *Huawei* (Huawei, 2017) muestra que realmente somos conscientes de menos de un 1% de las decisiones que tomamos a diario, de hecho, pasamos por alto un 99,74% de las decisiones que tomamos.

Es obvio que los sustanciales avances que se han producido en la última década en aprendizaje automático (*machine learning*) han producido sistemas que compiten con las conductas de las personas en situaciones que supongan desafíos, estén ambigüamente planteadas o requieran altas dosis de habilidad, como por ejemplo el reconocimiento del habla o de imágenes, los juegos, etc. A veces incluso nos superan.

Estos sistemas, a los que de manera genérica en lo que sigue denominaremos Sistemas Autónomos de Decisión (SAUD), como lugar común de sistemas basados en técnicas y modelos de Inteligencia artificial o Robótica o Sistemas autónomos, cuando se gestionan con técnicas de automatización inteligente pueden aumentar y en algunos casos sustituir, mediante sistemas completamente autónomos, la capacidad de actuar, de tomar decisiones, de las personas humanas. El hecho crucial es que esa “sustitución” de funciones podría producir, más pronto que tarde, pérdidas masivas de puestos de trabajo, la descalificación de las personas que los desempeñaban, efectos desconocidos en los sistemas que gestionarán o contextos de ingobernabilidad no deseados, por lo que resulta patente que:

- a) las cuestiones éticas relativas al comportamiento de los SAUD deben incluirse en su diseño tecnológico, para facilitar que más que como “factores de riesgo” o restricciones, se desempeñen como los motores tractores de la innovación, y
- b) en caso de producirse esa sustitución de funciones, que no produzca disfunciones, es decir, que el correspondiente SAUD 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 (Thaler and Sunstein, 2008).

Conscientes de estas dos exigencias sobre el carácter ético y el correcto funcionamiento de los SAUD, múltiples instituciones 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. Son muchas las iniciativas llevadas a cabo hasta el momento, y muy variadas las perspectivas a considerar según se contemplen entidades públicas o privadas, el entorno geográfico, el ámbito profesional, etc. Entre ellas, seguramente debido al interés social y a la trascendencia del tema, algunas pueden verse un tanto precipitadas en sus recomendaciones, o incluso sesgadas por estar movidas por intereses particulares pero, en cualquier caso, todas orientadas a conseguir que los SAUD funcionen y se desempeñen de manera tan conforme a los deseos de las personas como sea posible.

Como es prácticamente imposible recoger todas y cada una de esas iniciativas en un artículo como este, hemos optado por un enfoque limitado, pero multisectorial, que recoge los

puntos de vista de cuatro sectores concretos asociados a los SAUD: el productivo privado, el político europeo, el científico global y el más particular de España (figura 1).

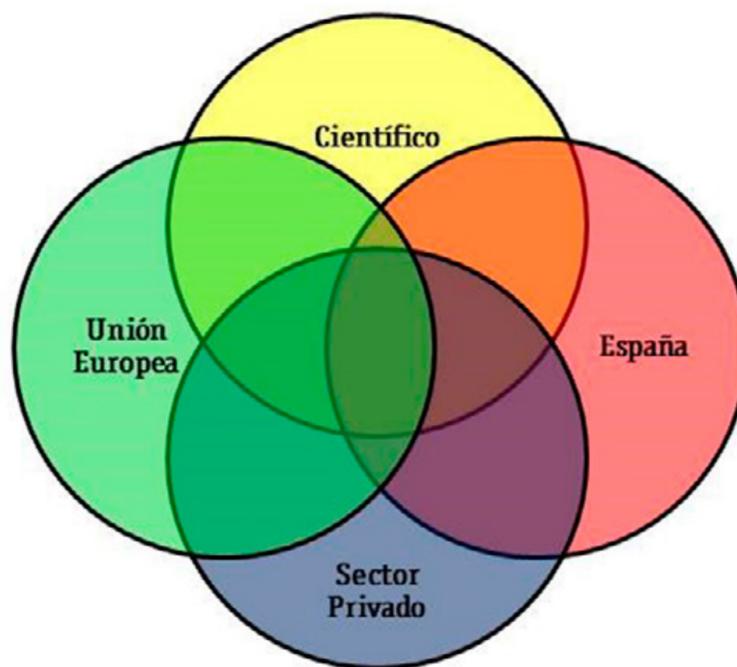


Figura 1: Sectores en los que se elaboran propuestas.

Por tanto en cada una de las siguientes secciones se aborda uno de esos ejes respectivamente. Así en la siguiente sección presentamos la perspectiva del sector empresarial privado. En la sección 3 abordamos el punto de vista político-científico de la Unión Europea, dedicando la sección 4 a mostrar la opinión de expertos científicos de ámbito general. Después, en la sección 5, recogemos la estrategia que actualmente se desarrolla en España y finalmente, en la sexta y última sección, recogemos algunos SAUD actualmente en explotación que se desempeñan en ámbitos en los que la ética juega un papel especialmente protagonista. El artículo concluye con un apartado de conclusiones.

RECOMENDACIONES DEL SECTOR PRIVADO EUROPEO SOBRE SAUD

Hace unos meses *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* (ACM), el *ACM Europe Policy Committee* (EUACM) y el *European Research Consortium for Informatics and Mathematics* (ERCIM) publicaron un informe (Larus, y otros, 2018) que recoge las medidas que se habrán de tener en cuenta para lograr un desarrollo equilibrado y eficaz de los SAUD en la Sociedad.

Dado que en la práctica sería peligroso responsabilizar de las cuestiones éticas relacionadas con los SAUD 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 SAUD 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.

RECOMENDACIONES DE TIPO TÉCNICO

1. Establecer medios, medidas y estándares para asegurar que los SAUD 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 SAUD. 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 SAUD. 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.

RECOMENDACIONES DE TIPO ÉTICO

2. Asegurar que la ética se mantenga a la vanguardia del desarrollo y la implementación de SAUD y que sea parte integral de esta. Al igual que con la salud y la biología, los países miembros y la Unión Europea 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 SAUD. También, como garante del interés público, debería crearse una nueva Agencia europea que supervisara el desarrollo y despliegue de los SAUD en toda Europa.
3. Promover el diseño de SAUD 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 SAUD.

RECOMENDACIONES DE TIPO LEGAL

4. Definir claramente la responsabilidad legal del uso e impacto de los SAUD. Los principios básicos que gobiernan actualmente el desarrollo de SAUD 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 SAUD. 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 SAUD. La agencia europea propuesta en la Recomendación 2 debe fomentar y facilitar este debate y proponer una legislación adecuada.

RECOMENDACIONES DE TIPO ECONÓMICO

5. Garantizar que las consecuencias económicas de la adopción de SAUD sean siempre consideradas en su totalidad. Entre sus primeras iniciativas oficiales, con el fin de emi-

tir 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 SAUD 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 SAUD y minimizar sus posibles interrupciones de tipo personal, social y económico sobre los individuos y las naciones.

RECOMENDACIONES DE TIPO SOCIAL

6. Imponer legalmente que se informe claramente a los usuarios de los SAUD sobre todas las prácticas de privacidad y adquisición de datos de sus implementadores. El aprendizaje autónomo 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 SAUD. 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.

RECOMENDACIONES DE TIPO EDUCATIVO

8. Fomentar la formación técnica universitaria relacionada con los SAUD. 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 interdisciplinarios, 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 SAUD y a los problemas que plantean.
10. Aumentar la conciencia y comprensión del público sobre los SAUD 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 formar a la opinión pública.

La Unión Europea también preocupada por este tema, pero convencida de que los SAUD 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 sobre las mismas a través del *European Group on Ethics in Science and New Technologies* (European Commission, 2018).

DECLARACIÓN DEL GRUPO EUROPEO DE ÉTICA EN LA CIENCIA Y LAS NUEVAS TECNOLOGÍAS

La Declaración exige la apertura de un proceso que allanaría el camino hacia un marco ético y legal común, reconocido internacionalmente, para el diseño, producción, uso y gobierno de los SAUD. También propone un conjunto de principios éticos fundamentales, basados en los valores establecidos en los tratados de la UE y la Carta de los Derechos Fundamentales de la UE, que recogemos sucintamente a continuación.

PRINCIPIOS ÉTICOS Y PRE-REQUISITOS DEMOCRÁTICOS

- a) **De dignidad humana:** el Principio de la dignidad humana, entendido como el reconocimiento del estado humano inherente a ser digno de respeto, no debería ser violado por las tecnologías autónomas o automáticas. Entre otras cosas, esto supone poner límites a las determinaciones y clasificaciones relativas a las personas, hechas sobre la base de algoritmos y SAUD, 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 inteligentes. 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, para tener derecho a asignar unas tareas al ser humano y otras a la máquina.
- b) **De autonomía:** este principio afectaría a la libertad del ser humano, lo que se traduce en la responsabilidad humana sobre los SAUD, y por lo tanto de su control y conocimiento, ya que no deben impedir la libertad de los seres humanos para establecer sus propios estándares y normas y poder vivir de acuerdo con ellos. 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 implicaría la transparencia y

- la previsibilidad de los SAUD, sin los cuales los usuarios no podrían intervenir o terminarlos si lo consideraran moralmente necesario.
- c) **De responsabilidad:** el Principio de responsabilidad debe ser fundamental para la investigación y aplicación de los SAUD, que solo deberían desarrollarse y utilizarse 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 tecnologías autónomas y automáticas plantea un gran desafío, hay que tomar conciencia de ello y por tanto las aplicaciones de los SAUD 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 SAUD 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 reste valor.
- d) **De justicia, igualdad y solidaridad:** los SAUD deben contribuir 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 SAUD, así como la distribución justa de los beneficios y la igualdad de oportunidades entre las sociedades y dentro de ellas. 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 SAUD, asegurar el acceso a las principales tecnologías basadas en I.A. y facilitar la capacitación en 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) **De democracia:** las decisiones clave para regular el desarrollo y la aplicación de los SAUD 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 facilitará 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 humanas 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) **De estado de derecho y rendición de cuentas:** 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 SAUD. 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 SAUD. Además, deberían existir sistemas eficaces para minimizar los daños.
- g) **De seguridad, protección e integridad física y mental:** la seguridad y la protección de los SAUD se materializan en tres formas:
1. Seguridad externa para su entorno y usuarios,
 2. Confiabilidad y solidez interna, por ejemplo contra la piratería y
 3. Seguridad emocional con respecto a la interacción hombre-máquina.
- Los desarrolladores de SAUD 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 automáticos 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) **De protección de datos y privacidad:** 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 de 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 sistemas autónomos 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 los SAUD. En vista de las preocupaciones con respecto a las implicaciones de los SAUD en la vida privada y la privacidad, se puede considerar el debate en curso sobre la introducción de dos nuevos derechos: 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) **De sostenibilidad:** la tecnología asociada a los SAUD 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.

Por otro lado, el *Future of Life Institute* es una institución que tiene por misión catalizar y apoyar la investigación y las iniciativas para salvaguardar la vida y desarrollar visiones optimistas del futuro, incluyendo formas positivas para que la humanidad siga su propio curso, considerando nuevas tecnologías y desafíos. Entre los temas de su interés viene priorizando el impulso y estudio de proyectos que fortalezcan los efectos beneficiosos de la I.A. Sobre esto se celebró en enero de 2017 la *Asilomar Conference on Beneficial A.I.* en California (EE.UU.) en el *Asilomar Conference Grounds*, a cuyos resultados relativos al tema que aquí se trata, se dedica la siguiente sección.

LOS 23 PRINCIPIOS DE ASILOMAR

La conferencia aludida 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 se celebraron, surgieron los denominados “Principios de Inteligencia Artificial de Asilomar” (AI/Robotics researches, 2017). La lista, consensuada por el 90% de los expertos convocados, y en la que hasta ahora han participado más de 890 investigadores en I.A. y otros 1445 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 SAUD.

A continuación recogemos los 23 Principios consensuados:

1. **De los objetivos:** el objetivo de la investigación en I.A. no puede ser producir SAUD sin más, sino crear inteligencia que beneficie a la humanidad.
2. **De la financiación:** las inversiones en investigación en I.A. debe ir acompañada de fondos que aseguren el uso beneficioso de la misma, incluyendo aspectos que puedan resultar espinosos en Informática, Economía, Derecho, Ética o estudios sociales, como por ejemplo:
 - a) ¿Cómo podemos hacer altamente robustos los futuros SAUD para que hagan justo lo que queramos sin funcionar mal o sin que sean hackeados?
 - b) ¿Cómo podemos aumentar la prosperidad a través de la automatización, manteniendo los recursos y los objetivos de las personas?
 - c) ¿Cómo podemos actualizar nuestros sistemas legales para que sean más justos y eficientes, puedan seguir el ritmo de la I.A. y sirvan para gestionar los riesgos asociados al uso de SAUD?
 - d) ¿Con que conjunto de valores deberían alinearse los SAUD y que status legal y ético deberían tener?
3. **De la relación entre Política y Ciencia:** hay que mantener un intercambio sólido, claro y constructivo entre los responsables de la política científica y los investigadores en I.A.
4. **De las formas de la investigación:** hay que fomentar una cultura de cooperación, confianza y transparencia entre investigadores y desarrolladores de SAUD.
5. **Del límite en la competición:** los equipos desarrollando SAUD han de cooperar activamente evitando los atajos en el cumplimiento de los estándares de seguridad.
6. **De la seguridad:** siempre y cuando sea aplicable y factible, los SAUD tienen que ser seguros y fiables durante toda su vida operativa.
7. **De la transparencia de los fallos:** si un SAUD causara perjuicios, ha de ser posible saber cuáles fueron las causas.
8. **De la transparencia judicial:** cualquier intervención de un SAUD 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 SAUD 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 SAUD altamente automáticos deberían 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 SAUD 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 SAUD la capacidad de utilizar y analizar esos datos.
13. **De la intimidad y libertad:** la aplicación de los SAUD a datos personales no puede restringir injustificadamente la libertad real o percibida de las personas.
14. **Del beneficio compartido:** las tecnologías asociadas a los SAUD han de beneficiar y favorecer a tantas personas como fuera posible.
15. **De la prosperidad compartida:** la prosperidad económica que creen los SAUD 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 SAUD, a fin de poder cumplir con los objetivos que hubieran seleccionado con anterioridad.
17. **De no inversión:** el poder que puedan tener los SAUD, más que invertir, tiene que respetar y mejorar los procesos cívicos y sociales de los que depende el bienestar social.
18. **De la carrera armamentística:** evitar participar en cualquier tipo de carrera armamentística en la que los SAUD puedan llegar a intervenir como armas letales.
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 SAUD.
20. **De la importancia:** los SAUD 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 SAUD, 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 SAUD 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.

La claridad, legibilidad y prospectiva del funcionamiento de los SAUD preocupa a usuarios, inversores e innovadores. Disponer de SAUD confiables y éticos, aparte de eventuales dificultades técnicas, conlleva algún problema más en el ámbito europeo, caracterizado por la aplicación consensuada de los derechos fundamentales durante décadas. La multinacionalidad europea requiere un marco legal, adaptado a cada país en particular, que facilite que los SAUD se utilicen de manera justa, transparente y responsable, combata los sesgos no intencionados y garantice que las inquietudes sociales y económicas relacionadas con la pérdida de empleos, la competencia desleal o el posible dominio de las máquinas se aborden de forma adecuada.

La siguiente sección será dedicada a describir como se está abordando esta situación en el caso particular español.

EL PANORAMA EN ESPAÑA

Lo mismo que en otros países, como Francia, Alemania, Finlandia, China, EE.UU., Reino Unido, India o Italia (Villani, 2018), (Harhoff, Heumann, Jentzsch, & Lorenz, 2018), (Steering Group of the Artificial Intelligence Programme, 2017), (Council, 2017), (National Science and Technology Council, 2016), (Department of Business, Energy & Industrial Strategy - Department for Digital, Culture, Media & Sport, 2019), (Niti Aayog, 2018), (Agencia Digital for Italy, 2018) y otros muchos (Dutton, 2018), el Gobierno de España presentó en marzo de 2019 la Estrategia Española de I+D+I en Inteligencia Artificial (Ministerio de Ciencia, Innovación y Universidades, 2019), que establece una serie de prioridades que serán enmarcadas en la nueva Estrategia Española de Ciencia, Tecnología e Innovación 2021-2028 y que deberán ser desarrolladas en iniciativas y actividades definidas y financiadas a través de los Planes Estatales de Ciencia, Tecnología e Innovación, movilizandando las sinergias entre los distintos niveles de la administración pública y mediante el codesarrollo del sector público y privado. También incluye una serie de recomendaciones que trascienden la I+D+I y reclaman la presencia de otros sectores y departamentos ministeriales debido a la naturaleza multidisciplinar y transversal de la I.A. y la revolución tecnológica y social que implica.

La Estrategia marca las seis siguientes prioridades:

1. Lograr una estructura organizativa que permita desarrollar un sistema de I+D+I en IA y medir su impacto.
2. Establecer áreas estratégicas en las que es necesario centrar los esfuerzos de las actividades de I+D+I.
3. Facilitar la transferencia del conocimiento y su retorno a la Sociedad.
4. Planificar las acciones de formación y profesionalización en el ámbito de la IA.
5. Desarrollar un ecosistema digital de datos y valorizar las infraestructuras disponibles.
6. Analizar la ética de la IA desde la perspectiva de la I+D+I.

Así mismo realiza las siete siguientes recomendaciones:

- Lanzar una Estrategia Nacional para la I.A. que permita el desarrollo e implementación de medidas específicas dirigidas a los sectores estratégicos nacionales. La evaluación y seguimiento de dichas medidas podrá ser realizada a través de un Observatorio Español de la I.A.
- Aprovechar la I.A. para alcanzar los objetivos marcados en la Agenda 2030.
- Diseñar e Implementar actuaciones específicas que impulsen la transferencia de conocimiento al entorno socioeconómico.
- Lanzar o adaptar los programas de fomento de vocaciones, no limitado a la I+D, así como la atracción, retención y recuperación de talento dirigidas a la I.A.
- Usar la I.A. para garantizar un uso óptimo de los datos abiertos. Crear un Instituto Nacional de Datos que permita planificar y definir una gobernanza sobre los datos procedentes de los diferentes niveles de la Administración Pública.
- Detectar las necesidades de adaptación y mejora de competencias en los distintos niveles de nuestro sistema educativo.

- Velar porque todas las actividades e iniciativas derivadas de los marcos estratégicos enfocados al desarrollo de la I.A., así como sus resultados cumplen con los compromisos éticos, legales y sociales de nuestro país y de nuestro entorno europeo.

En esa línea, en la Reunión sobre Inteligencia Artificial y Desarrollo Humano celebrada en la Universidad de Granada en diciembre de 2018, en colaboración con la Cátedra de Privacidad y Transformación Digital *Microsoft* de la Universidad de Valencia se propuso que la Estrategia Española contemple acciones en diferentes ámbitos concretos, como por ejemplo el socio-pedagógico, el educativo, el laboral, el empresarial, el investigador y el ético-legislativo.

Más específicamente:

1. Igual que está perfectamente asumida la existencia de Comités de Ética, que aseguran la protección de los derechos y el bienestar de los seres humanos que participan en los ensayos clínicos, procurando lograr la confianza del público mediante la inspección previa de los protocolos de los ensayos, en el ámbito de los SAUD hace falta instaurar comisiones que tengan misiones análogas y en las que puedan intervenir expertos de diferentes formaciones: humanistas, matemáticos, ingenieros, etc. Por otro lado hay que regular el diseño, comercialización y uso de los SAUD, con especial atención a la preservación de la privacidad, la no discriminación y la asunción de responsabilidades por los perjuicios de cualquier índole que puedan derivarse de un mal funcionamiento, uso o abuso de algún tipo de máquina inteligente.

En cualquier caso, y de acuerdo con los trabajos realizados por el Grupo de Expertos de Alto Nivel en IA de la Unión Europea (European Commission, 2019), en esas comisiones se debería verificar si el sistema que estuvieran analizando es un sistema en el que se puede confiar comprobando aspectos tales como:

- **Responsabilidad:** tiene que estar claro quién es el responsable ante los errores, fallos, etc.
- **Gobernanza de los datos:** hay que declarar la legislación que se ha seguido.
- **Diseño equitativo:** el SAUD tiene que estar concebido para pueda usarlo cualquier persona.
- **Control de la autonomía del SAUD:** interesará saber si está previsto algún tipo de control por humanos.
- **No discriminación:** si puede haber diferentes conductas ante situaciones similares, hay que dilucidar si estas pueden afectar a derechos humanos o principios éticos.
- **Respecto a la intimidad:** el flujo de información con datos personales debe estar continuamente bajo control y ajustándose a las leyes sobre protección de datos vigentes.
- **Respeto a la autonomía humana (y su ampliación):** se persigue conocer si el SAUD puede poner en riesgo a los usuarios (evitar la inducción).
- **Robustez:** se mediría la resistencia a posibles ataques, la fiabilidad y reproducibilidad, la precisión de los resultados a consecuencia del uso y control de datos y la existencia de opciones de marcha atrás, previstas para el caso de fallos.
- **Seguridad:** hay que explicitar las medidas de seguridad que se van a usar en el contexto en el que se vaya a emplear el SAUD, de modo que se tenga constancia de la seguridad del sistema.

- **Transparencia:** el propósito y la trazabilidad del SAUD, describiendo la técnica algorítmica empleada, la forma de verificación seguida y los resultados que se obtienen han de estar establecidos y ser comprensibles y accesibles.

Pero también serán trascendentales otras acciones en otros contextos, como por ejemplo pueden ser los siguientes:

2. El Pedagógico social, ya que la concienciación y formación de la ciudadanía en cuanto a la presencia, incidencia, repercusión y uso de los SAUD es prioritaria. Una ciudadanía bien informada y formada estará mucho mejor preparada para afrontar los cambios que la van a permear hasta transformarla, y sabrá aprovechar con garantías las oportunidades que se van a presentar.
3. El Educativo, puesto que hay que introducir nuevos hábitos docentes y pedagógicos en los diferentes planes de estudio que, lejos de insistir en procesos memorísticos, se dirijan a conseguir nuevas competencias que potencien las famosas cuatro ces: la creatividad, la crítica, la comunicación y la colaboración, por ser entornos intelectuales en los que los computadores no pueden competir con las personas.
4. El Laboral, para garantizar desde las instancias públicas el empleo sostenible, estable y de calidad para las personas, ya que siendo imparable el desembarco de los SAUD en este entorno, y por tanto la destrucción del empleo en la forma actualmente conocida, no hay que esperar que las organizaciones privadas, es decir, las empresas privadas, vayan a actuar de una forma diferente a la de aumentar sus ingresos independientemente de la naturaleza que tengan sus trabajadores.
5. El Empresarial, justificado por el continuo crecimiento del mercado de los SAUD (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* (Feldman, 2016), es probable que ese volumen de negocio 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, ya que esto no es más que el principio. 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.
6. El Investigador, incluyendo una serie de medidas relacionadas con la I+D+I en SAUD encaminadas a darle visibilidad y protagonismo en el marco político, aprobando su inclusión y reconocimiento presupuestario, como acción inexcusable.

Además sería necesario:

- Apoyar a los centros científico-tecnológicos especializados en la materia, así como fomentar la creación de nuevos centros, impulsando y desarrollando programas e incentivos para la colaboración público-privada y la transferencia de resultados de investigación que abarquen el ámbito público, el productivo y la sociedad en general.
- Impulsar acciones para minimizar la brecha de género existente en la educación, la investigación y la innovación en el área STEM en general.

- Diseñar programas e incentivos para la formación, retención y captación de talento en el ámbito de los SAUD, apoyando e favoreciendo el emprendimiento y las empresas de base tecnológica.
- Crear un Centro Nacional con la misión esencial de llevar a cabo una investigación de excelencia en el ámbito de los SAUD, así como de velar por el fortalecimiento, apoyo, coordinación y supervisión de la investigación en SAUD y sus aplicaciones, reforzando la confianza en su buen uso.

SAUD EN LA PRÁCTICA

Como es patente, los principales avances en SAUD en la última década han manifestado su capacidad como tecnología de propósito general y han impulsado la innovación en áreas de movilidad, salud, robótica doméstica y de servicios, educación y seguridad cibernética, por nombrar solo algunos.

Los desarrollos habilitados por esos sistemas tienen la capacidad de generar tremendos beneficios, no solo para los individuos sino también para la sociedad en general, y se tiende a que el abanico de aplicaciones vaya en aumento en los próximos años, al igual que ocurrirá con las ganancias que generen.

Los SAUD también tienen capacidades prometedoras a la hora de abordar y resolver desafíos trascendentales, como con el cambio climático o la salud y el bienestar mundial, como se expresa en los objetivos de desarrollo sostenible de las Naciones Unidas. En competencia con otros actores clave, cualquier país necesita aprovechar sus fortalezas, dar facilidades para incrementar la innovación y la adopción de tecnología y encontrar su propuesta única de venta en SAUD, para garantizar una ventaja competitiva y un desarrollo económico próspero a sus habitantes. Pero al mismo tiempo, la aplicación y el uso de los SAUD entrañan riesgos y desafíos asociados con los derechos humanos fundamentales y la Ética.

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 el futuro que le puede esperar de la mano de los SAUD y más concretamente de 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 cuestión es el conocido Dilema del Tranvía, formulado por la filósofa Philippa Foot en un artículo de 1967 (Foot, 1967) 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?

Pero no solo los coches autónomos. El problema se puede plantear en múltiples ámbitos. Pensemos por ejemplo en el conocido Dilema de los trasplantes (Sandel, 2015), según el cual un cirujano tiene cinco pacientes enfermos, que necesitan cinco trasplantes de órganos di-

ferentes o morirán. Pero también tiene un paciente ya recuperado y con órganos sanísimos. ¿Debería “utilizar” a este último para salvar cinco vidas?

La discusión ética sobre cómo resolver este tipo de problema está abierta desde hace mucho tiempo y probablemente así seguirá durante muchos años más. 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. Por eso estos problemas no tienen una solución universal, sino que dependen de cada usuario, de cada fabricante, de cada entorno regional y de variadas circunstancias que pueden, incluso, llevar a que ante una misma situación, un mismo sistema (aunque entrenado por diferentes agentes) produzca respuestas distintas. En consecuencia el debate sobre cómo deberían desempeñarse dichos sistemas autónomos es global, en la misma medida que lo son las premisas que deberían guiar su diseño, construcción y ubicación material, a las que habría que añadir la más importante de todas: que esos sistemas estén contruidos y diseñados sobre sólidas bases teóricas que minimicen los errores que un mal diseño pudiera producir, ya que las disfunciones en su actividad pueden tener consecuencias extraordinariamente importantes.

Todo ello sin menoscabo, paradójicamente, de que el coche autónomo sea necesario por alguna circunstancia, como ha sido el caso del primer dispensador de medicinas que se ha puesto a funcionar en el *Ninth Hospital* de Wuhan (China) a principios de marzo de este año, donde *Little Peanut*, construido por la empresa *Keenon Robotics* (Keenon Robotics, 2020), ha distribuido comida y material sanitario a personas en cuarentena.

Como en el marco de este trabajo es imposible un relato exhaustivo de todos los SAUD que actualmente están en explotación, nos centraremos en dos grandes áreas: la sanitaria y la jurídica, ambas especialmente sensibles en estos momentos en todo el mundo, con la única intención de dejar entrever las inmediatas aplicaciones que vamos a tener a nuestro alcance en el futuro más inmediato. La elección de estas dos áreas no es caprichosa ya que, mientras la primera está bajo el punto de mira de toda la Sociedad por las repercusiones de toda índole que ha traído la pandemia producida por el COVID-19, la segunda no le queda a la zaga por cuanto el logro de acuerdos entre partes o la imposición de penas, castigos y multas por determinados SAUD, cada vez es más frecuente.

En efecto, en lo que concierne a la primera de dichas áreas, es patente que la capacidad de predicción de epidemias es fundamental de cara a salvar vidas, por lo que disponer de SAUD que mejoren la efectividad de las predicciones que hacen los expertos en esa materia es crucial. Pues bien, a finales del pasado año, el SAUD *BlueDot* (Bluedot, 2020) fue capaz de alertar sobre la aparición de una neumonía no convencional en la ciudad china de Wuhan. Ya había actuado con anterioridad en 2014, cuando fue capaz de predecir con seis meses de antelación la llegada del virus *Zika* que produjo la temida alerta sanitaria. El problema que hubo con la alarma del COVID-19 fue que *BlueDot* la valoró con 3, en una escala de 1 a 5, sin apreciar completamente la importancia que iba a tener. No obstante, la OMS rectificó esa valoración al poco tiempo, recalificándola a nivel 5.

¿Superioridad de las personas sobre los SAUD? Más bien complementariedad, porque si *BlueDot* no hubiera detectado ese brote epidémico, quizás la reacción de la comunidad internacional hubiera llegado más tarde. La clave está en el desempeño de *BlueDot*, algo prohibido para las personas, ya que utiliza *Big data*, es decir, utiliza algoritmos de procesamiento del lenguaje na-

tural y de aprendizaje automático para seleccionar datos de cientos de miles de fuentes, incluidas declaraciones de organizaciones oficiales de salud pública, medios digitales, datos de billetes de aerolíneas, informes sobre la salud del ganado o datos demográficos de la población, siendo capaz de procesar rápidamente toneladas de información cada 15 minutos, las 24 horas del día.

Pero *BlueDot* no es el único SAUD que alertó sobre esa epidemia. También actuó en esa dirección *Metabiota* (Metabiota, 2020), que desde enero de 2020 está rastreando y analizando este brote de varias maneras, incluso a través de un sistema de vigilancia digital fiel a la fuente y un modelo mundial de propagación de enfermedades. De hecho *Metabiota* ha construido un modelo de pronóstico de propagación de la enfermedad a corto plazo, que incorpora las características conocidas actuales del virus y que predijo el 25 de febrero los países que se verían afectados en la semana siguiente: China, Italia, Irán y EE. UU.

El funcionamiento de *BlueDot* y *Metabiota* se apoya en la búsqueda de palabras claves en redes sociales y medios de comunicación, lo que provoca que a veces las decisiones que toman (en estos casos en términos de predicciones) no sean tan robustas como debieran o quisiéramos. Esta es la razón fundamental que recomienda que el uso de SAUD como estos vaya siempre acompañada de la supervisión de los expertos humanos en las respectivas áreas, de manera que la actuación conjunta de ambos agentes, la complementariedad de ambos como decíamos más arriba, produzca resultados mejores que los correspondientes por cada uno de ellos por separado.

Por otro lado, es evidente que la revolución digital ha llegado al terreno jurídico. Si la incorporación de la tecnología puede ser considerada como el último movimiento en la mejora del acceso a la justicia, este movimiento está llegando con olas sucesivas. Cuando todavía no se ha acabado de asimilar el potencial en la resolución de litigios asociado a las TIC, los últimos desarrollos en I.A. alimentan nuevas promesas para la mejora de la eficiencia y la calidad. Desde la creación de motores potentes para la búsqueda de decisiones se ha dado el paso hacia herramientas sorprendentes que, por ejemplo, ofrecen asesoramiento legal a las partes por medio de un robot, incluso con voz humana, permiten resolver un litigio sin contar con la intervención humana (ODR), predicen el resultado de un litigio o permiten identificar en los jueces perfiles más inclinados hacia determinadas soluciones. Todo ello facilitando, además, salvar la distancia geográfica, permitiendo la interacción en línea, las personaciones sin representación legal, la reducción de costes y la duración de los procedimientos. Así han aparecido SAUD que han hecho necesario plantear un completo rediseño del proceso tradicional.

El primer ejemplo se puso en funcionamiento en 2016 en el estado Columbia Británica, en Canadá, con la puesta en marcha del Tribunal de Resolución Civil (CRT por sus siglas en inglés). El CRT (CRT, 2020) es el primer tribunal *en línea* de Canadá, y es uno de los primeros ejemplos en el mundo de resolución autónoma de disputas en línea que se incorpora al sistema de justicia pública, resolviendo principalmente disputas:

- Derivadas de accidentes automovilísticos, por ejemplo, para compensar posibles daños físicos o materiales, para reparar fallos mecánicos o para determinar si una lesión es una “lesión menor” (todo con un tope de 50.000 USD, a partir del cual el accidente pasa a, digamos, una resolución tradicional, es decir, presencial y con asistencia personal).

- Producidas por pequeñas compras y ventas de bienes y servicios, alquileres de viviendas y reclamaciones sobre compañeros de alquiler, cargo-abonos de recibos y débitos de diversa índole o reparaciones caseras de poca entidad (fontanería, electricidad...).
- Asociadas a problemas en condominios con otros socios y causadas por interpretaciones distintas de actas, reglamentos, impagos, etc.

También en los Estados Unidos de América algunos Estados están experimentando con un *software* llamado *Matterhorn* (Matterhorn, 2020). De hecho más de 70 tribunales, en juzgados de primera instancia de diferente dimensión en diez estados, confían en la resolución de disputas en línea de Matterhorn, que se utiliza para resolver una amplia variedad de casos. Así mismo el Estado de Nueva York, junto con la *Judicial Division* de la *American Bar Association*, está desarrollando un proyecto piloto de tribunal virtual o en línea también para reclamaciones dinerarias (Larson, 2019) y en Australia (Tan, 2019) en septiembre de 2018 se implantó, como proyecto piloto, el *Victorian Civil and Administrative Tribunal* (VCAT) como una nueva estrategia digital para mejorar los servicios en línea y mejorar la eficiencia de los mismos. El VCAT implementó un prototipo, no exactamente un SAUD, pero sí un importante primer paso hacia su construcción definitiva, para la resolución de disputas, mediante el cual las partes pudieron hacer que sus asuntos fueran escuchados y resueltos por un miembro de VCAT en tiempo real utilizando la tecnología de video y uso compartido de archivos en sus propios dispositivos como alternativa a asistir a una audiencia en persona.

Con esta base, en la actualidad, el autor de este trabajo colabora en un proyecto (COTEC, 2020) financiado por la Fundación COTEC dirigido a generar conocimiento para identificar elementos y claves para la maduración de un debate y diálogo sobre los principios éticos a que ha de ser sometido el diseño, construcción y utilización de los SAUD para la resolución de litigios, con miras a conseguir que la mejora de la eficiencia y de la calidad pueda tener lugar sin detrimento de derechos fundamentales.

CONCLUSIONES

Todo lo descrito en este artículo justifica sobradamente la elaboración de planes estratégicos que planifiquen el desarrollo de los SAUD, a la vez que maximicen los beneficios de su uso y minimicen sus riesgos. Son muchas las acciones que habría que poner en marcha, pero entre ellas, sin ningún orden de prioridad, tendrían carácter necesario:

1. Elaborar un acuerdo mundial para la de defensa ante posibles consecuencias indeseadas del uso de los SAUD.
2. De acuerdo con la normativa de cada país, garantizar el derecho de las personas físicas y jurídicas a la intimidad y seguridad sobre los datos.
3. Crear en cada sector que sea susceptible de aplicación de los SAUD una Comisión de Garantías Éticas sobre su uso.
4. Incrementar la formación sobre los SAUD en todos los ámbitos sociales y niveles educativos, como garantía de conocimiento de riesgos y oportunidades.
5. Fomentar e impulsar nuevas acciones contra la brecha de género.

AGRADECIMIENTOS

Trabajo financiado por los proyectos *TIN2017-86647-P* (Fondos FEDER de la UE) y *Justicia en la era digital: Claves y elementos para una nueva arquitectura* (Fundación COTEC, 2020).

REFERENCIAS

- Agencia Digital for Italy. (2018). *White Paper on Artificial Intelligence at the service of citizens*. AGID. Obtenido de <https://ia.italia.it/assets/whitepaper.pdf>
- AI/Robotics researches. (2017). ASILOMAR AI PRINCIPLES. Obtenido de Future of life: <https://futureoflife.org/ai-principles/>
- Bluedot. (2020). *BlueDot*. Obtenido de <https://bluedot.global/>
- COTEC. (2020). *Proyectos*. Obtenido de Fundación COTEC: <https://cotec.es/proyectos/>
- Council, S. (2017). *New Generation of Artificial Intelligence Development Plan*. China: The Foundation for Law and International Affairs. Obtenido de <https://flia.org/wp-content/uploads/2017/07/A-New-Generation-of-Artificial-Intelligence-Development-Plan-1.pdf>
- CRT. (2020). *Civil Resolution Tribunal*. Obtenido de <https://civilresolutionbc.ca/>
- Department of Business, Energy & Industrial Strategy - Department for Digital, Culture, Media & Sport. (21 de May de 2019). *AI Sector Deal - Policy Paper*. Obtenido de GOV.UK: <https://www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal>
- Dutton, T. (28 de June de 2018). *An Overview of National AI Strategies*. Obtenido de Politics-AI: <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>
- European Commission. (2018). *Statement on A.I., Robotics and 'Autonomous' Systems*. European Group on Ethics in Science and New Technologies (EGE). Directorate-General for Research and Innovation, Unit RTD.01. Bruselas: European Union. Obtenido de Statement of AI Robotics: https://ec.europa.eu/info/research-and-innovation/strategy/support-policy-making/scientific-support-eu-policies/ege_en
- European Commission. (8 de April de 2019). *Ethic Guidelines for trustworthy AI*. Obtenido de Shaping Europe's digital future: <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>
- Feldman, M. (30 de August de 2016). *Market for Artificial Intelligence Projected to Hit \$36 Billion by 2025*. Obtenido de Top500: <https://www.top500.org/news/market-for-artificial-intelligence-projected-to-hit-36-billion-by-2025/>
- Foot, P. (1967). *The Problem of Abortion and the Doctrine of Double Effect. ... 1978. Virtues and Vices*. Oxford: Basil Blackwell.
- Harhoff, D., Heumann, S., Jentsch, N., & Lorenz, P. (2018). *Outline for a German Strategy for Artificial Intelligence*. Berlin: Beisheim Center.
- Huawei. (2017 de diciembre de 2017). *Huawei analiza la toma de decisiones de los europeos*. Obtenido de Tribuna Digital China: <https://www.tribunadigitalchina.com/huawei-analiza-la-toma-de-decisiones-de-los-europeos-con72965>

- Keenon Robotics. (2020). *Productos Robots Keenon*. Obtenido de <http://keenon.com/Product/pro2.html>
- Larson, D. A. (2019). Designing and Implementing a State Court ODR System: From Disappointment to Celebration. *Journal of Dispute Resolution* (2), 77-102.
- Larus, J., Hankin, C., Carson, S. G., Christen, M., Crafa, S., Grau, O., & al, e. (2018). *When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making*. Informatics Europe & EU ACM .
- Matterhorn. (2020). *Matterhorn online dispute resolution (ODR)*. Obtenido de ODR Solutions: <https://getmatterhorn.com/>
- Metabiota. (3 de February de 2020). *Monitoring and risk analytics for the 2019 novel coronavirus (COVID-19) epidemic*. Obtenido de Metabiota Risk Report No. 3: https://metabiota.com/sites/default/files/inline-files/Metabiota_Risk_Report_No.3-25Feb2020-COVID-2019_0.pdf
- Ministerio de Ciencia, Innovación y Universidades. (2019). *Estrategia Española de I+D+I en Inteligencia Artificial*. Secretaría General Técnica del Ministerio de Ciencia, Innovación y Universidades.
- National Science and Technology Council. (2016). THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN. U.S. Government. Obtenido de https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf
- Niti Aayog. (2018). *National Strategy for Artificial Intelligence- Discussion Paper*. Obtenido de https://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf
- Sandel, M. (2015). *M. Sandel. Justice: What's The Right Thing To Do? Episode 01 "THE MORAL SIDE OF MURDER*. Obtenido de Lecture of Harvard University: <https://www.youtube.com/watch?v=kBdfcR-8hEY>
- Steering Group of the Artificial Intelligence Programme. (2017). *Finland's Age of Artificial Intelligence*. Helsinki: Ministry of Economic Affairs and Employment-MEAE reports 47/2017. Obtenido de http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf.
- Thaler, R.H. and Sunstein, C.R. (2008): *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press.
- Tan, V. (August de 2019). Online Dispute Resolution for Small Civil Claims in Victoria: A New Paradigm in Civil Justice. *Deakin Law Review*, 24, 01-138. doi:<https://doi.org/10.21153/dlr2019vol24no1art873>
- Villani, C. (2018). *For a Meaningful Artificial Intelligence*. aiforhumanity.fr

