

EL DERECHO ADMINISTRATIVO EN LA ERA DE LA INTELIGENCIA ARTIFICIAL

Suzana Tavares da Silva

(Versión provisional)

RESUMEN: La revolución tecnológica que supone la introducción de la Inteligencia Artificial de forma cotidiana en la vida privada, social y política se presenta como una amenaza más compleja y preocupante que las que pudieron representar las anteriores revoluciones industriales y tecnológicas. Y ello, esencialmente, por su carácter disruptivo. La diferencia relevante que se da entre estas revoluciones es que los efectos de la Inteligencia Artificial no se circunscriben solamente a la economía, sino que se extienden de forma más directa sobre el comportamiento humano e implican un cambio en el modo en el que las personas perciben la realidad y sus propios valores. Esto hace que haya un riesgo de que se puedan perder los referentes existentes, que son los fundamentos de nuestra convivencia, que acabarán afectando al Derecho y, por supuesto, también al derecho administrativo.

PALABRAS CLAVE: Inteligencia artificial; futuro; disrupción social

SUMARIO: 1. Los conceptos. 2. El estado del arte en materia de regulación de la IA. 3. Los objetivos de la regulación. 4. Algunas reflexiones en torno a dos ejemplos de aplicación práctica de IA. 5. Reflexiones finales.

1. Los conceptos

El propósito de la *inteligencia artificial* (IA) es, esencialmente, el de crear una máquina con las capacidades del cerebro humano lo que significa que sean capaces de solucionar cuestiones, representar conocimientos mediante conceptos, razonar, aprender con lo nuevo y desconocido y moldear la realidad a las necesidades existentes¹. Como tal, es un sueño o un objetivo de la humanidad que se puede comparar con el afán de dominar la naturaleza o conquistar el espacio mediante el uso de la tecnología o de controlar la vida humana a partir de los conocimientos de la genética y de la biología. El avance de la ciencia ha contado siempre con limitaciones éticas para garantizar que los beneficios de los progresos tecnológicos se armonizan con los valores que se han desarrollado para garantizar nuestra convivencia colectiva, lo que ha hecho que surjan nuevas

¹ Russel and Norvig (2021).

ramas de la ética como la bioética, la ecoética, la ética del espacio y ahora también una ética de la IA².

Hasta ahora se habían desarrollado instrumentos que parcialmente realizaban actividades propias del cerebro humano, pero que no lo hacían de forma integrada. La IA, que se va a abordar en este trabajo, constituye en realidad una etapa más avanzada de lo que se conoce como la cuarta revolución industrial, la revolución digital³. Y no es poca cosa, porque amenaza con cambiar de forma profunda nuestra actual vida en comunidad.

En primer lugar, no se debe olvidar que la IA –a pesar de las contribuciones de las neurociencias, de las ciencias sociales, de la medicina, de la filosofía y del derecho– es un concepto de la ciencia computacional. Es el resultado en el que han confluído desarrollos realizados en áreas como la de los *algoritmos*, la *ciencia de datos*, el *aprendizaje de las máquinas*, el *big data* y la *producción distribuida de datos*.

El *algoritmo* es una secuencia de acciones ejecutables, normalmente de forma automatizada, que permite obtener un resultado. Para explicar un algoritmo se utiliza en general la metáfora de la cocina en la que el algoritmo corresponde a los pasos que hay que seguir en una receta que permita transformar los alimentos en nuestro plato favorito. Si seguimos con esta metáfora podríamos concluir que la Thermomix sustituiría a un *chef*, o, por lo menos, a un buen cocinero, cuando el algoritmo, garantizando un comportamiento estándar, sea eficiente y preciso en los resultados. Esta sería, además, una fórmula para *democratizar* el acceso a los platos más exquisitos. Bastaría con disponer de los ingredientes necesarios y que una persona creara el *algoritmo* (la receta) que luego se ejecutaría uniformemente por todos los autómatas. Pero el resultado, por lo menos en la cocina —para seguir con la metáfora—, no ha sido precisamente este. Los cocineros tradicionales que después se han convertido en grandes *chefs* no cocinan los platos favoritos como siempre, sino que hacen reinterpretaciones de las recetas con nuevas técnicas e introducen otros ingredientes, con lo que el resultado es mucho más caro. Hoy existen en todas las grandes ciudades restaurantes con muchas estrellas, que tienen una identidad cultural propia –la cultura del lugar, cuando existe, es distinta– que han hecho que surja una nueva “*aristocracia gastronómica*” y a la postre que haya una menor identidad cultural gastronómica común. Muy lejos, por lo tanto, de convertirse en una democratización de las recetas tradicionales que más gustaban.

² Dubber, Pasquale, Das (2020).

³ Los autores consideran que esta revolución industrial –de la que, además de la IA forman parte la denominada internet de las cosas, la biotecnología, la computación, la transición energética, etc..., se distingue de las anteriores, precisamente, porque su objetivo es un cambio en la humanidad, del modo como percibimos la realidad, vivimos, trabajamos y nos relacionamos con los demás (Schwab (2016)).

La *ciencia de datos* es otra área de las ciencias digitales en las que se estudian modos cada vez más sofisticados de producir información a partir de información agregada de datos, que a su vez se pueden ramificar en operaciones como la minería de datos (el diseño de procesos de recogida, almacenamiento y utilización para que se obtengan datos de mayor calidad), el *big data* (los procedimientos de gestión de un gran agregador de datos) y la *gestión de datos* (la protección de la información y el conocimiento que se pueden extraer de los datos).

A estos elementos se ha añadido desarrollos tecnológicos en minería de datos, entre los que destaca especialmente la *biometría* (los datos que genera o que se pueden generar a partir del cuerpo humano, como el ADN y las huellas digitales de la retina, las expresiones faciales y la voz), la mejora en estadística (estudio de probabilidades y ocurrencia de fenómenos), la valorización de las ciencias del comportamiento (psicología o antropología) y los nuevos hallazgos de la ciencia sobre el funcionamiento de la mente humana.

Sumados todos estos avances ha sido posible que se creen ahora algoritmos, que son procesos de ejecución automatizada, que a partir de una información correctamente seleccionada y con instrucciones que mimetizan los procesos del cerebro humano, pueden generar de modo automatizado predicciones, explicaciones y recomendaciones. Estos resultados son los que actualmente conocemos como *inteligencia artificial*.

Aunque en la actualidad se ha ido un poco más lejos. Como alertaron Kissinger, Schmidt y Huttenlocher certeramente en su obra de 2021, *The Age of IA* hasta un determinado momento se creía que la *inteligencia artificial* se desarrollaría a partir de la calidad del *input* de los datos y que los resultados se podrían siempre enmarcar dentro de respuestas a cuestiones u objetivos que se tenían que plantear al principio. Se pensaba que la IA no pasaría de ser una “capacidad de computación”. Es decir, que se iba a limitar a procesar información (*inputs*) con una capacidad superior a la humana. Posteriormente se han desarrollado tres experiencias que han demostrado que la IA no tenía necesariamente que ser así, sino que puede ser mucho más creativa cuando se asocia a procedimientos de aprendizaje de máquinas (*machine learning*).

El primer experimento revolucionario fue el *AlphaZero* de *Google DeepMind*, que consistía en un programa de juego de ajedrez al que se le había simplemente dado como *input* las reglas (y no secuencias de juegos bien sucedidas) y la orden de que ganara las partidas. Luego, el algoritmo, que se fue desarrollando solo, se ha convertido hoy en un adversario prácticamente imbatible para los humanos. Pero lo más interesante, no es que haya desarrollado secuencias de juego muy efectivas, sino que ha partido de esquemas de decisión que son distintos a los que desarrolla la mente humana, utilizando una lógica distinta. Este experimento ha puesto de manifiesto la diferencia entre un *algoritmo clásico*, que constituye una secuencia optimizada de pasos para llegar a un resultado, que es muy potente pero que se encuentra limitado en sus resultados y

objetivos, y un *algoritmo de aprendizaje*, que permite mejorar resultados imprecisos, esto es, con los que se pueden lograr resultados sin que existan unos *inputs* dirigidos a que se produzcan. No se trata de procesar información, sino de llegar a resultados a partir de representaciones, que son procesos de conocimiento que se asemejan a los del cerebro humano.

El segundo ejemplo que estos autores destacan es el del descubrimiento de una molécula para luchar con las bacterias multirresistentes. La IA, en esta experiencia, seleccionó entre 61.000 posibles muestras, la que podía ser la más eficaz, después de haber sido “cargada” con todo el conocimiento disponible y haberle pedido que hiciera esta selección. Se trata de una actuación que no hubiera podido hacerse utilizando el método científico tradicional o que hubiera sido sumamente costoso y largo en el tiempo. En este experimento intervino un *algoritmo de aprendizaje no vigilado* (al que no se le había impuesto ningún *input* basado en reglas de conocimiento anteriores), que hizo posible relacionar de una manera innovadora, siguiendo un camino “no convencional”, a la molécula que se buscaba con sus efectos sobre las bacterias.

Y el tercer ejemplo, más distópico todavía, fue el del *CPT-3 (Generative Pre-Trained Transformer* de tercera generación), que consiste en un generador de texto especulativo de “tipo conversacional humana” que funciona a partir de un tópico o de una cuestión incompleta, que se se obtiene tras haber consultado toda la información que se encuentra disponible *online*.

A partir de estos tres ejemplos se pueden extraer dos conclusiones importantes. La primera es que el aprendizaje de las máquinas (*machine learning*), de los algoritmos, es muchísimo más rápido que el aprendizaje de los humanos. La segunda, es que la IA acede a la realidad y a su comprensión de un modo distinto a como lo hacemos los humanos y que encuentra respuestas que los humanos no podrían alcanzar. Como explican Kissinger, Schmidt y Huttenlocher el problema no se encuentra en los cambios que se puedan generar con esta nueva tecnología –es una experiencia que ya se ha vivido antes con la máquina de vapor o el motor de combustión–, sino en el cambio filosófico que la IA representa, por el hecho de que la máquina haya superado a la inteligencia humana o pueda hacerlo muy rápidamente si hay éxito en el desarrollo de las *Generative Adversarial Network*, con sus redes neuronales y se consigue avanzar en una *General Artificial Intelligence*) o, peor aún, en que la máquina puede llegar a *manipular* a los humanos, dado que la realidad de la IA no es la misma realidad que la existe en los entornos humanos: Circunstancias que no se han dado en las anteriores revoluciones industriales o tecnológicas.

Si a todas estas circunstancias le añadimos todas las indefiniciones que ya se conocían de la física cuántica –como el famoso experimento del gato de Schrödinger– que transmiten una incertidumbre en relación con la verdad o la realidad de algo cuando no se cuenta con métodos científicos de comprobación; que nuestra mente es limitada en su modo de percibir la realidad si se la compara con una IA, y, por último, que las tecnologías actuales permiten desarrollar

imágenes, sonidos y videos falsos (como hace la tecnología de la red social *Tik Tok*) que no podemos saber si son verdaderos sin la ayuda de herramientas de comprobación, tenemos como sociedad un billete para entrar rápidamente en lo que algunos llaman la *sociedad de la pós-verdad*⁴.

Hoy el reto más importante se encuentra en la transformación que la *digitalización* produce *entre y en los humanos*. Los humanos somos herederos, en general, de la cultura Occidental. Por esta razón ajustamos nuestros comportamientos a las convicciones que se han venido formado a partir del conocimiento, que no es más que el resultado de la representación que cada uno hace de la información adquirida en su contexto cultural. Los humanos hacemos una representación de la verdad, de la virtud y de la realidad que se basa en la ética y que tiene asumido que la aprobación de sus comportamientos, para ellos mismos y para la sociedad, se apoya en estos fundamentos éticos. Investigaciones recientes que se han realizado en el ámbito de los comportamientos humanos han mostrado que la *generación digital* es distinta y que no tiene esa misma representación de las cosas basada en la ética ni que tampoco adecuan sus comportamiento a estos fundamentos. Es una generación que busca la aprobación de sus acciones de forma inmediata por los “*otros digitales*” más que por su propia conciencia. Es más, no tienen ni siquiera sus propias convicciones, sino que, en su mayoría, se apoyan en opiniones cambiantes, que se basan en los *inputs* que reciben de manera inmediata de una realidad que solo conocen parcialmente. Porque la IA se encarga de gestionar el acceso que tienen a la información de acuerdo con sus intereses, lo que les hace vivir en una burbuja y les impide enfrentarse a lo que es distinto o desconocido, que resulta esencial para su pleno desarrollo como humanos. Muchos autores defienden que nos encontramos ya con una revolución social en curso, parcialmente ya en marcha, de la que nos estamos dando cuenta demasiado tarde.

Más inquietante todavía es la *generación IA* que se está desarrollando en algunos países. Son niños que empiezan su vida relacionándose de “*forma igual*” con humanos y con herramientas de inteligencia artificial, que les cuidan y vigilan desde pequeños, enseñándoles a hablar, ayudándoles con los deberes del colegio y que puede llegar a producir unos efectos futuros que aún se desconocen, aunque a lo mejor un algoritmo sí!.

Una vez precisadas las cuestiones que se van a tratar hay que destacar que, al igual que sucedió en anteriores revoluciones, se está produciendo una transformación muy profunda en la vida colectiva de las sociedades y que, nada más que por eso, va a tener unos impactos profundos sobre la economía (previéndose que un enorme número de profesiones que hoy son muy valoradas por la sociedad puedan ser substituidas por herramientas de IA), la política (el modo en el que se ejerce influencia para obtener la confianza de los representados y se les rinde cuentas va a ser

⁴ McIntyre (2018).

muy distinto, como distinta va a ser la forma en la que se negociará entre Estados en el plano internacional) y también en las relaciones interhumanas.

Y, por último, hay que advertir de los riesgos de que la IA pueda llegar a ser utilizada para fines perversos⁵ lo que obliga a buscar una regulación, lo que constituye el principal motivo de reflexión de este estudio, que habrá de encuadrarse fundamentalmente en el ámbito del derecho administrativo. En concreto, se van a buscar respuestas a cuestiones como ¿qué regular?, ¿quién regula?, ¿quién controlará esta regulación? y, por último, ¿quién controlará a los reguladores? Lo que es indudable, en todo caso, es que la regulación ya va a llegar demasiado tarde porque es una realidad que se ha adelantado.

2. El estado del arte en materia de regulación de la IA

Tras la generalización del uso comercial de Internet, después de los años 90, se ha asistido a una expansión muy rápida de un nuevo poder, el de *las big tech*. Este *nuevo poder* se basa en que estas empresas constituyen una fuente de concentración de *poder económico* que obtienen ingresos muy elevados con la publicidad dirigida, porque esta publicidad es efectiva para influir sobre la demanda en todos los mercados, y, lo que resulta más relevante, en que son una *fuerza de poder sobre la sociedad*, porque tienen capacidad de influir sobre los comportamientos humanos a escala casi planetaria e incluso de interferir en los parámetros de la verdad, de la justicia o de lo que se considera correcto. Sería como si las normas jurídicas dieran lugar a una especie de “algonormas”, esto es que los comportamientos se produjeran con arreglo a reglas dictadas por un poder que no son las instituciones político-democráticas, sino por algoritmo sin rostro⁶, dando lugar a una despolitización de la vida en sociedad⁷.

Esta nueva realidad del *poder* ha despertado reacciones muy diferentes en los tres principales bloques económicos mundiales.

En Estados Unidos se sigue confiando en el poder de autorregulación del mercado y en que los responsables de cada una de estas empresas (*Amazon, Apple, Google, Meta, Microsoft*) se controlan mutuamente. Por lo que sería suficiente con el régimen existente de regulación de la libre competencia. Además, se confía a la *técnica* la definición de los *estándares de seguridad* del uso de la IA, configurando los parámetros necesarios para disponer de una IA fiable de forma que cumplan con unas exigencias que haga que sus sistemas sean discernibles, fiables, resilientes, objetivos, seguros, explicables o interpretables (que se pueda entender por los

⁵ La IA es un ámbito típico del riesgo (Kusche, 2023, 143).

⁶ Hyden (2020), 432.

⁷ Kusche (2023)157.

humanos), saludables (que no generen riesgos para la salud y para la integridad física), garanticen la privacidad y permitan la imputación de responsabilidad. Corresponde a la ciencia computacional, por tanto, desarrollar estos parámetros de legitimación para el uso “comercial” de estas tecnologías que, de alguna manera, se acercaría a una ética propia basada en cuatro elementos: *i) accountability*, que garantice la operabilidad permitiendo la imputación de resultados a efectos de responsabilidad; *ii) reliability*, que permita producir resultados correctos, con calidad y que incorporen una representación de la sociedad; *iii) efficiency*, que produzcan unos resultados que no se puedan alcanzar con otras técnicas no robotizadas; *iv) monitoring*, que permitan el uso de herramientas de control, igualmente robotizadas, como si se tratara de un “contra-algoritmo” que asegure confianza en los resultados producidos. Esta “confianza” implica la necesidad de realizar inversiones masivas en capacidad de computación que permitan mantener el liderazgo tecnológico de la IA, e, igualmente, conlleva fuertes exigencias en materia de seguridad⁸.

En China todas las empresas tecnológicas están controladas por el poder político del Estado (*Baidu, Alibaba, Tencent y Xiaomi*) y son utilizadas por este como herramientas de su política global. Aunque también se haya discutido allí la necesidad de contar con una regulación para la IA.

Ya en la Unión Europea, donde el poder de las tecnológicas norteamericanas está muy presente y no existe una tecnología propia, la discusión sobre la necesidad de un régimen público para la IA, no solo para regular la tributación de las actividades de estas empresas tecnológicas (sobre todo la publicitaria) que dan lugar a la acumulación de mucha riqueza, se ha centrado en garantizar que el uso que se pueda hacer de la IA no viole los estándares de garantía de los derechos fundamentales.

La Comisión Europea presentó, en abril del 2021⁹, una propuesta de regulación integral para la IA¹⁰. Se trata de una regulación, calificada como neutral, que se basa en una calificación de distintos niveles de riesgo. El objetivo de este nuevo régimen jurídico es el de permitir un desarrollo de la IA que haga posible que se generen beneficios económicos y sociales, pero que impida que se pueda utilizar para finalidades ilegítimas y en la que se asegure también que su funcionamiento va a respetar la matriz cultural que supone para Europa la salvaguarda de los derechos humanos¹¹. No se trata de regular los problemas jurídicos que ya se han abordado sobre

⁸ Taylor (2021).

⁹ COM(2021)206.

¹⁰ Desde Estados Unidos, la regulación de protección de datos y la regulación de la IA en Europa se perspectiva como una “constitución digital” – Bradford (2023).

¹¹ La regulación europea propuesta para la IA transpone para el plan normativo de regulación lo que ya se venía planteando en instrumentos de *soft law*: *i)* Comunicación “Inteligencia artificial para Europa” [COM(2018)237 final]; *ii)* Directrices Éticas para una IA fiable, 2018; *iii)* Recomendaciones políticas y de

la personalidad jurídica de los algoritmos¹², la responsabilidad civil de los autómatas¹³, los derechos de propiedad intelectual para el desarrollo de la IA¹⁴, el uso o prohibición de las tecnologías de IA en investigación criminal y policial¹⁵, o en educación, cultura y sector audiovisual¹⁶, o en actividades civiles y militares del Estado¹⁷, sino que lo que ahora está en juego es regular la prohibición y la admisión condicionada o incondicionada del uso de las tecnologías de IA según la calificación que se haga del riesgo asociado al uso de cada tipo o subtipo de esta tecnología.

De esta forma se contemplan cuatro niveles de calificación de riesgo: *i)* tecnologías de *riesgo bajo o mínimo*, que se pueden desarrollar sin condicionantes u obligaciones legales, aunque se recomiende la aplicación de códigos de conducta que proteja de estas tecnologías, que hoy se valora como no amenazadoras; *ii)* tecnologías de *riesgo limitado*, a las que se le imponen obligaciones de *transparencia*¹⁸; *iii)* tecnologías de *riesgo alto* que exigen una regulación y en las que se incluyen todas las que desarrollan herramientas para integrar elementos que están protegidos por una regulación armonizada de salud y seguridad (juguetes, transportes, dispositivos médicos, ascensores) y las que se encuentren entre las áreas calificadas como tal por la Comisión (identificación biométrica y categorización de personas, operación de infraestructuras críticas, educación, empleo, servicios públicos y servicios esenciales, aplicación coactiva de la ley, inmigración y asilo, administración de justicia y procesos democráticos) para las que se exige un control *ex ante* (registro, autoevaluación, heteroevaluación y sello de conformidad) y se establecen obligaciones *ex post* de evaluación o monitorización; y *iv)* tecnologías de *riesgo inaceptable*, que estarían prohibidas, como sucede con sistemas que puedan conllevar técnicas de manipulación, que exploten grupos vulnerables, que sean utilizadas por las administraciones públicas para su beneficio como elemento de promoción social y sistemas de identificación biométrica en tiempo real en espacio públicos para imponer el cumplimiento de la

inversiones para una IA fiable, 2019; *iv)* Libro Blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza [COM(2020) 65 final].

La OCDE también ha emitido “Recomendaciones para la IA”, primero en 2019 y en 2021.

¹² Los autores explican el riesgo que comporta la propuesta de aplicar la personalidad jurídica de las personas jurídicas a los instrumentos de IA, porque en este caso no estamos tratando de una organización de humano que actúan mediante esas personas jurídicas, pero con una racionalidad típica de los humanos. En el caso de la IA se hablaría de una entidad sin ningún sustrato humano que actúa con una racionalidad no humana, debiéndose reservarse la personalidad para expresiones de la humanidad (Yampolskiy *in* Thomson, 2021, 1-11).

¹³ De la propuesta de Directiva relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial [COM(2022)496 final]

¹⁴ Resolución del Parlamento Europeo, de 20 de octubre de 2020 [2020/2015(INI)]

¹⁵ Resolución del Parlamento Europeo, de 6 de octubre de 2021 [2020/2016(INI)].

¹⁶ Resolución del Parlamento Europeo, de 19 de mayo de 2021 [2020/2017(INI)].

¹⁷ Resolución del Parlamento Europeo, de 20 de enero de 2021 [2020/2013(INI)].

¹⁸ Se incluyen en este grupo sistemas como *chatbots*, *sistemas de reconocimiento de emociones*, *sistemas que utilizan biometría* y *sistemas que permitan manipular imágenes, audios o videos y crear (deepfakes)*.

ley (con excepción de los casos de terrorismo, protección de menores o persecución de sospechosos de crímenes).

Estamos ante una nueva área del derecho administrativo europeo. La aplicación de este régimen jurídico precisará de la existencia de una autoridad nacional de supervisión, que actuará, igual que sucede en otros dominios regulatorios armonizados por la Unión Europea (UE), en red con sus congéneres de los otros Estados miembros. El cumplimiento de las reglas en las actividades de riesgo alto y limitado quedará a cargo de las autoridades de regulación de los respectivos mercados y su violación dará lugar a la aplicación de sanciones administrativas.

Se trata de una propuesta de regulación que ha sido muy criticada por algunos sectores porque consideran que con ella la Unión Europea va a quedarse fuera del progreso que representa la IA, que le supondrá un retraso tecnológico y que puede comprometer la innovación y el tejido empresarial de las PYMEs (que constituyen el tejido empresarial típico de la UE), porque estas no podrán soportar los costes que supondrá el cumplimiento de estas medidas regulatorias. Otros autores, sin embargo, son muy favorables a este nuevo marco regulatorio porque creen que –al igual que ha sucedido con el régimen del reglamento general de protección de datos (Reglamento 2016/679)– la nueva regulación puede ser útil para suplir el déficit regulatorio que existe en Estados Unidos y que puede servir para crear un estándar regulatorio relevante en el plano internacional que garantice la armonización de la IA con el sistema de protección de derechos fundamentales y con la dignidad de la persona humana.

3. Los objetivos de la regulación

Volviendo a las cuestiones que se han planteado al principio –y sin tener ninguna seguridad acerca que se hayan tenido en cuenta efectivamente todos los ejes relativos a la IA, ya sea por su carácter dinámico (el proceso de la IA evoluciona a ritmos de gran incertidumbre, a veces muy rápido y a veces con bloqueos que ralentizan su progreso), ya sea por su carácter hermético para el derecho– habría que destacar dos cuestiones problemáticas: que la IA es una amenaza actual y real para la existencia humana tal como la conocemos hoy que demanda una intervención política; y que su regulación por parte del poder político debe asegurar el principio de la humanización de ciertos niveles o tipos de decisión pública, con especial consideración sobre las áreas propias de los servicios públicos, donde nuevamente el derecho administrativo tiene mucha relevancia.

La *regulación pública* de la IA como instrumento de *moderación* o *gestión* de un riesgo que amenaza a la existencia humana es algo para lo que el mundo parece haber despertado en el mes de noviembre de 2023. Las preocupaciones relativas a la regulación de la IA han salido ya de los foros científicos, de las publicaciones especializadas y de las páginas secundarias de los

periódicos, saltando a las primeras páginas de las publicaciones más difundidas tras la “crisis en la OpenIA”.

OpenAI se constituyó en 2015 en Estados Unidos como una organización no gubernamental dedicada al estudio y al desarrollo de la IA alineada con los valores humanos. Así aparece al menos en su página web. Se trata de una inversión muy importante de capital privado a la que han contribuido sus fundadores, que son mayoritariamente inversores de capital de riesgo: Sam Altman (CEO de la Y Combinator que es una “aceleradora de start-ups”), Reid Hoffman (creador de LinkedIn), Jessica Livingston (también socia de la aceleradora Y Combinator), Elon Musk (CEO de Tesla), Peter Thiel (fundador de Paypal) e Ilya Sutskever (especialista en ciencia de la computación y machine learning).

En 2019 la actividad pasó a tener carácter lucrativo y en la medida que el proyecto se fue desarrollando surgieron colaboraciones con proyectos parecidos de otras empresas, como la DeepMind de Microsoft. En noviembre de este 2023, Sam Altman fue despedido de la OpenIA por, supuestamente, no estar de acuerdo con los recelos que otros miembros del Consejo de Administración habían expresado en relación con los riesgos que la IA representa. Ello, a pesar de que el propio Altman había dicho en mayo de 2023, ante el Congreso de los Estados Unidos, con otros directivos de “laboratorios de investigación en la materia”, que la IA representaba un riesgo para la existencia humana¹⁹ y de que antes en marzo algunos de estos investigadores habían propuesto una moratoria de seis meses en el desarrollo de las redes neuronales más potentes, precisamente por el riesgo que su desarrollo representa al no tener ningún control de carácter público²⁰.

Lo cierto es que es bastante real el riesgo de que, entre otros peligros, sistemas como el ChatGTP, el Gemini y otros similares puedan empezar a realizar campañas de desinformación que alteren la sociedad a nivel mundial y en el que resulte difícil identificar la verdad para generar confianza. La crisis en la OpenIA se resolvió con el regreso de Altman, pero las alarmas han sonado ya en relación con los peligros que pueden darse con el desarrollo de esta nueva tecnología, más allá de los sesgos que puedan adoptarse en los programas de IA para la gestión de políticas públicas o la utilización que pueda hacerse en actividades muy sensibles. En efecto, habrá que cuestionarse si debe haber o no un control del poder político sobre esta nueva forma de poder.

¿Cuáles serían los objetivos de la regulación pública? Los autores se refieren a la *humanización* de la IA como el objetivo principal de su regulación. Si estamos ante una herramienta que puede estar dotada de capacidad para tomar decisiones de un modo automatizado habrá que garantizar,

¹⁹ V. NYT «A.I. Poses ‘Risk of Extinction,’ Industry Leaders Warn», 30.05.2023.

²⁰ V. NYT «Elon Musk and Others Call for Pause on A.I., Citing ‘Profound Risks to Society’», 29.03.2023.

en primer lugar, que esta herramienta puede *recrear la moralidad humana*, o, por lo menos, que no la vulnera. Es decir, que más que leer lo que pueda ser la ética y los escritos sobre ética es esencial que la IA sea capaz de integrar la ética en su proceso de decisión. Algunos operadores tecnológicos defienden que las herramientas de IA no disponen de intencionalidad y de conciencia porque se trata siempre de autómatas, por muy sofisticados que puedan ser sus algoritmos, y que, por eso, solo se puede hablar de una *ética soft* en este dominio (Kaas *in* Thomson, 2021, 49). Otros defienden que los algoritmos pueden ser programables para desarrollar juicios éticos ante diferentes escenarios, incluso utilizando características humanas como el libre arbitrio, la conciencia o las intenciones (Goram & Veiel *in* Thompson, 2021, 27ss).

La cuestión ha sido especialmente desarrollada en el ámbito de la programación de coches con conducción automatizada y robots que cuidan a enfermos y mayores de edad, en la que los programadores han tenido que recrear decisiones humanas, que pueden incluso conllevar arbitrariedades. El ejemplo más conocido es el de los coches con conducción automática. Los consumidores prefieren, en general, que los coches que se vendan tengan algoritmos en los que esté programado que se sacrificará antes a los ocupantes del vehículo cuando se encuentre en la situación de tener que proteger a peatones que caminan por la acera, pero se decantan por coches con dirección automática que protejan a los ocupantes cuando ellos son los que los compran. Este dilema que puede resolverse a través de medidas regulatorias *top-down*, que implica elegir una regla ética para todos los vehículos, se puede solucionar con un modelo de regulación *bottom-up*, en el que cada conductor podría programar libremente esta variable del algoritmo al inicio de cada viaje, con las correspondientes consecuencias que se podrían producir en el ámbito de la imputación de responsabilidad. Aunque pueda parecer extraña la solución *bottom-up* es la más coherente con lo que ahora mismo existe. Pero el conductor tendría que elegir una opción antes de saber todas las variables del caso. Pudiendo, por ejemplo, optar por proteger su vida, pero qué decisión habría tomado si en la vida real uno de los peatones que podían ser atropellados fuera su propio hijo. ¿Habría cambiado de opción? ¿Será el algoritmo capaz de incorporar todos los elementos de la voluntad del conductor para actuar como él en cada momento?

Pero la cuestión jurídica más relevante para nosotros en este ámbito es otra: ¿Deberá el régimen de imputación y responsabilidad en estos casos seguir las reglas del derecho civil o se tendrá que hacer una construcción jurídica nueva? y ¿las herramientas que trabajan con IA son productos de mercado o necesitan de una autorización pública para poder operar?

El estándar regulatorio de la Unión Europea soluciona, al menos, una parte de estas dificultades ya que no ha renunciado a los principios que inspiran el Reglamento General de Protección de Datos: la *privacidad* y la *soberanía*. La *privacidad* se garantiza no solo con normas jurídicas de conducta, también mediante reglas técnicas sobre construcción y diseño de sistemas de

almacenamiento de datos y de actuaciones con IA²¹. De acuerdo con el principio de *soberanía* la decisión tendrá que estar siempre reservada a los humanos. De esta forma, se garantiza que el control del resultado de la IA se va a hacer según criterios éticos humanos²². En todo caso, los desafíos a los que nos estamos enfrentando ya, y que en buena medida derivan de la permanente interacción entre los humanos y las herramientas automatizadas, obligan a replantearse los fundamentos de la ética. ¿Serán finalmente los autómatas un instrumento para superar las limitaciones éticas de los humanos?

En Portugal, gran parte de estos problemas no están aún en la agenda del legislador, ni tampoco en el debate político o social. Rastreando el derecho positivo, se podría citar la aprobación de la *Carta Portuguesa de Direitos Humanos na Era Digital*, aprobada por la Lei n.º 27/2021, del 17 de mayo, que, más que un régimen jurídico, es una normativa de principios en la que se contienen cuestiones muy distintas como las relativas al acceso a la administración digital y la protección post-muerte de “la personalidad digital” o el derecho al olvido. Puede que el retraso que existe en esta regulación y también en la asunción de este problema se deba a que nuestro país aún se encuentra pendiente de la aprobación del régimen jurídico aplicable al almacenamiento de los metadatos y su relevancia para usos como los de política criminal, que tiene como ejemplo claro la decisión del Tribunal Constitucional n.º 800/2023.

Por último, considero que este estudio no es el lugar, ni tampoco es el momento, para hacer una revisión del concepto de ética, aunque se trate de una cuestión central cómo se va a señalar a continuación. No obstante, estamos seguros de que hay que tener presente el concepto de la ética para identificar los principios y las reglas éticas que deben seguir quienes estén trabajando en el desarrollo de los sistemas de IA y, de lo que no hay tanto literatura, para saber el papel que le está reservado a la regulación pública. ¿Podría esta ser una forma de superar la imperfección humana?

Y hay igualmente muchas dudas sobre los límites de la *ética de la IA*, que es la otra dimensión del problema, para quienes piensan que es posible desarrollar una normativa distinta a la actual²³. Yo diría –probablemente por mi incapacidad de entender los algoritmos– que la UE tiene mucha razón en aplicar, más que el principio de precaución, una regla de prohibición del uso de la tecnología cuando el ser humano no sea *soberano*. Y lo afirmo, aunque no tenga una demostración cartesiana, porque creo que por muy perfectos que estos automatismos puedan llegar a ser y por

²¹ Una técnica regulatoria semejante a la del *ecodesign* de productos de la Directiva 2009/125/CE y de la propuesta de Reglamento llamado a sustituirla [COM(2022)142].

²² Asimov enunció tres reglas para la ética de los instrumentos de IA, como reglas que garantizarían su “no maldad”: *i*) no pueden hacer daño a los humanos, ni permitir que por su inacción ellos sufran cualquier daño; *ii*) deben estar preparados para recibir y obedecer a ordenes de los humanos, excepto cuando esa orden corresponda a una violación de la regla anterior; *iii*) deben proteger su integridad, excepto si esa protección implica la violación de las dos reglas anteriores (Kaas *in* Thomson, 2021, 48).

²³ Chesterman (2021), 173ss.

muy geniales que puedan llegar a estar programados para mimetizar el comportamiento humano, nunca llegarán a generar la *empatía* que es la característica única de la humanidad que dirige, condiciona y explica gran parte de nuestras actuaciones.

4. Algunas reflexiones breves en torno a dos ejemplos de aplicación práctica de IA

El primer ejemplo que quiero analizar es el de la *smart justice* o de la aplicación de la IA al sector de la justicia²⁴. Son muchos y distintos los proyectos ya en aplicación: desde sistemas en los que hay una decisión judicial robotizada, a los que no se prescinde de una intervención humana aunque se utilicen sistemas de “letrados digitales” y otros en los que hay digitalización y alguna *soft intelligence* como apoyo a la actividad de los magistrados. En todos ellos se intenta contradecir la idea defendida por Zalnieriute y Bell de qué la actividad de juzgar sería la única en la que no podríamos prescindir de los humanos²⁵.

Son muchos los desafíos, que ya han sido señalados, que implica la aplicación de la IA a los sistemas judiciales, empezando por la necesidad de la reconfiguración procesal, si es que efectivamente se generaliza su utilización²⁶. De momento, el análisis de las experiencias más avanzadas que se han dado en Australia, Estados Unidos, Reino Unido y Estonia han permitido identificar los siguientes problemas: *i)* la falta de un elemento humano, particularmente en procesos donde la evaluación con parámetros humanos es esencial, como sucede en derecho penal o en el derecho de la familia, siendo algo que se ha discutido mucho durante la pandemia por las limitaciones del uso de los medios audiovisuales durante las sesiones de la práctica de la prueba; *ii)* como afectan a la transparencia y a la confianza en las decisiones porque, por un lado, en un sistema judicial digitalizado hay un acceso total e inmediato a toda la información disponible que, sin embargo, en gran medida no está en lenguaje técnico-jurídico; pero, por otro lado, una parte de las decisiones robotizadas dictadas por los algoritmos no se puedan explicar de acuerdo con la lógica tradicional, particularmente las que se adoptan por intermedio de algoritmos de *deep-learning* y eso puede hacer casi imposible que se puedan recurrir ante magistrados humanos cuando no sean acatadas por las partes; y *iii)* las incongruencias que se puedan dar en las actuaciones judiciales, a las que nos referiremos más detenidamente.

Consciente de buena parte de estos problemas, la Comisión para la eficiencia de la justicia del Consejo de Europa adoptó, en 2018, la Carta Ética Europea para el uso de IA en los sistemas judiciales y sus entornos, en la que se prevén principios como el de la *compatibilidad de la IA con*

²⁴ Sourdin (2021).

²⁵ Sourdin (2021), 212.

²⁶ Zeleznikow (2021).

los derechos humanos, la no discriminación, la calidad y seguridad (la rastreabilidad de las fuentes de información, la certificación de datos y ambientes seguros), la *transparencia, imparcialidad y justicia y el uso controlado*. Más allá de los principios, es importante ver qué usos efectivamente se pueden hacer de la IA en el sector de la justicia para poder buscar, ante todo, en esta tecnología un apoyo para intentar resolver el problema generalizado del retraso en la producción de las decisiones judiciales.

En Australia y Reino Unido ya existe la posibilidad de que se dicten decisiones judiciales de forma totalmente robotizada. Se utiliza para decisiones de primera instancia y en procesos de pequeña cuantía, como expedientes de derecho civil (derecho del consumo) o sanciones administrativas en materia vial. En Australia ya es posible utilizar un expediente sin papel, en el que el demandante expone sus pretensiones en lenguaje común sin necesidad de abogado, el algoritmo notifica al demandado que puede hacer lo mismo y, al final, el mismo algoritmo dicta una decisión que se puede recurrir ante un magistrado humano. Este tipo de decisión puede ser importante en áreas muy masificadas, en aquellas en las que la resolución de la controversia no afecte a cuestiones complejas y que, por esta razón, una vez fijada el alcance de la ley, un algoritmo puede repetirla en infinidad de casos semejantes. Pero esta solución no es válida para decisiones innovadoras, en las que las decisiones anteriores no sirven como parámetro y donde la actividad de concreción del derecho exige que intervengan las ya mencionadas características humanas.

En Estados Unidos se utiliza el modelo de *letrados virtuales* con los que un magistrado puede “discutir” un proyecto de decisión para que la IA le ayude a valorar la “consistencia” de la decisión, señalando divergencias con otras decisiones, que pueden ser anteriores del mismo tribunal, de tribunales superiores o, incluso, de tribunales internacionales que pueden emplearse por el gran avance que se ha producido en las herramientas de traducción, que permite compáralo con decisiones de tribunales supremos de otros países) y que permiten detectar, discrepancias, oscuridades o contracciones que puedan darse en la fundamentación jurídica. Si estas herramientas actúan como efectivos ayudantes para la elaboración de una decisión más correcta desde un punto de vista de los cánones jurisprudenciales (coherencia y claridad de fundamentación, uniformidad con jurisprudencia anterior, agotamiento de las cuestiones planteadas), podrían ser útiles. Pero existe el riesgo real de que esta “tutoría de revisión” pueda presentar sesgos como, por ejemplo, que solo indique la necesidad de revisión o que solo aprecie problemas de fundamentación en determinados casos o cuando la decisión apunte a un determinado resultado.

En efecto, no nos podemos imaginar un sistema en el que estos sesgos no existan, ya que el derecho positivo (la ley) es un producto sociocultural y la jurisprudencia una actividad de aplicación de ese derecho positivo, por lo que las soluciones anteriores no serán siempre un

parámetro de su corrección. Y a eso hay que añadir el hecho de que la decisión judicial es fundamentalmente una decisión de un caso concreto para lo que es esencial entender perfectamente el problema y los argumentos de las partes, sin que se pueda solucionar una cuestión de una forma abstracta. Por esa razón un buen magistrado es el que consigue formular con claridad las cuestiones planteadas por las partes y decidir de forma concreta cómo se solucionan y no, desde luego, el que maneje de una forma sublime las instituciones jurídicas. En esta medida, parece que los algoritmos serían más buenos substituyendo la actividad desarrollada por un catedrático de derecho que la de un magistrado, porque el algoritmo tardará más –si lo logra– en desarrollar esa capacidad humana que ha de tener un magistrado para identificar en el plano jurídico las voluntades, las intenciones y los deseos de las personas en conflicto, incluso cuando se trate de expedientes que se desarrollan de forma escrita únicamente. La actividad judicial exige conocer el contexto moral y cultural en el que se sitúan las partes en conflicto, lo que solo se consigue con la experiencia de la vida, que es un elemento del que la IA carece.

Habría, no obstante, que esperar a ver los resultados de estas experiencias de IA en la justicia para poder saber si la razón artificial sirve para dictar soluciones de interpretación y aplicación del derecho correctas a la luz del método jurídico y que sean también justas en el plano de los principios.

Como reflexión adicional me parece importante que en relación con algunos procedimientos administrativos la intervención de magistrados humanos en su control posterior me parece aún más necesario cuando se trate de procedimientos administrativos que se han tramitado mediante IA, que es ya una realidad muy difundida, ya sea en hacienda, en procedimientos de asilo e inmigración o en otro tipo de actos administrativos masificados. Para el control de estas decisiones es fundamental un *gatekeeper humano* que pueda neutralizar, más que errores en la interpretación y aplicación de la ley a los supuestos del caso concreto, los sesgos y las injusticias que pueden estar detrás de las decisiones robotizadas y la forma en la que resuelven las cuestiones donde haya discrecionalidad.

Por otro lado, es razonable que estas experiencias avancen en áreas como el derecho civil y en aquellas parcelas jurídicas en las que ya se admite el recurso a medios alternativos para solucionar disputas, como el arbitraje. Esta es, por ejemplo, la experiencia que se ha dado en Estonia en materia de contratos civiles. La situación es distinta en el ámbito del derecho administrativo donde resulta muy complejo integrar en un algoritmo el juicio del control de la discrecionalidad.

Como estas novedades se están expandiendo a un ritmo acelerado, lo mejor será que se creen herramientas de carácter macro, de análisis de información agregada (cuantificación de expedientes sobre determinado tema o sobre determinada normativa, se creen mejores motores de búsqueda de decisiones anteriores, de otros países o de tribunales internacionales) y que su

utilidad se vincule a la gestión procesal de la organización de la actividad jurisdiccional en procesos en masa, pero que no se centre en el apoyo a decisiones concretas de los expedientes. Pero incluso en este ámbito de la IA aplicada a la *judicial analytics*, muy difundida en Estados Unidos, han surgido preocupaciones, como se ha visto con la medida adoptada en Francia en 2019 que ha prohibido el uso de software de análisis de las decisiones judiciales y que incluso se plantea que se pueda considerar como delito que se utilicen para prever el sentido de decisiones judiciales futuras.

Hay también ejemplos de IA que utilizan los abogados cuando pleitean ante los tribunales. En este caso las herramientas de IA se pueden utilizar para detectar lagunas jurídicas y contradicciones en la legislación vigente, que, como muchos señalan, son herramientas que ya son utilizadas en algunas áreas jurídicas como el derecho mercantil, el derecho de los contratos y el derecho tributario para sacar beneficio de estos *huecos y disfunciones* del ordenamiento que permita obtener resultados lícitos, aunque sean contrarios al intereses nacional (Ej. Buscar una solución fiscal que permita una reducción en el pago de impuestos). Como respuesta a estos riesgos se ha propuesto que los Estados creen servicios de IA de “legal safety”, que operarían en un modo semejante a los sistemas de ciberseguridad pero aplicado al sistema jurídico, señalando esos *huecos regulatorios*.

Hay otra área en la que también existen experimentos *disruptivos* de la IA como es en la *enseñanza*. Primero, porque la enseñanza ha dejado de ser una vía esencial para la transmisión de información, una vez que la información es algo que hoy está disponible de modo instantáneo a través de motores de búsqueda, Eso es ya hoy una realidad y un problema para la *enseñanza* de los llamados nativos digitales. Los programas de enseñanza, sin embargo, no han cambiado para ajustarse a la accesibilidad de la información en tiempo real. Los responsables políticos en el ámbito educativo no se han dado cuenta aún de que estamos ante un cambio cualitativo parecido al que sufrió la sociedad con la invención de la imprenta y la producción de libro. En aquel momento la enseñanza pasó de ser una difusión de información a la que solo algunos accedían de forma directa a otra en la que todos pudieron acceder a los originales. Y este cambio fue el que permitió el desarrollo filosófico que es la base de nuestra civilización, del Renacimiento, del Iluminismo y de la reforma en la Iglesia que dinamizó a la difusión de las ideas y, de esta manera, a la *enseñanza universal*²⁷.

Hoy estamos ante un nuevo momento fundacional de la *enseñanza*: ¿Qué sentido tienen y tendrán las investigaciones de *master* y doctorado? ¿Qué sentido tiene exigir como habilitación para una carrera de investigación científica la capacidad de leer y entender fuentes escritas y, a partir de ello, hacer un escrito original con una idea nueva, si las herramientas de IA lo van a hacer mejor?

²⁷ Kissinger, Schmidt y Huttenlocher (2021) 41.

¿y si necesitamos de una herramienta de IA para saber si el texto que estamos discutiendo es un original del candidato o resultado de la IA, sería esa tesis mejor que la que haría la IA? Del mismo modo que las Universidades se han construido con el objetivo de difundir conocimiento y formar las elites de una sociedad, habrá que preguntarse ahora qué sentido tiene la Universidad en el mundo de la IA, o de qué forma se puede acomodar la IA a la enseñanza universitaria, sabiendo que es una herramienta que sustituye o tiene capacidad de sustituir a los mejores pensadores. Habrá que definir un nuevo sistema de formación de la educación superior en ciencias sociales con un nuevo marco para la enseñanza teniendo en cuenta la irrupción de la IA.

5. Reflexiones finales

La IA en términos jurídicos y de impacto en la vida social no es solamente una tecnología, sino que es una herramienta que va a influir en los comportamientos de grandes masas de población (con la consiguiente posibilidad de manipulación social) con un alcance muy superior a las de las dictaduras políticas de mediados del siglo XX.

Para entender el riesgo potencial que tienen hay que tener en cuenta datos muy relevantes, como el de que los usuarios de plataformas y redes sociales son mucho más que la población de los países donde están operativas y que su control tecnológico está en manos de un número muy reducido de personas, cuyos intereses no son siquiera controlados, y que van desde el ánimo de lucro (como Facebook, Google o Uber) a objetivos políticos (como las chinas Baidu o WeChat). Las reglas que gobiernan esta vida digital no son democráticas y no tienen la densidad de las que aprueban nuestros parlamentos, a pesar de que muchas personas reparten su vida entre el mundo real y el digital, entre la vida en sociedad reguladas por reglas que incorporan valores culturales y la vida en redes digitales que no tiene estos filtros; y eso cambia su personalidad, cambia el modo en cómo se relacionan con los otros y está erosionando, a una velocidad record, los valores de nuestra civilización como la tolerancia, el respeto, la democracia y la solidaridad social, especialmente en Occidente que lleva décadas intentado mejorar.

El cambio más profundo se encuentra, en mi opinión, en la ética y en el modo en cómo nos relacionamos con esas reglas en el plano personal y social. Voy a poner un último ejemplo: el derecho tributario y la obligación de pagar impuestos es algo que está presente desde que nos hemos organizado como comunidad y es una herramienta fundamental para atender a los costes de los servicios públicos y del Estado, que incluye al gobierno y la administración, la seguridad, la defensa o la resolución de conflictos como el servicio de justicia que es el primer bien público de una comunidad política organizada. Pero sabemos que como supone una pérdida de patrimonio para el contribuyente, este siempre intenta pagar lo mínimo posible, aunque disfrute y valore

aquellos servicios públicos. Todo el derecho tributario está construido sobre la idea de que la obligación de pagar es un deber fundamental que el Estado impone por la fuerza, que la cuantía de los impuestos es una cuestión política e ideológica, cuya definición se legitima y se fija por el parlamento, y que los conflictos entre el órgano que impone el cumplimiento de la obligación tributaria y el contribuyente se solucionan en los tribunales. En este juego de intereses, en el que los contribuyentes intentan pagar lo mínimo posible y los órganos del Estado buscan corregir excesos y fraudes, existe un equilibrio gobernado por espacios de libertad económica y por las reglas del debido procedimiento, que los tribunales igualmente están obligados a garantizar. En los últimos años, hemos asistido a una fuerte digitalización de los servicios de hacienda, recurriendo incluso a la IA, cuyo objetivo claro es aumentar los ingresos públicos. No hay duda de que la finalidad –así enunciada– es legítima, pero lo que no es legítimo es que esa finalidad se imponga desconociendo las reglas del debido procedimiento o que se pueda alcanzar sin respetar derechos fundamentales como la privacidad. Por mucho que la falta de ingresos de los Estados sea un problema grave y se diseñen estrategias para combatir este fenómeno –ya sea con pedagogía (educación para la ciudadanía), ya sea mediante esquemas de compensación (premios para quien cumple correctamente sus obligaciones) o regulatorios (incrementando las retenciones fiscales)– no hay un consenso social, ni hay una habilitación constitucional, para que los servicios de hacienda puedan actuar como un *Big Brother* de los contribuyentes o, peor aún, como un adivinador de los impuestos que debemos de pagar antes incluso de que se haya producido el hecho imponible.

Regresando al tema de la ética, es importante recordar que el individuo –que nosotros– somos seres imperfectos y que uno de los rasgos de *humanidad* es exactamente el perfeccionamiento que vamos buscando a lo largo de nuestras vidas, con las experiencias que acumulamos con nuestros aciertos y nuestros errores y el modo en que todo eso cambia nuestra percepción de los demás, de las cosas y de nuestras ideas sobre la verdad, lo correcto y lo justo. La IA no es humana en este sentido y no veo qué ventaja puede haber en explorar ese camino. Además, como destacaba Stephen Hawking al señalar los peligros de la IA, el cerebro humano está biológicamente programado para desarrollarse hasta un determinado punto y luego extinguirse y la IA tiene un desarrollo ilimitado, también en cuando a sus recursos, una vez que identifica la forma de ser más eficiente para poder “crecer”. Es fácil, por esta razón, que pueda llegar a tener más fuerza que el cerebro humano y que acabe incluso desarrollando una “vocación predatoria” contra el ser humano.

No soy una fanática, desde luego, de la IA, pero tampoco creo que podamos evitarla. Confío en que, al igual que pasó con avances en el conocimiento de la genética, también aquí los individuos seamos capaces de establecer límites al desarrollo y a la utilización de la IA para garantizar que seguimos en un entorno de *imperfecta humanidad*.

BIBLIOGRAFIA

BRADFORD, A. «Europe's Digital Constitution», *Virginia Journal of International Law*, (2023), 64, pp. 1-68.

CHESTERMAN, S., *We the Robots? Regulating AI and the Limits of the Law*, Cambridge University Press, Cambridge, 2021.

DiMATTEO, L., PONCIBÓ, C., CANNARSA, M. (Ed), *The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press, Cambridge, 2022

DUBBER, M. D., PASQUALE, F., DAS, S. (Ed), *The Oxford Handbook of Ethics of AI*, Oxford University Press, ,Oxford, 2020

HYDÉN, H., «AI, Norms, Big Data, and the Law», *Asian Journal of Law and Society*, (2020), 7, pp. 409-436.

KERRIGAN, C. (Ed.), *Artificial Intelligence*, Edward Elgar, Cheltenham, 2022

KISSINGER, H., SCHMIDT, E., HUTTENLOCHER, D., *A Era da Inteligência Artificial*, D. Quixote, Lisboa, 2021

KUSCHE, I., «Artificial Intelligence and/as Risk», *AI – Limits and Prospects of Artificial Intelligence*, transcript, Bielefeld, 2023, pp.143-161.

LIU, A. C., LAW, O., LAW, I., *Understanding Artificial Intelligence. Fundamentals and Applications*, Wiley, San Diego, 2022.

MCINTYRE, L., *Post-Truth*, The MIT Press, Massachusetts, 2018.

RUSSELL, S., NORVIG, P., *Artificial Intelligence: A Modern Approach, Global Edition*, 4.^a ed., Pearson Education, 2021.

SCHWAB, K., *The Fourth Industrial Revolution*, World Economic Forum, Geneve, 2016

SOURDIN, T., *Judges, Technology and Artificial Intelligence*, Edward Elgar, Cheltenham, 2021.

TAYLOR, M., “Is the Internet Trying to kill us?”, *Fletcher Sec. Review* (2021) 102.

THOMPSON, Steve John (ed.), *Machine Law, Ethics, and Morality in the Age of Artificial Intelligence*, IGI Global, Hershey

WHITE HOUSE, *AI Bill of Rights. Making Automated Systems work for the American People*, October 2022.

ZELEZNIKOW, J., « Using Artificial Intelligence to provide Intelligent Dispute Resolution Support», *Group Decision and Negotiation* (2021), Springer, 30, pp.789–812.