

Fernando H. LLANO ALONSO, *Homo Ex Machina.*
Ética de la inteligencia artificial y Derecho digital
ante el horizonte de la singularidad tecnológica,
Tirant lo Blanch, Valencia, 2024, 248 pp.

MIGLE LAUKYTE
Universitat Pompeu Fabra

Palabras clave: inteligencia artificial, ética, filosofía del Derecho, algoritmos, sesgos, discriminación, transhumanismo
Keywords: artificial intelligence, ethics, legal philosophy, algorithms, bias, discrimination, transhumanism

Hace más de cuatro años tuve la oportunidad de escribir la recensión de otro libro del Prof. Fernando H. Llano Alonso, bajo el título “*Homo Excelsior. Los límites ético-jurídicos del transhumanismo*” (Tirant lo Blanch, Valencia, 2018, 238 pp.). Ahora, con *Homo Ex Machina*, tengo entre las manos ya su segundo libro, en lo que espero represente un inicio de una serie sobre la temática de los avances tecnológicos, humanismo, Derecho y ética. No diré nada más sobre la relación entre estos dos libros, ya que el Prof. Stefano Pietropaoli –quien al escribir el *Prólogo* finaliza con una bella frase al llamar a este libro “una bandera que ondea con orgullo al duro viento de [...] campo de batalla [de la lucha por la dignidad humana]”-explica perfectamente la relación entre las dos monografías. Me limito a observar que uno de los muchos hilos que entrelazan el pensamiento entre los dos libros es la presencia del pensamiento de José Ortega y Gasset (1883-1955) y sus ideas, como, por ejemplo, la del ser humano como centauro ontológico. El centauro ontológico es humano si y solo si existe la técnica que le permite crearse su mundo en paralelo a la naturaleza: mundo tecnológico, por lo tanto, es la dimensión alternativa, otra realidad donde escapamos, donde nos realizamos, donde nos expresamos y exploramos. Esta idea como ninguna otra nos lleva a entender mejor no solo la continua relevancia del pensamiento de José Ortega

y Gasset, sino también sus reverberaciones en los pensadores y filósofos (del Derecho) contemporáneos, que tienen que enfrentarse y dar sentido a los nuevos retos, responder a unas nuevas preguntas y situar la nueva realidad de este mundo paralelo en el contexto de la actualidad.

Dicho esto y tornando al libro en cuestión -libro *Homo Ex Macchina*- Fernando H. Llano Alonso demuestra una y otra vez su profundo conocimiento del desarrollo del pensamiento humano, tanto antiguo como contemporáneo, tanto mundial como nacional, tanto filosófico como filosófico-jurídico. Además, brilla por su habilidad para conectar este pensamiento -situarlo, enlazarlo o, mejor dicho, enraizarlo- con la actualidad tecnológica, evidenciando que lo que vivimos no es algo que nos ha pasado de un día para otro (como parece ser, por ejemplo, en el caso de la Inteligencia Artificial (IA), que explotó con la IA generativa, pero que ya estuvimos observando y analizando desde el siglo pasado), sino que es una natural continuación de la conversación entre ciencias naturales, ciencias exactas y las humanidades, entre los avances tecnológicos y los valores humanos, entre lo que somos, lo que nos hacen creer que somos y lo que queremos ser.

Este libro es también un *J'accuse...!* elegante contra lo que el autor mismo llama “la revolución pedagógica, mercadotécnica y científico-tecnológica,” que busca sustituir “contenidos cognitivos por sus contenedores” (pp. 29-30) en nuestras universidades. Esta realidad que estamos viviendo en tiempo real y a la que muchos intentamos oponer resistencia en las instituciones de educación. Parece que algo se está moviendo ya: “los contenedores” finalmente se ponen en duda. Obviamente, el Derecho tampoco se salva de estas dinámicas del mercado y disruptión tecnológica, pero dichas dinámicas son peligrosas solo si no son acompañadas por los ideales del humanismo: parafraseando a Pinker (2018), citado por el autor, este acompañamiento es lo que nos ayuda a distinguir entre progreso de la humanidad del dominio de la misma. En la versión en la que la humanidad viene dominada por la tecnología, no hay sitio para los humanos, ni para sus valores, y de consecuencia tampoco para el Derecho.

Por lo tanto, estamos frente a un libro que argumenta -entre muchas otras cosas- que para tratar los avances tecnológicos en la manera que no perjudique la humanidad tenemos que reforzar la relación entre las humanidades (la ética, la filosofía, la literatura, ...) y el Derecho. Es decir, que la tecnología en sí no soluciona problemas ni es la respuesta, sino que el ser humano necesita más herramientas-y algunas ya las tiene dentro, es decir

su creatividad, su empatía, su humanidad, compasión, sus valores, sus sentimientos, su vulnerabilidad-para tomar las decisiones sobre el futuro sea cual sea el ámbito de sus decisiones. Y en esta tarea tan compleja, donde se contraponen y se enfrentan tantos poderes y muchos intereses, el Derecho, con sus garantías, valores, principios, es esencial.

El autor dice que el objetivo de este libro es una teoría alternativa al iustecnicismo y la verdad es que es un libro que aboga por el humanismo tecnológico donde ambas partes -avance tecnológico e ideas humanistas- se encuentran y encajan, y además crean un terreno fértil para un nuevo Derecho, abierto a otras disciplinas y conocimientos y, gracias a esta abertura, se encuentra listo y preparado para afrontar los retos de las sociedades contemporáneas.

La obra está organizada en 6 capítulos y cada uno se merece una recensión más detallada porque cada uno lleva ideas con las que uno puede estar más o menos de acuerdo, pero todas son merecedoras de atención y consideración. Seguramente muchas de las ideas son conocidas a los que estén familiarizados con estos temas desde hace algún tiempo, pero no se puede prescindir del papel educativo y divulgativo de este libro: muchos aún no están plenamente al corriente del vasto abanico de cuestiones que trata este libro, ni están preparados para razonar críticamente sobre las mismas.

El autor divide los capítulos de la siguiente forma. Los capítulos relativos a la Ética (capítulo 1 y 3), al Derecho (2 y 4) y a la Política (5 y 6) si bien estas distinciones son más bien declarativas: estos temas se solapan, se sobreponen, se funden y lo hacen aún más cuando se trata del campo de tecnologías disruptivas que tan fuertemente han desestabilizado la vida humana.

En el Capítulo 1 “*Singularidad tecnológica, Metaverso e identidad personal: del Homo Faber al Novo Homo Ludens*”, Fernando H. Llano pone el enfoque en la identidad personal y la sitúa dentro de los avances neurocientíficos, que más que cualquier otra tecnología cuestiona la esencia humana: al final, es en el cerebro y no en ningún otro sitio, donde nace, se desarrolla y se muere la identidad de cada uno de nosotros. Para resguardar en alguna manera este proceso creativo-evolutivo de la identidad individual, hemos llegado a hablar de los neuroderechos y, en particular, de la protección del derecho a la identidad mental como una nueva versión de (o continuación del ya existente) derecho a la identidad personal.

Los matices de la identidad personal -atención a no confundirla con la identidad humana (la diferencia explicada por Rafael de Asís en la nota a pie

de página 2)- se descubren a través de un abanico de nuevos derechos digitales relacionados con el espacio virtual. Dichos derechos digitales muchas veces son reducidos erróneamente a los derechos de la privacidad y protección de datos, pero son muchos más. Lo podemos leer en la normativa de protección de datos española, dedicada a la “protección de datos y garantía de los derechos digitales” y también la “Carta de Derechos Digitales” adoptada por primera vez por España y un poco más tarde también en la UE bajo el título “La Declaración sobre los Derechos y Principios Digitales” donde la UE se compromete con (entre otras cosas) “velar porque se ofrezca a las personas que viven en la UE la posibilidad de una identidad digital accesible, voluntaria, segura y fiable que proporcione acceso a una amplia gama de servicios en línea”.

Fernando H. Llano Alonso parece estar de acuerdo con los derechos digitales (y la identidad digital), pero mucho más reacio a incluir entre los mismos los neuroderechos. Entre los entusiastas y los críticos a esta sugerencia de ampliar la lista de los derechos humanos e incluir una nueva categoría, el autor de este libro parece estar más con los críticos, si bien es verdad que deja la puerta abierta (en línea con la tradición filosófico-jurídica del Prof. Antonio Enrique Pérez Luño quien en 1993 decía en esta misma revista que “Una concepción generacional de los derechos humanos implica, en suma, reconocer que el catálogo de las libertades nunca será una obra cerrada y acabada. Una sociedad libre y democrática deberá mostrarse siempre sensible y abierta a la aparición de nuevas necesidades, que fundamenten nuevos derechos”)¹.

La verdad es que tenemos que tener en cuenta el escenario en que los avances tecnológicos no nos ofrecerán esta posibilidad de nuevos derechos, sino que nos empujarán a exigirla: si hoy hablamos de posibilidad y oportunidad de expandir el listado (con todas las consecuencias negativas que esto podría conllevar a los derechos ya existentes), el día de mañana podremos encontrarnos en una situación de la urgente necesidad de tenerlos y –lo más importante cuando hablamos de derechos– disponer de herramientas para hacerlos reales tanto en el mundo físico, como en el digital.

El interesante pasaje que se hace en este capítulo aviene entre la identidad personal y el ensimismamiento y la autoconciencia como actos de la liberación humana de las cosas y los ruidos del mundo. Contrastá con esta

¹ A.E. PÉREZ LUÑO, “El concepto de los derechos humanos y su problemática actual”, *Derechos y Libertades*, núm 1, 1993. pp. 179–198.

visión, pues, la visión del Metaverso como realidad que absorbe todo, identidad humana incluida. En este contexto del Metaverso, que representa la versión nueva de la realidad, la identidad humana se fragmenta y es por eso que, si bien otros neuroderechos pueden ser puestos en duda, el neuroderecho de la identidad mental parece ser el más importante y urgente para garantizar. El Metaverso aún no ha desplegado con toda la fuerza, pero cuando lo haga -y ya hemos aprendido con la IA generativa que esto puede pasar de un día para otro- con sus avatares y (como los llama el autor) "no-cosas" podría ya ser tarde. Podría ser tarde, porque el mundo virtual siempre influye más sobre el mundo real: lo diseña, lo moldea, a veces parece que lo come y lo escupe en una nueva versión, más pobre, más limitada, más algorítmica y menos crítica. Redes sociales, criptomonedas, ... son solo algunos de los ejemplos de estas dinámicas: en el mundo del Derecho, por ejemplo, esto se refleja en la presencia de los contratos autoejecutables, siendo así que, en sentido jurídico de contratos tienen más bien el nombre pero no la esencia de encuentro de voluntades.

Lo más importante es que no somos del todo conscientes de que lo que antes considerábamos ser un espacio digital, es decir un lugar, una zona, ya no es solo un espacio, sino que es un espacio con otras entidades (avatares, robots, agentes, etc.) y nuestra tarea es situar el ser humano en este espacio y entre estas entidades sin que el ser humano pierda su esencia, que es su identidad personal y humana. Al final, todo se reduce a la pregunta un poco Hamletiana de si el papel del ser humano es el papel de objeto o de sujeto.

El Capítulo 2 "*Justicia digital, algoritmos y Derecho: de la predictibilidad del Big Data al mito del juez-robot*" nos lleva directamente a hablar de la justicia, o, mejor dicho, de Derecho a través del sistema de la justicia y su compleja relación con la IA. Hay que notar que este capítulo enfoca sobre todo en la IA predictiva -el autor aquí se refiere a la justicia predictiva que "reduce el Derecho a un pronóstico de un resultado cierto"- y su hipotético impacto sobre el sistema de la justicia, y sobre todo acerca del trabajo y del papel del juez (humano).

Es interesante observar cómo la posibilidad de sustituir al juez con un sistema de IA nos empuja a hacernos preguntas que a veces van más allá del sistema de la justicia: ¿dichas preguntas se quedarían sin formular, sino fuera por la IA? Sin duda esta necesidad que tenemos de hacernos preguntas esenciales sobre la sociedad, sus instituciones, estructuras, principios y valores, papeles y responsabilidades cada vez que nos encontramos ante un

nuevo avance tecnológico que promete irrumpir nuestra realidad (jurídica, económica, etc.) conocida, está entre los mejores efectos secundarios de toda la revolución digital. El problema verdadero se presenta cuando, si bien nos hemos formulado las preguntas, no trasferimos estas reflexiones en el diseño y despliegue de la tecnología, sobre todo cuando se tratan de unos sectores tan sensibles como la justicia.

Dicho esto, es también verdad que ninguna profesión jurídica -ni abogado, ni procurador, ni el fiscal, por nombrar algunas- haya sido sustituida tantas veces (por lo menos a nivel hipotético) por IA como el juez. La pregunta de por qué eso es así merecería un trabajo de investigación aparte: ¿no sería más fácil sustituir al fiscal? ¿Por qué enfocar al juez? La respuesta podría ser que sustituir el juez-última autoridad en el proceso para tomar las decisiones es mucho más complicado que sustituir cualquier otro profesional de Derecho y, por lo tanto, una vez encontrada manera de hacerlo con el quien es más complicado, sería todo mucho más fácil con los restantes profesionales jurídicos. No sé si esta hipótesis mía es correcta, pero es obvio que la tarea del juez representa sin duda el reto más desafiante para la IA (y para los que la desarrollan): el trabajo de "hacer justicia" para las partes del juicio² es la llave para muchas otras tareas en distintas áreas de la actividad humana, pero por encima de esto, también hacer justicia se basa en tener la confianza de la sociedad. Si la IA un día la llega a tener en cuestiones tan fundamentales, podremos decir que nos ha conquistado al cien por cien?

Sea cual sea la respuesta a esta pregunta, Fernando H. Llano dedica una parte considerable del segundo capítulo a tratar la diferencia entre la justicia de los algoritmos predictivos (justicia predictiva) y la justicia basada en reglas y principios. Sin duda, el autor está más en favor de esta última (basada en principios y reglas), mientras vea la primera como una herramienta para la "desinstitucionalización del derecho y la deshumanización de la Justicia en aras de un solucionismo tecnológico que nos asegura la certeza o la precisión de los resultados" (p. 85). En este sentido, Fernando H. Llano se une a una corriente de los pensadores fuertemente críticos con el tecno-solucionismo como única y mejor solución a los problemas complejos, sean estos de naturaleza económica, política, social, ambiental o, como en nuestro caso, jurídica. Hemos visto fallos de esta visión en Internet gracias al trabajo de

² D'AMATO, "On the Connection Between Law and Justice". *Faculty Working Papers. Paper 2*, 2011 <http://scholarlycommons.law.northwestern.edu/facultyworkingpapers/2>

Evgeny Morozov³. Parece que tenemos dificultades abandonar esta visión: desde Internet-centrismo hemos pasado a la IA-centrismo y en este capítulo tenemos la explicación de cómo el techno-solucionismo se aplicaría en el ámbito judicial, donde la realidad ya no es la que es, sino sustituida por una recreada artificialmente y, por cierto, llena de falsas certezas.

Fernando H. Llano argumenta su oposición a la justicia predictiva a través de los principios de transparencia, la legalidad, la causalidad (sustituida en el mundo algorítmico por la correlación) entre otros, pero también subraya que lo que de verdad nos ayudará a entender la coexistencia entre los avances tecnológicos y el Derecho son dos paradigmas que nos llevarán verso “el nuevo modelo de Derecho en que el factor humano cohabitará con la IA legal y la justicia digital” (p. 97). El primero es el paradigma del humanismo tecnológico, mientras que el segundo es paradigma posthumanista. Sin ninguna duda, el autor es partidario del primero.

En el Capítulo 3 “*De máquinas y hombres. Tres cuestiones ético-jurídicas sobre la inteligencia artificial*” volvemos a los temas ético-jurídicos donde el discurso enfoca los grandes temas como la singularidad, conciencia, personalidad, identidad (¡de nuevo!) y responsabilidad. En este capítulo el autor se apoya en la literatura de la ciencia ficción para formular las preguntas al centro de este capítulo. Este método de tratar la ética jurídica a través del pensamiento literario es una manera de recordar la estrecha relación –¿a lo mejor no tan popular en los ambientes académicos continentales?– entre el Derecho y la literatura, porque, como ya lo observó la filósofa Martha Nussbaum, “la narrativa y la imaginación literaria no solo no se oponen a la argumentación racional [del Derecho], sino que pueden aportarle ingredientes esenciales”⁴.

Con la primera pregunta sobre la singularidad tecnológica, el autor expone los argumentos en favor y en contra de esta posibilidad. Ambas posiciones tienen argumentos muy válidos y la posición intermedia es a lo mejor la más prudente en este contexto, sobre todo observando la velocidad del avance tecnológico. Pero la más interesante es la segunda pregunta –la pregunta sobre la conciencia– que da el sentido a la primera: ¿esta singularidad puede ser consciente? ¿Tener la voluntad? ¿Parecerse a la conciencia humana?

³ E. MOROZOV, *To Save Everything, Click Here – The Folly of Technological Solutionism*, PublicAffairs, New York, 2013.

⁴ M. C. NUSSBAUM, *Justicia Poética. La imaginación literaria y la vida pública*, Editorial Andrés Bello, Barcelona, trad. Carlos Gardini. 1997, p. 16.

Para tratar esta cuestión nos adentramos en el mundo de Isaac Asimov y sus leyes de la robótica que han dado forma y contenido a numerosos análisis filosófico-jurídicos sobre las posibilidades, necesidades y responsabilidades de la IA en relación con los seres humanos (y al contrario, las de los humanos en relación con la IA). Pero Fernando H. Llano Alonso también entrelaza los relatos de Asimov con cuestiones jurídicas más profundas, cuál es la Teoría General del Derecho, y argumenta que al final la pregunta de base en la literatura de ciencia ficción es la pregunta sobre la conciencia robótica. Para analizarla, el autor adapta la clasificación de Margaret A. Boden quien en el trabajo de 2017⁵ propuso dividirla entre conciencia artificial funcional y conciencia funcional fenoménica. La revisión de distintas visiones al fin y al cabo lleva a que lo que nos ayudaría a entender mejor la cuestión es crear un doble plano educativo-cultural, basado en educación humanista, por un lado, y ético-jurídico, inspirado en las ideas de regulación antropocéntrica y antropogénica, por otro lado. Probablemente no estamos preparados para tratar estas cuestiones y separarlas de las meras hipótesis que siempre parecen lejanas y, por lo tanto, de una relevancia relativa.

Siempre navegando entre distintas obras de literatura, el autor llega a la tercera pregunta sobre la personalidad robótica y los derechos humanos (y no humanos). Esta cuestión, que tuvo sus momentos de mayor atención por parte del mundo académico hace algunos años sigue sobre la mesa porque la posibilidad de tener las máquinas que son algo más que meros objetos nunca ha desaparecido (véase por ejemplo los trabajos del autor estadounidense David J. Gunkel y también de los españoles como Prof. José Julio Fernández Rodríguez y Prof. María del Mar Méndez Serrano). Desde el punto de vista puramente jurídico, si bien distintas tendencias abundan, el autor hace hincapié en el papel de las instituciones internacionales, y sobre todo de la UE para equilibrar y alinear los principios fundamentales de nuestra convivencia y los avances en la IA y robótica. Sin ninguna duda, el original “no” a los derechos de los robots encuentra nuevos obstáculos y nuevos argumentos que llevan a ver la idea de los derechos no humanos y la personalidad jurídica de no humanos desde una nueva óptica.

En el Capítulo 4 “*Homo ex machina: el ser humano ante la inteligencia artificial y el gobierno de los algoritmos*”, tenemos delante el capítulo que ha dado nombre al libro y donde el autor aborda el tema de interacción humana con

⁵ M. A. BODEN “Artificial Intelligence” en J. AL-KHALILI (ed.), *What’s Next? Even Scientist Can’t Predict the Future – Or Can They?*, Profile Books, London, 2017, pp. 118–128.

distintas formas de IA y robótica. Es decir, nos alejamos de las hipótesis del capítulo 3 –conciencia artificial, personalidad robótica y demás– y enfocamos más bien las posibles configuraciones de control a través de las cuales el ser humano podría interactuar con los sistemas artificiales. En este sentido, las opciones son cuatro: la de control humano (*Human-in-the-Loop*), la de interacciones y transferencias de control entre el ser humano y la máquina (*Human-in-the-Loop for Exceptions* y *Human-on-the-Loop*) y aquella donde el control humano no existe (*Human-out-of-the-Loop*). Una vez analizadas estas opciones, el autor llega a concluir que sea cual sea la manera que adoptemos de interactuar con las distintas máquinas y sistemas que cada vez nos rodearán más, hay que garantizar que dichas interacciones lleven a las decisiones humanas auténticas, que se basen en (y surgen de) la transparencia algorítmica, que es la manera de garantizar nuestra participación y control de la IA. La transparencia, entendida no solo como posibilidad de acceder, sino también como la necesidad de entender cómo las máquinas llegan a los resultados, es la llave de todo.

En lo que resta del capítulo el autor entra también en el debate sobre como proteger el coto vedado a la intromisión y mal uso de la IA: la reserva de la humanidad probablemente no es suficiente, si bien el legislador europeo, adoptando el término de la supervisión humana de los sistemas de IA de alto riesgo obviamente piensa distintamente. Es decir, la supervisión humana en sí misma es problemática y tenemos que ser bien conscientes de sus limitaciones.

Por cierto, esta reserva de la humanidad choca no solo con la IA en sí, sino también –y probablemente, sobre todo– con la presunción humana de la perfectibilidad de la IA: el ejemplo sobre el uso de la IA en medicina es emblemático en este sentido y la falta de conocimientos por parte de seres humanos del funcionamiento, de las capacidades y también de los límites de la IA es el mayor obstáculo para hacer la reserva de la humanidad más efectiva. Eso nos lleva de nuevo a la importancia de la transparencia de la IA y no es casual que el legislador europeo incluya en la nueva ley de la IA el artículo sobre la alfabetización en la materia de la IA (artículo 4).

El Capítulo 5 “*Globalización y revolución tecnológica ante la crisis de paradigma humanista*” sitúa la IA y los avances tecnológicos en su conjunto, que han permitido el crecimiento de las Big Tech, dentro de los procesos de globalización, que en los últimos años se ha alineado y mutuamente reforzado con la proliferación de las ideas del transhumanismo. Bajo la bandera de la

cuarta revolución industrial y El Foro Económico Mundial como fondo, el autor nos advierte sobre el riesgo de tecnoutopía que justificaría el cambio del paradigma del humanismo al paradigma técnico-científico, tan caro a los transhumanistas.

Sigue un análisis detallado de las ideas de transhumanismo, ampliamente explicadas por su representante, el fundador del Foro Económico Mundial, Klaus Schwab, quien mejor que nadie representa estas ideas del nuevo inicio para la humanidad. Según él, dicho inicio se basa en la quiebra de la gobernanza mundial, descalificación y obsolescencia de las instituciones internacionales y la necesidad de un nuevo contrato social. Obviamente, siendo el fundador del Foro Económico Mundial, Klaus Schwab está en una posición de privilegio para comunicar su punto de vista y tiene la credibilidad y autoridad que pocos foros pueden otorgar y pocas instituciones pueden ofrecer, y por lo tanto, muchas de sus ideas han encontrado apoyo entre los gobiernos de muchos países.

El autor trata los mandamientos de Schwab y los desenreda con una argumentación clara y contundente, apoyándose también en otros autores como Shoshana Zuboff quien en el 2020 ha dado la vuelta al mundo con su libro de denuncia de lo que está pasando en el mundo globalizado por las tecnologías de vigilancia⁶. Su conclusión es invertir más-y no menos, como sostiene Schwab y sus seguidores-en la gobernanza global democrática, sobre todo a través del apoyo a la UE como un actor internacional que no solo se hace oír, sino también se hace seguir gracias al “efecto Bruselas” explicado tan bien por Anu Bradford en su libro de 2020⁷. Hemos podido observar este efecto en muchos ámbitos, como por ejemplo, la protección de datos, protección ambiental y alimentaria: esperamos poder repetir con la Ley de la IA que, a pesar de todas las deficiencias que tiene, lleva un enfoque de garantías para los consumidores.

El capítulo acaba con las reflexiones sobre el papel de la UE en crear un espacio tanto tecnológicamente competitivo y avanzado, como respetuoso con los derechos de los ciudadanos y los valores y los principios ético-jurídicos que están en la base de los Tratados. Obviamente, la nueva Ley de la IA ocupa un espacio especial en esta visión: los tiempos legislativos no han permitido al autor tratar el texto finalizado ni tampoco ver la última versión,

⁶ H. ZUBOFF, *La era del capitalismo de la vigilancia*, Paidós, Barcelona, 2020.

⁷ A. BRADFORD, *The Brussels Effect. How the European Union Rules the World*, Oxford University Press, New York, 2020.

pero las otras iniciativas-desde el Reglamento General de Protección de datos, Directrices éticas para una IA fiable, La Ley de Servicios Digitales, entre muchas otras iniciativas-claramente representan el enfoque Europeo que es distinto de la visión de la UE y de otras regiones y continentes. Pero exactamente por este motivo -porque es distinto y enfocado en los valores que la mayoría de la humanidad, tan imprevisible y contradictoria, sin embargo, considera esenciales- no podemos cansarnos en promoverlas, explicarlas y divulgarlas en los foros que compiten con cada uno de nosotros como ciudadanos.

El último capítulo del libro, el capítulo 6 *"Inteligencia artificial, discriminación y sesgos algorítmicos"* se dedica a este problema tan inmenso y denunciado por muchísimos autores en todo el mundo. El autor analiza en detalle cómo surgen sesgos algorítmicos, sus conexiones con los sesgos humanos, que a veces son tan obvios y omnipresentes que se vuelven invisibles -no hace falta ni mencionarlos, si bien diría que los relativos al género tal vez son los más difundidos- y en la misma manera son transferidos al mundo digital y en las aplicaciones y sistemas de IA, que los repiten, hacen crecer y difunden más y mejor.

Lo que personalmente considero fascinante en el discurso sobre IA discriminatoria y sesgada es que la IA nos ha hecho ver nuestros sesgos, prejuicios y actitudes discriminatorias. Muchísimas veces no nos hemos dado cuenta de los sesgos, no los hemos pensado, no hemos revisado críticamente nuestras maneras de actuar; en fin, hemos hecho algo, pero no suficiente para detectarlos y eliminarlos. Lo que sí que hemos hecho es transferirlos -programarlos, digitalizarlos, convertirlos en lenguaje máquina- a la IA. Y luego, cuando IA ha generado las respuestas discriminantes y claramente en contra de la igualdad de las personas, nos hemos dado cuenta: muchas veces, escuchando los medios de comunicación, nos viene la impresión de que la discriminación es de la IA, pero no es así. La discriminación entendida como una injusticia representa un problema estructural que llevamos como una mochila a lo largo de la historia de la humanidad.

Y es por eso que el autor encuentra la raíz de IA sesgada en sus programadores que "no son plenamente conscientes de problema de prejuicios, ni tan siquiera de sus propios prejuicios, ...[ellos] no imaginan ni han reflexionado lo bastante sobre las consecuencias de la tecnología..." (p. 187). Leyendo este capítulo se cristaliza el lado oscuro de la IA y, de consecuencia, la del ser humano. Por ejemplo, los sistemas como COMPAS y HART re-

presentan los fallos en el diseño de la IA, pero también son ejemplos de los fallos humanos –¡de los profesionales del Derecho por cierto!– en utilizarlas sin cuestionarlas, sin poner en duda su funcionamiento, sin titubear ni un momento si estos sistemas pueden perjudicar a las personas reales. También hay que tener en cuenta que dichas personas perjudicadas están sujetas a estos juicios –no solo por el juez humano sino también por el sistema de la IA– no solo por las decisiones propias, que también, sino por los fallos del sistema de asistencia y servicios públicos en general. Los fallos de la sociedad son también, y no solo, fallos algorítmicos.

Seguramente necesitaremos años y años de alfabetización tecnológica y sensibilización de la sociedad para abandonar de una vez por todas las falsas creencias de que la tecnología (IA incluida) es neutra, porque nunca lo es; que la tecnología sirve a todos de igual manera; que es solo una herramienta para ayudar a las personas en sus tareas y nada más; que nunca se equivoca y siempre tiene los datos exactos, correctos y precisos para llegar a las decisiones; qué vulnerabilidad social es solo la cuestión de la brecha digital y no un problema más profundo que va más allá de lo digital, entre otras.

Así llegamos a las consideraciones sobre el estado algorítmico (algoritmocracia) y el *novo homo ludens* que intercambia las comodidades proporcionadas por el uso de dichos sistemas con sus derechos, sobre todo el derecho a la privacidad y protección de datos. Ante esta visión desoladora y pesimistamente realista, podemos decir que de alguna manera la respuesta a esta posibilidad es el enfoque de la UE, que se basa en los valores éticos europeos, derechos humanos, democracia, Estado de Derecho y otros principios, que encontramos siempre en cualquier herramienta de *soft* o de *hard law* sobre estos temas. Y si bien este enfoque tiene sus puntos débiles, ¿qué opciones o alternativas desde el punto de vista ético-jurídico tenemos? Parafraseando a Winston Churchill –quién sobre la democracia ha dicho que es el peor sistema del gobierno, a excepción de todos los demás que se han inventado y que son mucho peores–, hay que admitir que la visión de la UE es la peor forma de gobernanza de la IA, con la excepción de todas las demás que existen (o ni siquiera existen).

El epílogo se merece una reflexión aparte: ahí volvemos a los orígenes, a Ulises y a Homero, para una vez más confirmar su atemporalidad y relevancia para los tiempos actuales, donde todos somos Ulises expuestos a las sirenas digitales y todos luchamos para encontrar la manera de resistir. El verdadero problema son aquellos que no se oponen a las sirenas: pero Homero

ya nos ha contado lo que pasa con ellos y Fernando H. Llano Alonso nos lo recuerda de nuevo con este libro, que espero se leerá y se discutirá no solo en las aulas de filosofía del Derecho de nuestras universidades, sino también en otros foros.

Espero también que el autor esté preparando el siguiente volumen de *Homo* y la pregunta es: ¿de qué tipo de *Homo* hablaremos? No lo tengo claro, pero estoy segura de que el autor nos proporcionará muchas ideas, que nos dará esperanza en el ser humano y no le quitará el papel central en la escritura de su futuro.

MIGLE LAUKYTE
Universitat Pompeu Fabra
e-mail:migle.laukyte@upf.edu