

**VIVIR CON ROBOTS.
REFLEXIONES ÉTICAS,
JURÍDICAS, SOCIALES Y
CULTURALES**

***LIVING WITH ROBOTS. ETHICAL,
LEGAL, SOCIAL, AND CULTURAL
REFLECTIONS***

Mario Toboso Martín

Consejo Superior de Investigaciones Científicas
<http://orcid.org/0000-0001-9284-3643>
mario.toboso@csic.es

María Amparo Grau Ruiz

Universidad Complutense de Madrid
<http://orcid.org/0000-0002-0482-2816>
grauruiz@ucm.es

Cómo citar este artículo/Citation: Toboso Martín, Mario; Grau Ruiz, María Amparo (2021). Vivir con robots. Reflexiones éticas, jurídicas, sociales y culturales. *Arbor*, 197(802): a623. <https://doi.org/10.3989/arbor.2021.802001>

Copyright: © 2021 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia de uso y distribución *Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0)*.

ROBÓTICA E INTELIGENCIA ARTIFICIAL: ROBÓTICA INTERACTIVA

La robótica estudia el diseño y la construcción material de robots. Junto con las interfaces para la interacción virtual entre humanos y máquinas, las redes neuronales o los sistemas multiagente constituyen en la actualidad tecnologías disruptivas que se desarrollan rápidamente, y cuyo potencial de transformación del ser humano, de la naturaleza y de la sociedad es enorme. La inteligencia artificial (IA) y la robótica son tecnologías digitales que, como ha señalado Vincent Müller (2020), en un futuro próximo tendrán un impacto significativo en el desarrollo de la humanidad. Traen a primer plano cuestiones importantes acerca de qué deberíamos hacer con estos sistemas, qué deberían poder hacer, o no hacer, esos mismos sistemas, sobre los riesgos implicados en nuestra convivencia con ellos y sobre la manera de gestionarlos.

Desde las primeras conceptualizaciones imaginarias, antes de que la tecnología los hiciera materialmente posibles, los robots nos han preocupado, porque nunca hemos tenido claro si nos servirían realmente de ayuda o supondrían en algún momento una amenaza. La posibilidad de crear vida e IA nos confronta con nosotros mismos. En la robótica, la figura de Frankenstein sirve de arquetipo de la ciencia ficción distópica, en la que la vida artificial creada por el ser humano llega a convertirse en un peligro. Especialmente la ciencia ficción de la primera mitad del siglo XX compartía esta visión distópica, como en la película *Metrópolis* de Fritz Lang o en la famosa obra de teatro *R.U.R. Rossum's Universal Robots* de Karel Capek, estrenada en 1920, en la que apareció por primera vez la palabra *robot* (INBOTS, 2019b: 35)¹. Los robots nos colocan ante un espejo en el que mirarnos y constatar nuestras propias luces y sombras. Guardan una dialéctica prometeica y al mismo tiempo de caja de Pandora.

Al igual que otras tecnologías, los robots están culturalmente incardinados. En consecuencia, en la interacción entre humanos y robots (*Human-Robot Interaction*, HRI), la aceptación social debe desempeñar

un papel fundamental. La confianza social se convierte en un elemento clave. Las personas llegan a aceptar y a confiar en las nuevas tecnologías cuando, de una manera pragmática, las evalúan favorablemente al satisfacer una necesidad determinada o un conjunto de necesidades. Esta evaluación práctica dependerá de valores que pueden ser de naturaleza muy diversa: técnicos, instrumentales, económicos, culturales, sociales, éticos, o incluso religiosos; pero también dependerá de las habilidades funcionales para su uso, de la educación o alfabetización tecnológica de las personas usuarias, de sus expectativas sobre la tecnología en cuestión y de las actitudes, favorables o desfavorables, hacia la misma (INBOTS, 2019b: 35, 38)

El desarrollo de una robótica inteligente no solo implica incorporar técnicas de IA a los robots para que se relacionen de la manera óptima entre ellos y con los humanos. La inteligencia a la que debe apelarse en mayor medida en el proceso actualmente en curso, en el que estamos inmersos, es la inteligencia verdaderamente humana. La apuesta decidida por comprender las características y la magnitud de este proceso y poder hacer propuestas con sentido para resolver los problemas que pueda traer consigo requiere conocer y compartir experiencias para fomentar una apropiación social del conocimiento y de las prácticas que puedan ser más provechosas al respecto.

Mientras que la IA puede consistir únicamente en software, los robots son máquinas físicas que se mueven, exploran e intervienen sobre el entorno mediante conjuntos de sensores y actuadores (Müller, 2020). Los robots interactivos son robots que interactúan muy estrechamente con las personas. De manera más precisa: son sistemas artificiales, dotados de un cuerpo físico y de elementos de IA, que se sitúan y operan en entornos de actividades humanas, en condiciones de proximidad muy estrecha con las personas (INBOTS, 2019b: 8). Consideradas como un campo en pleno desarrollo científico y tecnológico, las investigaciones sobre la interacción humano-robot prestan una atención notable a las cuestiones éticas, a los estudios de percepción y aceptación, a las dinámicas

1 INBOTS es el acrónimo del proyecto europeo de investigación Inclusive Robotics for a Better Society (CSA H2020, 780073). El objetivo general de este proyecto es crear un espacio de investigación que reúna a expertos para debatir y establecer un paradigma de investigación e innovación responsable para la robótica interactiva, que potencie el liderazgo de la Unión Europea en este campo. Para ello, el proyecto coordina y apoya las acciones dirigidas a tender puentes entre los siguientes cuatro ámbitos de expertise, que abarcan a todas las partes interesadas en la robótica interactiva: el ámbito técnico, el empresarial, el ámbito ético, jurídico y socioeconómico, y el ámbito de los usuarios finales, de los responsables políticos y del público en general. El proyecto se ha llevado a cabo por un consorcio de veinticinco entidades participantes, de doce países europeos. Algunas de las contribuciones incluidas en este monográfico son el resultado de la participación de sus autoras y autores en este proyecto.

de comunicación e interacción funcional entre ambas partes y a los valores e intereses implicados en esta peculiar relación social entre humanos y robots, incluyendo los aspectos que tienen que ver con su trabajo colaborativo (Arnold y Scheutz, 2017; Müller, 2020)

Los robots interactivos no son meros instrumentos. Constituyen mediaciones que surgen de un contexto sociocultural y sociotecnológico determinado, que ofrecen nuevas posibilidades para el funcionamiento humano. En este sentido, puede decirse que, al igual que otras tecnologías, no solo transforman la materia, sino también nuestros hábitos, costumbres y relaciones. Por ello, si se entiende la robótica interactiva de una manera sistémica, será difícil dudar de su capacidad transformadora sobre los sistemas, ya sean estos sociales o naturales (O'Brolchain y Grau Ruiz, 2020). Esta capacidad genera conflictos de valores asociados al proceso de su desarrollo tecnológico e implantación social. Cuando las innovaciones tecnológicas inciden marcadamente en la sociedad, las controversias derivadas suelen plantearse en términos de la dicotomía entre actitudes tecnófilas y tecnófobas (INBOTS, 2019a: 118). En el caso particular de la robótica interactiva, los contextos sociales en los que esta se aplique determinarán, delimitarán y harán más complejo el alcance del conflicto axiológico mencionado, creando, por ejemplo, controversias entre subsistemas de valores en ámbitos de aplicación como el asistencial, el sanitario o el militar.

Al igual que cualquier otro artefacto tecnológico, los robots interactivos incorporan valores procedentes de los diferentes agentes y comunidades implicadas en su diseño, fabricación, comercialización, uso, etc. Cada comunidad particular puede tener su propio discurso acerca de los posibles beneficios o perjuicios de tales robots. La confrontación entre los diferentes discursos mediante debates públicos abiertos debería constituirse en una condición necesaria (UNESCO, 2017; European Group on Ethics, 2018) para que la introducción de la robótica interactiva en las sociedades llegue a favorecer realmente el bienestar y la justicia social (INBOTS, 2019a).

DESAFÍOS QUE PLANTEA LA IMPLANTACIÓN SOCIAL DE LA ROBÓTICA INTERACTIVA

Hoy en día, algunos sectores críticos para el desarrollo económico cuentan con sistemas específicos de gobernanza. Por ejemplo, en el sector bancario, tradicionalmente han supervisado el cumplimiento de normas especiales sobre contabilidad, seguridad

y estabilidad económica los Bancos Nacionales o el Banco Central Europeo. Por otro lado, la situación de vulnerabilidad generalizada que puede ocasionar su quiebra a particulares y empresas hace que las entidades bancarias se sometan periódicamente a pruebas de *stress*, simulándose diferentes escenarios económicos para conocer la situación hipotética en la que se encontrarían. La inyección de capital acaba siendo la herramienta empleada más comúnmente en la práctica para tratar de resolver males hasta cierto punto impensables *a priori* o inevitables.

La previsible dependencia creciente de la robótica para el ejercicio de actividades económicas y para el desenvolvimiento normal de la vida de las personas en sociedades tecnológicamente desarrolladas debería hacernos pensar si, en esa tesitura de posible vulnerabilidad, cabría arbitrar algún sistema de gobernanza similar al bancario, reforzándolo y haciendo acopio de caudales mínimamente suficientes para el caso de eventuales desastres. En este orden de cosas, no son de extrañar algunas propuestas sobre el registro de robots, que llevaría aparejado el pago de las respectivas pólizas de seguro y la creación de fondos para abordar posibles contingencias.

De manera alternativa, se podría pensar en tomar como referencia otros sistemas vigentes en la Unión Europea que, de hecho, ya influyen de manera más o menos directa en campos afines. En concreto, interesaría analizar los sistemas de gobernanza aplicables en el mundo de los datos, por estar muy estrechamente ligados al funcionamiento de los robots interactivos en los que opera la IA. Así mismo, por el impacto que la robotización tiene en la competitividad, en tanto en cuanto el uso inadecuado de la financiación pública puede alterar la competencia en los mercados, sería oportuno prestar atención a los sistemas de gobernanza en esta materia, por sus efectos en la desigualdad interpersonal e interterritorial.

Lejos ya de ser los protagonistas abstractos de mitos o del folclore popular, de la literatura de ciencia ficción o del cine, vemos, pues, cómo los robots se vienen adentrando en importantes ámbitos prácticos, y su repercusión en campos de conocimiento como el derecho y la economía, la ciencia de datos, la sociología o la filosofía. La robótica interactiva ofrece también posibilidades prometedoras en ámbitos como la educación, la medicina, los cuidados, la convivencia o la administración, pero surgen, igualmente, importantes interrogantes éticos, jurídicos, políticos, sociales, económicos y culturales. Desde el punto de vista filo-

sófico y jurídico ¿debemos concebir a los robots (incluidos los que se dotan de IA) como sujetos morales o agentes de pleno derecho? ¿Es aceptable, sostenible, este enfoque? ¿Existen aproximaciones alternativas? ¿Qué consecuencias tendrían estos planteamientos en cuestiones jurídicas tan importantes como la atribución de responsabilidad? ¿Cuáles son las consecuencias del auge de la robótica interactiva para el derecho laboral, para el futuro del trabajo, para los mecanismos de protección de la propiedad industrial e intelectual? ¿Qué elementos del derecho fiscal y de la responsabilidad social de las empresas deberían revisarse en este contexto? (INBOTS, 2019a).

En todos estos casos, el peso específico de cada factor dependerá de la voluntad humana, ya sea esta manifestada formalmente en el plano colectivo o a través del margen del libre albedrío individual. En la actualidad, se hace patente, más que nunca, la necesidad de detectar alternativas, de debatir las opiniones encontradas y de generar consensos para apostar por soluciones suficientemente equilibradas en un entorno globalizado. En consecuencia, una aproximación razonable y sensata a estos temas requerirá de un diálogo continuado de naturaleza interdisciplinar e intercultural.

Si el motor de búsqueda de Google, la publicidad segmentada de Facebook o el funcionamiento de asistentes virtuales como Alexa y Siri son posibles gracias a la potencia actual de los algoritmos de IA, presentes igualmente en los vehículos autoconducidos, en la actividad policial predictiva o preventiva, y en las armas autónomas que permiten ejecutar acciones letales sin la necesidad de la intervención humana, no se deben perder de vista las complejas implicaciones éticas inherentes a estas y otras aplicaciones (Coeckelbergh, 2020). Muchas de estas implicaciones se relacionan con la privacidad, con la responsabilidad y con la delegación o sustitución en la toma de decisiones mediada por algoritmos, ligada a un cuestionable ideal de transparencia acompañado de insalvables sesgos de todo tipo, asociados a la calidad de los datos. El avance, no solo de la IA sino también de las tecnologías robóticas arroja, pues, nuevas preocupaciones éticas acuciantes: si acaso fuese posible, ¿deberían programarse los robots para seguir códigos éticos?; si existiese la posibilidad de

crear vínculos emocionales con robots, ¿cuáles serían los riesgos implicados, y cómo gestionarlos?; ¿podría la robótica llegar a transformar, no solo la sociedad, sino también el significado de la ética y el carácter de sus prácticas? (Lin, Abney y Beckey, 2011).

A medida que las tecnologías robóticas se vuelven más autónomas, es necesario identificar los valores y principios éticos que deberían regular su interacción con las personas. Si toda tecnología es el resultado de las decisiones de una pluralidad de agentes, que hacen que esta llegue a ser lo que finalmente es, y estas decisiones implican valores muy diversos, entre los cuales suelen prevalecer los valores técnicos y económicos, ¿no debería ampliarse el marco de valores en tales decisiones para llegar a involucrar a la mayor representación posible de agentes sociales: usuarios, consumidores, comunicadores, políticos, responsables de los productos tecnológicos, grupos sociales afectados por ellos...? (INBOTS, 2019a: 118).

La presencia de la ética en este campo de investigación tecnológica se da a través de dos vías. Por un lado, la denominada «ética de los robots», que se refiere a los problemas éticos que el uso de los robots plantea para las personas y los entornos de actividades humanas. Por otro lado, las preguntas por el estatus ético de los robots y por la posibilidad de dotarles de cualidades morales remite a la denominada «ética de las máquinas». La ética de la IA y de la robótica representan un ámbito incipiente dentro del terreno de la ética aplicada, con un dinamismo apreciable que llega a ofrecer cuestiones muy relevantes pero, por el momento, deficitariamente planteadas, por carecer de marcos generales bien establecidos que permitan abordarlas de una manera integral (Müller, 2020)².

La mirada se pone a menudo en los sectores particulares en los que la confluencia de la robótica y la IA puede tener previsiblemente un mayor impacto, como las manufacturas, la medicina, la agricultura, la logística y el transporte o el consumo (Comisión Europea, 2014), confiándose en los mecanismos de supervisión preexistentes en cada uno de ellos. Así viene sucediendo hasta la fecha. Sin embargo, el intento de fijar la mirada, de manera introspectiva, en el propio

2 Müller señala, sin embargo, excepciones, como la labor del European Group on Ethics in Science and New Technologies (2018), algunas investigaciones destacadas acerca del impacto social de estas tecnologías (Floridi et al., 2018; Taddeo y Floridi 2018; Taylor et al., 2018; Walsh, 2018; Bryson, 2019; Gibert, 2019; Whittlestone et al., 2019), y el trabajo experto llevado a cabo en espacios en los que se generan, finalmente, recomendaciones políticas, como el High-Level Expert Group on Artificial Intelligence, dependiente de la Comisión Europea (AI HLEG, 2019) o el Institute of Electrical and Electronics Engineers (IEEE, 2019).

sector de la robótica inteligente, podría resultar enormemente útil. Al menos en lo que a la relativa invisibilidad de su análisis sectorial respecta, pues es fácil observar que habitualmente se ha pasado por alto.

En este punto, parece obvio que los resultados e impactos (negativos y positivos) de los procesos de robotización deban analizarse; pero, simultáneamente, ello no exige de analizar también los procesos en sí mismos. Por eso, más allá del ritmo de implantación de los robots y de su adecuado funcionamiento, en el fondo, debe preocuparnos no solo el hecho de que no dañen a otros (como desde el derecho romano se postulaba (*neminem laedere* o *alterum non laedere*), sino también atender a que se aproveche la oportunidad de utilizarlos para el logro del bien común, sin desperdiciar su potencial en este sentido.

Uno de los principales retos en la investigación, el desarrollo e implantación social de nuevas tecnologías como la robótica es la desconfianza e incluso el rechazo de la sociedad. El compromiso público es esencial para que la investigación alcance las mejores soluciones posibles en cuestiones éticas delicadas, y para garantizar la aceptación y la confianza del público en dicha investigación. El enfoque de *Investigación e Innovación Responsables (Responsible Research and Innovation, RRI)* tiene como objetivo hacer que la sociedad en su conjunto participe de la ciencia y la innovación desde sus fases iniciales, involucrando a todos los actores posibles, y alineando los procesos y los resultados de la investigación con los valores, necesidades y expectativas sociales. De acuerdo con la RRI, se deben utilizar metodologías inclusivas y participativas en todas las etapas de los procesos de investigación e innovación. Un elemento importante para avanzar en el camino hacia una robótica realmente inclusiva será identificar y analizar los mecanismos de evaluación que se dan a lo largo del proceso de innovación de la robótica interactiva, que se orientan a favorecer la satisfacción de valores sociales relevantes, como, en este caso, la inclusión (INBOTS, 2019a).

Desde una perspectiva interdisciplinar, característica de los estudios de Ciencia, Tecnología y Sociedad (CTS), se han elaborado las siguientes conclusiones acerca de la hoja de ruta para la implantación social de la robótica interactiva (INBOTS, 2019b: 8): 1) La aceptación de los robots interactivos no puede producirse de una manera aislada del contexto social. 2) El significado práctico y el valor de las tecnologías, como los robots interactivos, no son predecibles independientemente del contexto social en el que se

pretenden implementar. 3) La aceptación potencial de los robots interactivos no se puede separar de desarrollos tecnológicos anteriores, ya que no se trata de un desarrollo tecnológico rupturista, sino evolutivo. Desde la perspectiva de las ciencias sociales, el desarrollo actual de los robots interactivos puede considerarse como la continuación de una automatización en curso, que se remonta a varias décadas atrás y en la que los ciudadanos han ido aprendiendo y aceptando gradualmente nuevas formas de comunicación e interacción mediadas por diversas tecnologías.

En este número monográfico se tratan algunos de los retos principales que plantea este nuevo escenario tecnológico. Entre otras cuestiones: cómo desarrollar unas tecnologías robóticas inclusivas, sostenibles y socialmente responsables; qué consecuencias tendrán sobre el mercado laboral o la legislación; qué impacto tendrán el aprendizaje automático y el aprendizaje profundo en los procesos de toma de decisión mediante algoritmos en ámbitos tan importantes como el empleo, la medicina, la justicia o el acceso a derechos humanos fundamentales; qué papel jugarán los valores, las creencias, las emociones, el lenguaje y los sesgos cognitivos en esos procesos algorítmicos de aprendizaje y decisión; de qué manera estas tecnologías, cada vez menos emergentes y más presentes, van a afectar, en definitiva, al funcionamiento de los sistemas sociales.

LA GOBERNANZA DE LA ROBÓTICA INTERACTIVA

Ante esta situación de cambio tecnológico, conviene plantearse qué esquemas de gobernanza son deseables y factibles para garantizar el bienestar de las personas. A la hora de proteger los derechos humanos, sociales y políticos, es absolutamente necesario establecer sistemas de control, principalmente en la toma de decisiones o al desarrollar determinadas actividades en materia sensible. Cómo arbitrar la gobernanza del modo más eficaz es perfectamente discutible, pues puede fiarse al autocontrol, crearse un sistema novedoso o tratar de combinar otros ya existentes que puedan ser relevantes, teniendo alguna conexión más o menos directa. En cualquier caso, pensamos que queda fuera de toda duda que la supervisión ha de ser, en última instancia, esencialmente humana. Conviene, sin embargo, afrontar una importante restricción de partida: el sistema de gobernanza que se diseñe ha de ser ciertamente eficaz y suficientemente garantista, pero no debería frenar las ventajas que la innovación en la robótica pueda traer consigo.

En un primer estadio, es lógico que el derecho sea fundamentalmente de corte reactivo con respecto a los posibles inconvenientes que se vayan descubriendo. Así, suele inicialmente atenderse a los intereses de las partes implicadas en algunos supuestos de hecho con normas sobre responsabilidad y seguros, e incluso sobre protección de datos ante posibles violaciones del ordenamiento jurídico, ofreciendo niveles de protección variable en función de las circunstancias del caso concreto. Con frecuencia, la idea de remediar, o en su defecto paliar o compensar el daño, procurando los fondos necesarios es la principal solución en la que se inspira el legislador, anticipando para ello una estimación cuantitativa. La forma de reunir las cuantías monetarias precisas variará según el riesgo y el impacto esperado.

En un momento posterior, con un planteamiento más maduro, quizás cabría dar un paso más y orientar el derecho de forma proactiva. Esto supondría articular medidas de fomento que incentivasen los comportamientos que la sociedad valore como útiles por reportar beneficios no solo de orden económico, sino también medioambientales, de buena gobernanza y sociales. Hacia este punto, pensamos, debería avanzar cualquier regulación futura de la robótica, enfilando hacia la dirección trazada mediante la adopción de las diferentes medidas normativas y, en su caso, el margen dejado a la autorregulación. La posibilidad de adaptación continua habría de incorporarse al sistema de gobernanza para corregir, en su caso, el rumbo eficazmente. Una buena regulación redundará en una mayor seguridad jurídica y, por ende, en una mayor aceptación social de la robótica.

Para diseñar un sistema de gobernanza puede contarse, de entrada, con instituciones privadas o públicas (o de ambos tipos), con esquemas centralizados o descentralizados, teniendo en cuenta el número de niveles a la hora de valorar los costes (económicos, de tiempo, de complejidad en las operaciones) y las ventajas que supongan. Sería conveniente optimizar la creación de instituciones, limitándolas a las estrictamente necesarias según sus objetivos y el oportuno reparto competencial (conforme al territorio o la materia, evitando solapamientos que puedan dar lugar a conflictos). Deberían aprovecharse, además, las sinergias que puedan surgir al considerar enfoques cooperativos con otros organismos. La ambición del sistema de gobernanza habría de ser alcanzar al mayor número posible de países y empresas, sin perder efectividad. Igualmente importa señalar las tareas que se asignan al organismo

en cuestión en el marco de gobernanza (por ejemplo, relativas a la fiscalización, decisión o ejecución).

Entre las normas europeas que ya han establecido algunos sistemas de gobernanza de posible interés para el sector de la robótica (la inteligencia artificial y los datos masivos) figuran las siguientes:

- Reglamento (CE) No. 1060/2009 del Parlamento europeo y del Consejo de 16 de septiembre de 2009 sobre las agencias de calificación crediticia.
- Reglamento (UE) No. 1095/2010 del Parlamento europeo y del Consejo de 24 de noviembre de 2010 por el que se crea una Autoridad Europea de Supervisión (Autoridad Europea de Valores y Mercados), se modifica la Decisión 716/2009/CE y se deroga la Decisión 2009/77/CE de la Comisión.
- Reglamento (UE) 2016/679 del Parlamento europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE.
- Reglamento (UE) 2017/2394 del Parlamento europeo y del Consejo de 12 de diciembre de 2017 sobre la cooperación entre las autoridades nacionales responsables de la aplicación de la legislación en materia de protección de los consumidores y por el que se deroga el Reglamento (CE) No. 2006/2004.

A la vista de su contenido, dependiendo del grado de compromiso de los Estados miembros, bien podría establecerse una agencia europea, o bien contar con una red de autoridades estatales coordinada por la Unión Europea. En el primer caso, la adopción de decisiones podría hacerse recaer en un Comité de Responsables Europeos, del que formarían parte las autoridades nacionales (con derecho a voto) junto con un representante de la Comisión Europea (sin derecho a voto). Debería también aclararse qué decisiones requerirían ser aprobadas por mayoría simple o cualificada. En el segundo caso, podría establecerse una red de cooperación entre autoridades nacionales coordinadas a nivel europeo, que podría apoyarse en la investigación y ejecución en las estructuras nacionales ya disponibles. La Comisión Europea pondría en conocimiento de la red los casos en los que estimase pertinente su actuación. Para la fijación de la autoridad competente podría tenerse en cuenta la sede

principal de la empresa afectada (con el riesgo de que se frecuentaran los lugares con estándares más bajos) o permitir que las autoridades implicadas eligieran la nacional mejor posicionada para coordinar el caso, preparar un borrador de decisión y discutirlo con las restantes en la búsqueda de consenso, previendo mecanismos de desbloqueo en su caso por un comité europeo. Debería, especificarse el alcance territorial de los efectos de la decisión adoptada

Otros modelos de gobernanza híbridos pueden idearse como fruto de combinaciones en el reparto de tareas entre las autoridades europeas y nacionales. Deberían sopesarse, en cualquier caso, los cambios organizativos que comporte el sistema a aplicar, evitando duplicar estructuras en cada Estado y que se incurra en costes innecesarios en momentos de recuperación económica tras la crisis provocada por la pandemia COVID-19. La gobernanza ha de responder al criterio de centralización económicamente eficiente sin menoscabar la descentralización jurídica necesaria, de acuerdo con el principio de subsidiariedad que ha de guiar el grado de la intervención de las instituciones europeas³. Es pues el momento oportuno para reflexionar sobre la Resolución del Parlamento Europeo de 20 de octubre de 2020⁴ que contiene propuestas en torno a las autoridades supervisoras del cumplimiento de principios éticos y obligaciones jurídicas. La Comisión Europea la ha considerado en la formulación de su Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de inteligencia artificial) y se modifican determinados actos legislativos de la Unión⁵, que ya ha sido transmitida al Consejo de la Unión Europea⁶.

Con independencia de la forma particular en la que se esboce el sistema de gobernanza, conviene resaltar que cuando cualquier debate se centra excesivamente en la autonomía del robot, en cierto sentido,

se desenfoca. Lo esencial es la autonomía humana a la que este tipo de herramienta técnica debe servir. Finalmente, su impacto habrá de juzgarse no solo desde una perspectiva meramente individualista, sino siendo también considerada desde la preocupación de la sociedad en su conjunto.

CONTRIBUCIONES A ESTE MONOGRÁFICO

En la contribución que abre este monográfico, Daniel López Castro ilustra la presencia creciente de la robótica en nuestras vidas como resultado de una transición, en curso, desde la robótica industrial, operante en entornos cerrados y predecibles, hacia la robótica interactiva que opera en entornos abiertos y complejos, en los que el ser humano pasa a estar presente. Esta transición, de lo industrial a lo personal, pero también a lo social, amplía el alcance de la investigación tradicional a múltiples disciplinas y trae consigo nuevos desafíos, tanto éticos como tecnocientíficos. En el artículo lleva a cabo una cartografía de los mismos, a partir de tres dimensiones prácticas atribuidas normalmente a los robots, igual que a las personas: sentir, pensar y actuar. Sobre la base narrativa de un conjunto amplio de entrevistas a personas expertas en diversos campos y aplicaciones de la robótica y la IA, muestra las posibilidades, las implicaciones y las limitaciones de tal atribución.

El artículo de Christi Amesti Mendizábal y Natalia Zardoya Jiménez atiende a los retos jurídicos y éticos que plantea la robótica interactiva y la introducción de robots inteligentes en entornos sociales. Los retos principales giran en torno a la posibilidad de que tales robots se consideren como máquinas capaces de aprender y de tomar decisiones con cierto grado de autonomía, e interferir con la independencia y libertad humanas, que deberían trazarse como líneas fundamentales en el diseño de la actuación de los robots inteligentes. Se plantea además el dilema sobre la manera de configurar

3 Para financiar debidamente el funcionamiento del sistema de gobernanza, debería subyugarse la posibilidad de exigir las correspondientes tasas de registro o de supervisión, o ambas. Estas deben ser proporcionales al coste en que haya incurrido la autoridad competente.

4 Con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas (2020/2012(INL)). Disponible en: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_ES.html [Fecha de consulta 23 de noviembre de 2021].

5 Dirección General de Redes de Comunicación, Contenido y Tecnologías, COM/2021/206 final, Bruselas, 21 de abril de 2021. Disponible en el siguiente enlace: <https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52021PC0206> [última consulta 23 de noviembre de 2021].

6 Expediente interinstitucional: 2021/0106 (COD). Documento ST 8115/21 INIT de envío por la Secretaria General de la Comisión de la COM/2021/206 final a las Delegaciones, Bruselas, 23 de abril de 2021. Disponible en el siguiente enlace: <https://data.consilium.europa.eu/doc/document/ST-8115-2021-INIT/ES/pdf> [última consulta 23 de noviembre de 2021]. Sigue abierto el procedimiento legislativo ordinario en curso 2021/0106(COD), esperando ahora que se pronuncien algunos Comités del Parlamento Europeo.

un sistema de supervisión por parte de las personas, que sea lo suficientemente garantista y que, al mismo tiempo, respete y fomente el potencial de innovación propio de estas tecnologías, mediante iniciativas orientadas hacia una regulación más proactiva que reactiva de los robots y de la IA en la sociedad.

El artículo de María Yolanda Sánchez-Urán Azaña se sitúa en el terreno de conflicto entre la robotización y el empleo. Considerado el trabajo como un factor principal de inclusión social, que conlleva promover la garantía de los derechos de los trabajadores, junto a la libertad de empresa (incluyendo la libertad de innovación en aras de la mayor productividad), se presenta como objetivo la integración real y efectiva de todos los ciudadanos. Si el aumento de la robotización genera una mayor riqueza en los países, esta podría ser utilizada por los gobiernos para reducir la desigualdad a través de políticas redistributivas, mejorando la situación de los ciudadanos. Urge, pues, prestar más atención a la empleabilidad de las personas que a los puestos de trabajo en sí, tratando de superar el efecto de sustitución de humanos por robots y apoyar el uso de los robots colaborativos en el lugar del trabajo.

En el siguiente artículo, Pablo Jiménez Schlegl ofrece primero una panorámica general de los paradigmas de aprendizaje de sistemas artificiales: métodos supervisados y no supervisados, describiendo los tipos de fuentes de datos actualmente disponibles para la comunidad robótica. En este sentido, atiende a continuación al sesgo observable en los conjuntos de datos de imágenes, originado por la anotación humana de las mismas que, utilizada en el entrenamiento de clasificadores, puede llegar a transmitir los prejuicios de los anotadores, especialmente en la categorización de personas. Si esta cuestión de los sesgos presentes en los entornos humanos no se aborda adecuadamente, el aprendizaje automático aplicado a la robótica puede provocar comportamientos totalmente inadecuados de los robots.

Por su parte, en su contribución Ricardo Morte Ferrer, en contra de la consideración generalizada, mantiene que tanto el derecho como las normas, de diferente tipo, pueden y deben ser aplicables a las nuevas

tecnologías (computación en la nube, *big data*, Internet de las cosas, robots, IA). Centrándose en la IA, plantea criterios para que esta se desarrolle y aplique de forma adecuada a los derechos fundamentales en general y a la privacidad en particular. Recurriendo a los objetivos de protección (disponibilidad, confidencialidad, integridad, transparencia, no encadenabilidad y capacidad de intervenir) y a los principios de la ciberética, incluyendo una referencia especial a la sostenibilidad y la resiliencia de este tipo de sistemas, profundiza en ideas como la posibilidad y necesidad de controlar los sistemas de IA (hasta el punto de poder detenerlos) y de auditarlos.

El monográfico se cierra con una contribución que aborda la temática de los Objetivos de Desarrollo Sostenible (ODS) en relación con la implantación social de la robótica. Aníbal Monasterio Astobiza ofrece en su artículo numerosos casos de aplicación de la robótica y la IA encaminados al cumplimiento de los ODS. Frente a una narrativa distópica frecuente en torno a la IA, cuyo avance y desarrollo se presentan habitualmente en medios de comunicación y en la cultura popular en términos de amenazas y riesgos (desempleo tecnológico, sistemas de armas autónomas letales, robots asesinos, etcétera), se valora de manera crítica y constructiva, en esta contribución, el papel de la IA en el logro del bien común (AI4SG), su aplicación para la mejora del bienestar de las personas y para el desarrollo hacia sociedades más prósperas.

En definitiva, cabe concluir que este asunto transdisciplinar es cabal abordarlo con una mirada abierta, capaz de interconectar prudentemente el conocimiento, para ayudar a comprender el mundo presente y pensar en la trascendencia futura de las decisiones que se adopten en materia tan sensible.

AGRADECIMIENTOS

Este trabajo se ha realizado en el marco de los siguientes proyectos de investigación: INBOTS (H2020, 780073), EXTEND (H2020, 779982) y AudIT-S (RETOS PID2019-105959RB-I00 Ministerio de Ciencia, Innovación y Universidades).

REFERENCIAS

- AI HLEG (2019). High-Level Expert Group on Artificial Intelligence: Ethics Guidelines for Trustworthy AI. *European Commission*. Disponible en <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>
- Arnold, Thomas y Scheutz, Matthias (2017). Beyond Moral Dilemmas: Exploring the Ethical Landscape in HRI. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. Vienna: ACM Press, pp. 445-452. Disponible en <https://hrilab.tufts.edu/publications/landscape.pdf> <https://doi.org/10.1145/2909824.3020255>
- Bryson, Joanna J. (2019). The Past Decade and Future of AI's Impact on Society. In *Towards a New Enlightenment: A Transcendent Decade*. Madrid: Turner - BBVA. Disponible en <https://www.>

- bbvaopenmind.com/en/articles/the-past-decade-and-future-of-ais-impact-on-society>
- Coeckelbergh, Mark (2020). *AI Ethics*. Cambridge MA: The MIT Press. <https://doi.org/10.7551/mitpress/12549.001.0001>
- Comisión Europea (2014). *Strategic Research Agenda for Robotics in Europe 2014-2020*. euRobotics. Disponible en <https://www.eu-robotics.net/sparc/upload/topic_groups/SRA2020_SPARC.pdf>
- European Group on Ethics in Science and New Technologies (2018). Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems. Luxembourg: Publications Office of the European Union. Disponible en <https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf>
- Floridi, Luciano; Cows, Josh; Beltrametti, Monica; Chatila, Raja; Chazerand, Patrice; Dignum, Virginia; Luetge, Christoph; Madelin, Robert; Pagallo, Ugo; Rossi, Francesca; Schafer, Bukhard; Valcke, Peggy y Vayena, Effy (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28 (4): 689-707. Disponible en <https://link.springer.com/content/pdf/10.1007/s11023-018-9482-5.pdf> <https://doi.org/10.1007/s11023-018-9482-5>
- Gibert, Martin (2019). Éthique Artificielle (Version Grand Public). In Kristanek, Maxime (ed.). *L'Encyclopédie Philosophique*. Disponible en <<https://encyclo-philo.fr/item/199>>
- IEEE (2019). *Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems* (First Version). Disponible en <<https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf>>
- INBOTS (2019a). INBOTS White Paper on Interactive Robotics Legal, Ethics and Socioeconomics Aspects. Deliverable 2.1. Disponible en <http://inbots.eu/wp-content/uploads/2019/07/Attachment_0-1.pdf>
- INBOTS (2019b). INBOTS White Paper on Interactive Robotics Strategies to Increase Public Awareness and Acceptance. Deliverable 6.1. Disponible en <http://inbots.eu/wp-content/uploads/2019/07/Attachment_0-5.pdf>
- Lin, Patrick; Abney, Keith y Beckey George A. (2011). *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge MA: The MIT Press.
- Müller, Vincent (2020). Ethics of Artificial Intelligence and Robotics. *Stanford Encyclopedia of Philosophy*. Disponible en <<https://plato.stanford.edu/entries/ethics-ai>>
- O'Brolchain, Fiachra y Grau Ruiz, M. Amparo (2020). Environmental Impact of Robotics: Ethical Concerns and Legal Alternatives. In Walter Leal Filho; Anabel Marisa Azul; Luciana Brandli; Amanda Lange Salvia y Toni Wall (eds.) *Industry, Innovation and Infrastructure. Encyclopedia of the UN Sustainable Development Goals*. Springer. https://doi.org/10.1007/978-3-319-71059-4_147-1
- Taddeo, Mariarosaria y Floridi, Luciano (2018). How AI Can Be a Force for Good. *Science*, 361 (6404): 751-752. Disponible en <http://science.sciencemag.org/content/sci/361/6404/751.full.pdf> <https://doi.org/10.1126/science.aat5991>
- Taylor, Steve; Pickering, Brian; Boniface, Michael; Anderson, Michael; Danks, David; Folstad, Asbjorn; Leese, Matthias; Müller, Vincent C.; Sorell, Tom; Winfield, Alan y Woolard, Fiona (2018). Responsible AI – Key Themes, Concerns and Recommendations for European Research and Innovation: Summary of Consultation with Multidisciplinary Experts. Disponible en <<https://www.ngi.eu/wp-content/uploads/sites/48/2018/07/Responsible-AI-Consultation-Public-Recommendations-V1.0.pdf>>
- UNESCO (2017). Report of COMEST on robotics ethics. Disponible en <<https://unesdoc.unesco.org/ark:/48223/pf0000253952>>
- Walsh, Toby (2018). *Machines That Think: The Future of Artificial Intelligence*. Amherst, Mass.: Prometheus Books.
- Whittlestone, Jess; Nyrupe, Rune; Alexandrova, Anna; Dihal, Kanta y Cave, Stephen (2019). Ethical and Societal Implications of Algorithms, Data, and Artificial Intelligence: A Roadmap for Research. Cambridge: Nuffield Foundation, University of Cambridge. Disponible en <<https://mk0nuffieldfounpg9ee.kinstacdn.com/wp-content/uploads/2019/12/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf>>