

José Antonio Rodríguez García: Laicidad e Inteligencia Artificial: La implementación y el alineamiento de la laicidad como valor en los sistemas de Inteligencia Artificial, Publicaciones Cetinia, Editorial Dykinson Universidad Rey Juan Carlos, Madrid, 2024, 356 páginas.

Oscar Celador Angón
Universidad Carlos III de Madrid

La monografía del profesor Rodríguez García, titulada *Laicidad e Inteligencia Artificial: La implementación y el alineamiento de la laicidad como valor en los sistemas de Inteligencia Artificial*, publicada en la editorial Dykinson, es el resultado de una de sus líneas de investigación más recientes. El autor ya avisó de la relevancia de esta línea de investigación con ocasión de las publicaciones de sus trabajos: «¡El futuro ya está aquí! Derecho e Inteligencia artificial», publicado junto a la profesora Mar Moreno Rebato, en la Revista Aranzadi de derecho y nuevas tecnologías, núm. 48, 2018; y «Value Awareness and Process Automation: A Reflection Through School Place Allocation Models», publicado junto a Joaquín Arias, Mar Moreno y Sascha Ossowski, en septiembre de 2023 en la obra colectiva «Value Engineering in Artificial Intelligence». Asimismo, esta investigación es el resultado de la participación del profesor Rodríguez García en el Grupo de alto rendimiento de Inteligencia artificial de la Universidad Rey Juan Carlos y del CETINIA (Centro de Investigación para las Tecnologías Inteligentes de la Información y sus Aplicaciones).

Teniendo en cuenta la trayectoria del autor, tanto con carácter general desde la perspectiva académica, como en el contexto específico de las nuevas tecnologías y la Inteligencia Artificial (IA), no es sorprendente que la monografía que a continuación recensamos derroche calidad, originalidad y, mucho más importante, destaque por su capacidad para introducir al lector en una temática en constante evolución y adaptación a los continuos retos y demandas que plantea la sociedad tecnológica del siglo XXI.

El libro se divide en cuatro bloques. El primer bloque ofrece un concepto de IA, explica cuáles son los valores que están presentes en los sistemas de IA, y en concreto los analiza desde diferentes perspectivas como por ejemplo el reglamento 2024/1689 de 13 de junio de 2024 de la Unión Europea sobre inteligencia artificial. El segundo bloque se centra en el estudio del principio de laicidad, así como en la explicación de cuál es su papel en nuestro ordenamiento jurídico; asimismo, en esta parte de la monografía la laicidad se analiza teniendo en cuenta cuál es su papel en los principales Tratados de Derechos Humanos, y su conexión con los valores religiosos y transhumanistas en los sistemas de IA. Además, se estudia la relación entre la laicidad y el denominado «Derecho computacional», así como la implementación de este principio en los sistemas de IA mediante el denominado «algoritmo laico». El tercer bloque presenta casos de uso específicos, tales como la manipulación de la conciencia, la igualdad y no discriminación (sesgos algorítmicos), y la libertad de expresión en relación con la IA, especialmente en las redes sociales. También se exploran cuestiones como en qué medida los sistemas de IA pueden tener reconocidos algunos derechos fundamentales, como la libertad de conciencia (religiosa o no), el uso de la IA en contextos educativos, y el derecho a utilizar (o a no utilizar) los sistemas de IA, especialmente en el ámbito sanitario. Finalmente, la obra estudia el posible uso de sistemas de IA por parte de grupos religiosos. El cuarto bloque se dedica a las conclusiones del estudio, y en ellas el autor nos ofrece sus valoraciones personales y críticas con la temática objeto de estudio. Se trata de un esquema de trabajo sencillo y eficaz, qué ha permitido al autor acometer su investigación con garantías de éxito.

De acuerdo con este planteamiento, nos encontramos ante un tratado sobre las aplicaciones que la IA puede tener en el contexto de la garantía y el ejercicio del derecho a la libertad de pensamiento, conciencia y religión. Con este objeto, el autor en la primera parte de su monografía introduce al lector, de forma hábil y rigurosa, en los principios, valores éticos y regulación que ordenan el uso de la IA. A continuación, con el objeto de que el lector contextualice el debate acerca del uso de la IA y las creencias y convicciones en el marco constitucional, analiza, desde la perspectiva del principio de laicidad, los principales puntos de encuentro entre la IA, las creencias y convicciones personales, las confesiones religiosas, los valores transhumanistas y los valores religiosos. Y, posteriormente, estudia los «casos de usos» que están directamente relacionados

con el contenido de la libertad de conciencia y con la igualdad y su correlato de no discriminación por motivos ideológicos y o religiosos.

La IA se ha desarrollado a una velocidad vertiginosa. En un periodo muy corto de tiempo. Lo que era pura ciencia ficción y veíamos con un futurible en las películas se ha convertido en una realidad, y está presente en las principales facetas de nuestra cotidianidad. La IA presenta numerosas ventajas e inconvenientes. Por una parte, La IA ha permitido automatizar y acelerar procesos cotidianos de especial relevancia en nuestro día a día. Pero, por la otra, su uso inapropiado puede atentar contra los derechos fundamentales; pensemos, por ejemplo, en los efectos que la IA tiene en las redes sociales y su capacidad para manipular imágenes, datos e información de especial relevancia para la formación de la opinión pública.

Los cuatro tipos de IA que analiza el autor permiten entender el retó al que nos enfrentamos: sistemas que piensan como humanos; sistemas que actúan como humanos; sistemas que actúan racionalmente pero que tratan de emular de forma racional el comportamiento humano; o sistemas que piensan racionalmente y qué tal limitar o emular el pensamiento lógico racional del ser humano. Siendo esta la principal misión de la IA, rápidamente surge la cuestión de en qué medida es necesario limitar su capacidad para emular al ser humano, de acuerdo con unos principios morales y éticos básicos que sean coherentes con el orden público y los principios constitucionales. En este apartado, en mi opinión, brilla con especial relevancia el esfuerzo que hace el autor por conectar los valores éticos, el derecho y la IA.

El capítulo de la monografía dedicado al análisis y estudio del principio de laicidad de los poderes públicos es especialmente interesante, toda vez que en esta parte de la monografía se aprecia la trayectoria académica del profesor Rodríguez García y su capacidad para analizar principios jurídicos complejos, como ocurre con el principio de la laicidad. En palabras del autor: «lo que cualquier estado laico y, por ende, democrático, debe garantizar es que los derechos humanos se plasmen en leyes y que estas recojan los valores de la ética laica, que todos podemos y debemos compartir. Además, la laicidad garantiza la independencia de los poderes públicos de la moralidad religiosa [...] No se puede basar la moral ni la ética universal en valores éticos o morales que son particulares. Como se comprobará en un capítulo de

este Bloque II, los valores religiosos podrán ser implementados en los sistemas de IA, con carácter general, solamente, en la medida que coincidan con los valores comunes universales, como son los derechos humanos y la laicidad, como garantía de los mismos. Es decir, que cuando los valores morales como el respeto al derecho de otros a profesar sus ideas religiosas, la imparcialidad frente a la pluralidad religiosa y el no tratar de imponer las convicciones religiosas personales al resto de la sociedad se llevan al plano de la ética pública, se llama laicidad» (pp. 89-90).

La parte de la investigación dedicada a estudiar la laicidad y el derecho a formar libremente la conciencia, así como a responder a la cuestión de en qué medida los sistemas de IA pueden manipular la conciencia, me han parecido extremadamente interesante. En este apartado confluyen una serie de bienes jurídicos cuyo alcance es preciso delimitar como, por ejemplo, el derecho a no revelar las propias ideas, creencias, convicciones, pensamientos y opiniones; el derecho a que no se manipulen nuestras ideas, creencias, convicciones, pensamientos y opiniones; y el derecho a no ser penalizado por nuestras propias ideas, creencias, convicciones, pensamientos y opiniones, siempre que permanezcan en nuestro foro interno, ya que, como señala el autor, se trata de la posibilidad de leer nuestras ideas, pensamientos y emociones, lo cual puede conllevar que se nos penalice por ello, o bien que no se nos permita acceder a determinados contenidos o servicios por este motivo. En este apartado el autor se centra, entre otros debates de especial interés, como el que ha surgido con ocasión de los neuroderechos, que son nuevos derechos humanos que protegen la privacidad e integridad mental y psíquica tanto consciente como inconsciente de las personas del uso abusivo de las neurotecnologías.

En esta parte del trabajo también me ha aparecido muy novedoso el planteamiento que el autor hace de la denominada «vigilancia policial predictiva», debido al uso de sistemas de IA para prevenir el delito con ocasión de atentados terroristas en Estados Unidos y en Europa. De ahí que los poderes públicos hayan pedido a las plataformas de redes sociales que utilicen sus algoritmos para intentar identificar perfiles de potenciales terroristas, así como adoptar medidas para identificar cuentas que generen contenido que pueda promover el fanatismo, la violencia o el radicalismo.

La monografía finaliza con un último capítulo donde, a modo de bloque IV, el autor ofrece 8 conclusiones en las cuales analiza de forma sistemática la temática objeto de investigación. Se trata de conclusiones serias y rigurosas que permiten al profesor Rodríguez García explicar de forma sobresaliente su posición acerca de las diferentes temáticas analizadas en la monografía, de enorme interés para los estudiosos de las Ciencias de la computación con carácter general, y en concreto para todos aquellos que se quieren estudiar la IA desde la perspectiva jurídica.

Con esta obra, el profesor Rodríguez García introduce a los estudiosos del Derecho Eclesiástico del Estado en el mundo de la IA, toda vez que, como el autor aclara, estamos ante un proyecto en constante revisión. En otras palabras, las conclusiones a las que llegamos hoy en materias tan tecnológicas, pueden estar obsoletas dentro de unos meses cuando surja una nueva modalidad de IA o está evolucione para adaptarse a los nuevos retos que le plantea la sociedad.

La propuesta del profesor Rodríguez García de elaborar un «algoritmo laico» que esté implementado en los sistemas de IA me parece muy interesante. En su planteamiento el autor diferencia entre los sistemas de IA públicos y privados, así como entre aquellos sistemas de IA que tienen trascendencia pública y aquellos que carecen de la misma y no tienen repercusión en el debate político y público. De acuerdo con esta lógica, el autor propone diferenciar cuatro tipos de sistemas de IA atendiendo a los espacios donde estos se utilicen o a quienes sean sus usuarios. En primer lugar, tendríamos los sistemas de IA institucionales, que son aquellos que utilizan los poderes públicos y que deben implementar el denominado algoritmo laico con el objeto de garantizar en condiciones de igualdad el ejercicio de la libertad de conciencia de los individuos. En segundo lugar, los sistemas de IA utilizados por las organizaciones con fines religiosos para sus actividades religiosas no tienen que implementar el algoritmo laico, debido a que la laicidad incluye el reconocimiento del derecho a la autonomía de las organizaciones con fines religiosos, y se trata de grupos soportados en una ética privada, que no tiene por qué ser ideológica o religiosamente neutral. En tercer lugar, los sistemas de IA que se utilizan en el ámbito privado, o en espacios donde el individuo establece a título personal sus reglas, tampoco deben implementar el algoritmo laico; y sus usuarios deben de ser libres para decidir si quieren mantener el algoritmo tal y como está implementado por defecto o modificar su configuración para atender a

sus necesidades personales. Y, en cuarto lugar, tendríamos los sistemas de IA que, en palabras del autor, actúan en un espacio de contornos confusos, y que por lo tanto están sujetos a fuertes tensiones de interpretación y regulación.

El reglamento comunitario 2024/1689 de 13 de junio de 2024, está llamado a tener una especial importancia en este ámbito, ya que con el mismo se pretende, entre otras cosas, generar lo que el autor denomina el «efecto Bruselas» que, entre otras iniciativas, intenta conservar la influencia geopolítica de la Unión Europea y promover los valores éticos y humanistas implícitos en el artículo 2 del tratado de funcionamiento de la Unión Europea. Cómo señala el autor: «Europa es la conciencia del mundo, es la idea que se transmite» (p. 303). El reglamento no considera sistemas de alto riesgo aquellos que no puedan generar daños o perjuicios a los derechos fundamentales de las personas físicas, así como aquellos que no hay influyan de forma sustancial en la toma de decisiones. Pese a esto, los sistemas de IA que no son de alto riesgo pueden someterse a códigos de conducta con el objeto de que tengan las nuevas condiciones de uso que los sistemas de IA de alto riesgo.

Desde la perspectiva del principio de laicidad, el autor señala que «el alineamiento es la evolución del desarrollo de los sistemas de IA y que esos actúen de conformidad con unos determinados valores», lo cual no será compatible con el valor universal de la laicidad. Según el profesor Rodríguez García, «para combatir la discriminación, los sesgos, la manipulación algorítmica, etc., es necesario la auditoría de los sistemas de IA, dónde se incluye el derecho a la explicación. No se puede eximir a los implementadores de los sistemas de IA de la responsabilidad por las violaciones de los derechos humanos, de la no discriminación y de la laicidad y, para ello, se debe exigir que expliquen de forma accesible y sencilla la decisión que han tomado los sistemas de IA» (p. 307). Desde la perspectiva de la garantía y protección del ejercicio de los derechos humanos nos encontramos ante una propuesta impecable.

Finalmente, el autor justifica su preocupación por la autorregulación ética de los sistemas de IA, debido a que la mayoría de los sistemas de IA no les será de aplicación directa el reglamento 2024/1689 anteriormente referido.

En resumen, en un contexto tecnológico, económico, social y cultural, que reclama el uso de la IA, parece evidente la necesidad, que reclama el autor en las páginas finales de su libro, de incluir el estudio de la IA en los diferentes planes de estudio, de forma que los discentes conozcan de primera mano las ventajas y desventajas que tiene su uso, en la medida en la que esta puede influir de forma determinante en el libre desarrollo de su personalidad y en el ejercicio de los principales derechos y libertades fundamentales.

Una vez más, el profesor Rodríguez García ha demostrado su capacidad para abordar con éxito el estudio de temáticas complejas, caracterizadas, como ocurre en este caso, por la ausencia de trabajos de investigación, pese a la relevancia y a la actualidad de la temática. Por todo ello, felicitamos al autor y recomendamos la lectura de su sugerente trabajo.

