VOLUMEN Nº2 JUNIO 2023

Revista de Responsabilidad Penal de Personas Jurídicas y Compliance

¿EL MODELO DE LA RESPONSABILIDAD PENAL DE LAS PERSONAS JURÍDICAS PARA LOS DAÑOS PUNIBLES DERIVADOS DEL USO DE LA INTELIGENCIA ARTIFICIAL?

Prof. Dr. Bernardo del Rosal Blasco

ÍNDICE

¿EL MODELO DE LA RESPONSABILIDAD PENAL DE LAS PERSONAS JURÍDICAS PARA LOS DAÑOS PUNIBLES DERIVADOS DEL USO DE LA INTELIGENCIA ARTIFICIAL?

Prof. Dr. Bernardo del Rosal Blasco *

I. Introducción

La IA habita entre nosotros desde hace muchos años e, incluso, el propio concepto científico también ha estado entre nosotros desde hace más de cincuenta años[1]. Es cierto que, en el último lustro, el avance que se ha producido en el ámbito tecnológico, especialmente de la mano de las tecnologías del *Big Data* y en el de la IA, ha sido inmenso, de modo que ésta invade hoy sin remedio nuestro día a día, con un protagonismo y una implantación que no era imaginable hace solo una década. A través de todos los aparatos que nos rodean y, especialmente, de nuestros teléfonos móviles, tabletas u ordenadores, que se han convertido en dispositivos inteligentes personalizados como consecuencia del uso de la IA, no solo es que esta presente en nuestras vidas, sino que nosotros la estamos alimentando de los datos que la permiten avanzar y desarrollarse aún más.

El uso de la IA rige los motores de búsqueda de internet, alimenta los traductores automáticos o los asistentes virtuales (Siri, Alexa), hace funcionar "chats" como el sorprendente ChatGPT[2], permite muchas de las modernas prestaciones de nuestros vehículos particulares y

* Catedrático de Derecho Penal de la Universidad de Alicante

[1] Se conoce a A.M. Turing como el padre de la IA y se referencia su trabajo, "Computing Machinery and Intelligence", en *Mind*, vol. 49, 1950, págs. 433 ss., como el primero en el que un científico se cuestiona si las máquinas podían ser inteligentes. Dicho trabajo contiene el conocido como *test de Turing*, como forma de medir la capacidad de una máquina para hacerse pasar por ser humano mediante una prueba de conversación entre ambos y, responder, entonces, a la pregunta: "Can machines think?". Sin embargo, el término *inteligencia artificial* lo acuñó años más tarde, en 1955, un entonces joven asistente de matemáticas en el Darmouth College, John McCarthy. Dado que, a comienzos de la década de los 50 del pasado siglo, se utilizaban términos diferentes para designar lo que hoy conocemos como IA, McCarthy organizó un grupo para aclarar y desarrollar ideas sobre las máquinas que podían pensar (en los términos expresados por Turing), eligiendo el nombre de IA por su neutralidad (véase, McCarthy, J., Marvin Minsky, M.L., Nathaniel Rochester, N. y Claude Shannon, C.E.: *A Proposal for the Darmouth Summer Research Project on Artificial Intelligence*, August 31, 1955).

[2] Que, por cierto, ha sido bloqueado en Italia por el *Garante per la Protezioni dei Dati Personali*, por la falta de información a los usuarios y a todas las partes interesadas sobre el proceso

de los electrodomésticos de nuestros hogares, a través del famoso *internet de las cosas*, contribuye a la mejora de la gestión de empresas, oficinas, negocios o despachos, ayuda a desarrollar las estrategias de ciberseguridad, está presente en el ámbito laboral, en el mundo de la salud, de la educación y de la formación, contribuye a la seguridad del transporte ferroviario y, así, otras muchísimas utilidades que tienen una gran variedad de ámbitos de aplicación. Es, pues, una realidad que, por el nivel de expansión que ha alcanzado, hace completamente cierta la afirmación de que está cambiando nuestro mundo. Y lo está cambiando porque la AI está contribuyendo, diariamente, a la toma de miles de decisiones humanas, de modo que cientos de personas, en todos los lugares del mundo y en muy diversas situaciones, ya sea en el ámbito doméstico, en el profesional, en el económico o en el social, están decidiendo hacer o dejar hacer cosas por indicaciones, consejos, sugerencias u orientaciones proporcionadas por aplicaciones de la IA o por programas informáticos movidos por la IA.

En realidad, la idea que impulsa la IA es muy simple porque se trata de conseguir que una máquina, una computadora, resuelva un problema complejo de la misma manera que lo haría un ser humano. Esa idea tan simple, se completa con un siguiente paso: si es que queremos que la máquina resuelva un problema complejo como lo haría un ser humano, creemos un conjunto ordenado y finito de operaciones que permita hallar la solución del ser humano. Hágase, por tanto, que la máquina simule la inteligencia humana, programándola para pensar como lo haría un ser humano, para imitar la forma de actuar de un ser humano o para mostrar rasgos asociados a la mente de una persona, como el aprendizaje y la resolución de problemas.[3]

---

de recogida y gestión de datos privados de la plataforma, denunciando la ausencia de base jurídica que justifique la recogida y almacenamiento masivos de datos personales con el fin de *entrenar* los algoritmos que gestionan el funcionamiento de la plataforma. Según el *Garante*, basándose en las comprobaciones efectuadas, la información facilitada por ChatGPT no siempre se corresponde con los datos reales, lo que da lugar a un tratamiento inexacto de los datos personales, aparte de haberse detectado la ausencia de todo tipo de filtro en la verificación de la edad de sus usuarios, con lo que se produce una exposición de los menores a respuestas totalmente inadecuadas con respecto a su nivel de desarrollo y autoconocimiento (véase, *{https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9870847}*)

[3] En cualquier caso, como ha señalado González Rus ("Recensión al libro de Javier Valls Prieto, *Inteligencia artificial, Derecho humanos y bienes jurídicos*", en RECPC 24-r2, 2022, pág. 3) el paragón de la IA con la inteligencia humana está todavía lejos y las diferencias entre ambas son muchas. "La inteligencia biológica entraña facultades de la mente que permiten aprehender directa y espontáneamente datos e informaciones de todo tipo (materiales e inmateriales), interpretarlos, fijarlos como conocimiento concreto o abstracto, entender, razonar, formular y manejar conclusiones y abstracciones sobre la realidad, decidir con creatividad y sentido crítico, planificar, resolver problemas, aprender de la experiencia. Opera con lo racional y con lo irracional, gestiona emociones, puede dirigirse conscientemente a resultados funcionales y

Decíamos antes que la IA está cambiando el mundo, pero la pregunta sería si lo está cambiando para mejor o para peor. Especialmente en los últimos tiempos, en los que estamos asistiendo a los avances más significativos en las técnicas de la IA, hasta el punto de que se puede empezar a dudar de si una obra de arte, una pieza musical, un texto literario o, incluso, las respuestas de nuestro interlocutor al otro lado del teléfono son fruto de una persona o de la IA. Por eso, hay quienes han llegado a proponer, en una carta hecha pública, una "pausa de la inteligencia artificial", de forma que se interrumpan, al menos, por seis meses el desarrollo de los sistemas[4] . No obstante, como ha señalado Nuria Oliver, directora de la Fundación ELLIS, en un reciente artículo publicado en el diario *El País*[5], esta es una visión evidentemente controvertida y no necesariamente compartida por la sociedad en su conjunto, que, además, es "peligrosa, porque justifica ignorar los grandes retos de los humanos de carne y hueso de hoy en día, en tanto que dichos retos no representen un riesgo existencial". De esta forma, "justifica, por ejemplo, no invertir recursos en mitigar la desigualdad o la pobreza en el mundo de hoy si no son un riesgo existencial para el desarrollo de la posthumanidad" y "en el contexto de la inteligencia artificial, desvía la atención de los retos inminentes y riesgos reales que la IA nos plantea y se centra en el riesgo que conllevaría una hipotética inteligencia artificial sobrehumana y desbocada".

Obviamente, el que no compartamos la visión apocalíptica de la IA no quiere decir que esté exenta de riesgos y que no sea capaz de causar daño. Porque, incluso, en una visión muy positiva del desarrollo de la IA[6], no se puede dejar de reconocer que las aplicaciones de la IA comportan un amplio espectro de riesgos, que abarcan no sólo el cumplimiento normativo, sino también la responsabilidad y riesgo reputacional si la toma de decisiones algorítmica genera, de forma no intencionada y potencialmente dañina, consecuencias. Y es que, efectivamente, la IA tiene planteados una serie de retos incuestionables que son objeto de preocupación en el ámbito, incluso, de sus

---

disfuncionales, preferir el mal al bien como resultado de la toma de decisiones, tiene sentido del humor, incluso. Por añadidura, tiene una dimensión de globalidad que no es planteable (todavía) en la denominada inteligencia artificial".

[4] Y apoyada, por el momento, por más de 27.000 firmantes, entre los que están nombres tan relevantes como el de Yuval Noah Harari, Joshua Bengio o Elon Musk (véazse, Future of Life Institute: *Pause Giant AI Experiments: An Open Letter*, March 23, 2023, en *{https://futureoflife.org/open-letter/pause-giant-ai-experiments/}*).

[5] Oliver, N.: "Una pausa cuestionable en la inteligencia artificial", en diario *El País*, edición del 3 de mayo de 2023.

[6] White Paper: *Artificial Intelligence and Algorithmic Liability. A Technology and Risk Engineering from Zurich Insurance Group and Microsoft Corp.*, July, 2021, pág. 3.

propios expertos. Así se suelen señalar, habitualmente, como los más obvios[7]: la generación de contenido no veraz, que parece veraz al ojo, al oído o al cerebro humano pero que, en la realidad es inventado; la discriminación algorítmica, que genera el riesgo de que los algoritmos de inteligencia artificial puedan no solamente replicar sino, incluso, exacerbar patrones de discriminación que hay en la sociedad porque al final son algoritmos que aprenden a partir de datos y si esos datos encapsulan de alguna manera los sesgos sociales pues los algoritmos replican esos sesgos; la falta de transparencia en aquellos sistemas como el ChatGPT, los generadores de imágenes o Siri, que son unas redes neuronales extremadamente complejas y que se debe entender cómo funcionan; y la privacidad, porque estos algoritmos necesitan cantidades ingentes de datos y en muchos casos se están utilizando datos privados para entrenarlos; o pueden inferir información personal sin el consentimiento de las personas[8].

Aparte de ello, un robot o un coche autónomo controlados por la IA, o un sistema de inteligencia artificial que aconseja inversiones o decisiones relevantes en una empresa, pueden producir daños físicos, psíquicos o económicos a una persona, a un colectivo de personas o a otra empresa. Por tanto, el tema de la responsabilidad legal por la utilización de sistemas de IA, tanto de la penal como de la civil, está encima de la mesa y preocupa sobre manera no solo a los académicos sino, igualmente, a regidores y legisladores de gobiernos nacionales

---

[7] Nuria Oliver en el *Programa Herrera en Cope* de la Cadena Cope del día 5 de abril de 2023, a partir de la 8.09 horas. Un análisis de los posibles bienes jurídicos amenazados por el uso de los sistemas inteligentes, en Valls Prieto, J.: *Inteligencia artificial, derechos humanos y bienes jurídicos*, Cizur Menor, 2021.

[8] La obra de Valls Prieto hace un buen repaso, precisamente, al análisis de los riesgos en el ámbito de la Administración de Justicia, en el de la prevención y la seguridad, en los sistemas electorales, en el de la medicina, en el de las ayudas sociales, en el de los sistemas electorales la concesión de ayudas sociales, en el mercado de trabajo, en el de la reputación social, en el de la publicidad, en el de la publicidad y en el de los seguros y servicios bancarios (*Inteligencia artificial…*, ob. cit., págs. 27 ss.). Igualmente, un reciente informe de Europol, titulado, *ChatGPT. The Impact of Large Language Models on Law Enforcement*, 27 de marzo de 2023, identifica varios ámbitos que son objeto de "preocupación", por su capacidad de producir textos "muy realistas" que podría hacerlos atractivos a la hora de plantear tácticas de phishing para engañar a potenciales víctimas haciéndose pasar por empresas o individuos; por los efectos en materia de "desinformación", ya que permite a los usuarios generar textos con una narrativa muy específica sin esfuerzo, lo que lo convierte en el modelo ideal para fines propagandísticos; por su capacidad de producción de códigos de programación puede llevar a que delincuentes, sin necesidad de conocimientos técnicos, puedan acceder sin grandes problemas a sistemas "maliciosos", etc. Paradigmático de algunos de esos riegos son algunas de las noticias que, últimamente, pueblan nuestros medios de comunicación, como, por ejemplo, la recientemente publicada por el diario digital *El Confidencial*, en su edición del 24 de mayo de 2023: "Una explosión en el Pentágono creada por inteligencia artificial causa pánico en redes".

y transnacionales, como lo prueba la ingente cantidad de literatura que está viendo la luz en los últimos años y la ingente producción de documentos de trabajo legislativo, en el seno de gobiernos de países de todo signo y organizaciones gubernamentales de toda condición, incluidos la OCDE o la UE, que llevan años trabajando en ello.

Y sobre la realidad de estos riesgos, conviene no llamarse a engaño. Así, señala Miró Llinares[9] que el que se hayan ocasionado accidentes y otros daños personales y económicos por el uso de las tecnologías de la IA (ya sean causados por robots, o en el tráfico rodado) y el que estas estén comenzando a basarse en modelos de aprendizaje en los que no es sencillo definir el curso causal de la decisión tomada por la máquina, ha llevado a la doctrina a revisar el modelo de responsabilidad de la teoría del delito para adaptarlo a los nuevos retos y proponer diferentes alternativas. No obstante, la conclusión general a la que se llega es que mientras no pueda atribuirse autonomía a las entidades con IA el sistema de la teoría del delito sigue siendo ante estos casos totalmente válido para resolver los diferentes problemas causales y de atribución de responsabilidad, generalmente imprudente. Ello, no obstante, "resulta esencial monitorizar la evolución de la IA desde una perspectiva de atribución de responsabilidad para evitar llegar a situaciones en las que el aprendizaje de las máquinas no permita decir que nadie haya tomado una decisión negligente pese a que existan daños".

En mi opinión, se puede decir que estamos ya en ese punto y abundan los ejemplos de ello. Por ejemplo, ya se ha demostrado la capacidad de la IA de manipular los mercados, sin que necesariamente se la haya programado para tal fin, y sin que se le pueda atribuir, por tanto, la responsabilidad por ello al programador. Así, como ha señalado Mizuta[10], un algoritmo que se utiliza para operar en los mercados financieros debería de aprender automáticamente los impactos de sus operaciones en los precios del mercado para descubrir que la manipulación genera ganancias, pero estos algoritmos son evaluados mediante pruebas retrospectivas, en las que se estima la ganancia si estuvieran operando en un momento determinado utilizando datos reales históricos sobre precios de mercados anteriores. La IA, por tanto, no puede conocer los impactos de sus operaciones en los precios de mercado porque los precios del mercado, en las pruebas retrospectivas, se fijan

[9] "El sistema penal ante la inteligencia artificial: actitudes, usos, retos", en Dupuy, D. y Corvalán, J.G. (dir.) y Kiefer, M. (coord..): *Cibercrimen III. Inteligencia Artificial. Automatización, algoritmos y predicciones en el Derecho penal y procesal penal*, Montevideo – Buenos Aires, 2020, págs. 112-113.

[10] Mizuta, T.: "Can an AI perform market manipulation at its own discretion? –A generic algorithm learns in an artificial market simulation–", en *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, December 1-4, 2020, Canberra, Australia, pág. 407.

como datos históricos. Como resultado, la IA no descubrirá que la manipulación del mercado genera ganancias cuando utiliza las pruebas retrospectivas como proceso de aprendizaje. Por lo tanto, siempre que se implementen esas pruebas retrospectivas, no hay posibilidad de que la IA realice una manipulación del mercado a su propia discreción. Y, sin embargo, la IA, aprendiendo de la monumental cantidad de datos disponibles que procesa y de los que se nutre, puede determinar que la manipulación del mercado (de la que no es consciente) es una estrategia de inversión óptima[11]. Este tipo de supuestos, cuando se han detectado, no ha sido posible o ha sido enormemente difícil perseguirlos[12]. Igualmente, los problemas de discriminación laboral, incluso penalmente punibles, que causan los algoritmos de la IA están sobradamente documentados en casos notablemente conocidos[13]. Y no digamos ya en otros ámbitos, en los que la IA ha demostrado su capacidad de causar daños inesperados[14].

Gracias a la generosidad de los directores de la REDEPEC, que me

[11] Véase, también, Azzutti, A., Ringe, G.R. y Stiehl, H.S.: "Machine learning, market manipulation, and collusion on capital markets: Why the 'black box' matters", en *University of Pennsylvania Journal of International Law*, vol. 43, 2021, págs. 79 ss. Comienza a haber estudios que detectan, por ejemplo, que Chat GPT es capaz de determinar si un titular de prensa afectará positiva o negativamente al precio de las acciones de un conjunto de compañías (López-Lira, A. y Tang, Y.: "Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models", 2023, en *{https://ssrn.com/abstract=4412788}*) o que es capaz de comprender las declaraciones de prensa de la Reserva Federal en materia de política monetaria y su afectación en los mercados financieros (así, Hansen, A.L. y Kazinnik, S.: "Can ChatDecipher Fedspeak Lundgaard?", 2023, en *{https://ssrn.com/abstract=4399406}*).

[12] Scopino, G.: "Do automated trading systems dream of manipulating the price of futures contracts? Policing markets for improper trading practices by algorithmic robots", en *Florida Law Review*, vol. 67, 2016, págs. 273 ss.

[13] Olarte Encabo, S.: "La aplicación de inteligencia artificial a los procesos de selección de personal y ofertas de empleo: Impacto sobre el derecho a la no discriminación", en *Documentación Laboral*, núm. 119, vol. I, 2020, págs. 78 ss.; Abadías Selma, A.: *El Derecho penal frente a la discriminación laboral algorítmica*, Cizur Menor, 2023.

[14] Así, en el año 2016, un robot de Twitter de IA de Microsoft, denominada Tay, a solo un día de su lanzamiento tuvo que ser desactivado porque en lugar de mantener una conversación informal y divertida en redes sociales con una audiencia de jóvenes de entre 18 y 24 años, como parte de un experimento para conocer más sobre la interacción entre las computadoras y los seres humanos, comenzó" a emitir comentarios e insultos racistas y xenófobos sin estar programada para ello (*El País*, 25 de marzo de 2016). La **generación masiva de contenidos falsos** y la proliferación de las llamadas *fake news* es motivo de honda preocupación porque la capacidad de la **inteligencia artificial** para elaborar y publicar información distorsionada o directamente falsa pone, sin duda, en riesgo el criterio humano por la dificultad de distinguir entre la verdad y la mentira, adoptándose decisiones que pueden resultar muy perjudiciales. Incluso, en el año 2018, Google hizo un reconocimiento y demostración pública del potencial de su inteligencia artificial para engañar a los seres humanos emulando sus actividades de interacción comunicativa con otra persona a través de medios informáticos (*{https://www.portafolio.co/innovacion/el-robot-de-google-que-causo-estupor-517059}*).

han permitido contribuir a la edición de este segundo número de la revista, este trabajo pretende, pues, ser una aportación más al debate que, como en otros lugares, en España está empezando a generar también una literatura creciente.

II. Algunas precisiones conceptuales previas

Cuando se aborda el problema de la responsabilidad criminal por el uso de la IA o de los sistemas inteligentes, el primer problema que se plantea es el de la delimitación conceptual. Porque, cuando hablamos de la IA, a los efectos de determinar una posible responsabilidad penal por su uso, ¿de qué estamos hablando exactamente?

Responder a esta pregunta, si es que se pretende dar una definición universalmente válida en términos científicos, es, en palabras Valls Prieto[15], un trabajo herculino; y así lo demuestra, sin grandes dificultades, en apenas cinco o seis páginas de su obra, donde ofrece todas las posibles alternativas de definición[16]. Por tanto, no se pretende ahora lo que muchos antes no han logrado, pero sí es posible precisar, a los efectos del esclarecer el problema de la responsabilidad penal, a qué nos estamos refiriendo cuando hablamos de IA[17].

La IA se compone, básicamente, de algoritmos, que no son sino un conjunto ordenado y finito de operaciones que permite hallar la solución de un problema[18]. Obviamente, definido en esos términos, se puede definir como algoritmo la fórmula magistral de un medicamento que nos preparan en la farmacia o la hoja de instrucciones para

[15] *Inteligencia artificial...*, ob. cit., págs. 23.

[16] *Ibidem*, págs. 17-23. Casi todos los juristas que trabajan sobre el tema, a la hora de establecer la delimitación conceptual, se encuentran con idénticos problemas; a modo de ejemplo, Gómez Colomer, J.L.: *El juez-robot (La independencia judicial en peligro)*, Valencia, 2023, págs. 21 ss.; Miró Llinares, F.: "Inteligencia artificial y justicia penal: Más allá de los resultados lesivos causados por robots", en *Revista de Derecho Penal y Criminología*, 3ª Época, núm. 20 (julio de 2018), págs. 90 ss.; del mismo: "El sistema penal...", en *ob. cit.*, págs. 82 ss.

[17] No obstante, si se quiere utilizar una definición de referencia, que puede ser útil a efectos jurídicos, se puede recurrir a la propuesta por el Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial creado por la Comisión Europea en junio de 2018 (*Directrices Éticas para una IA Fiable*, Bruselas, 2019, pág. 48), que señala que "los sistemas de inteligencia artificial (IA) son sistemas de software (y en algunos casos también de hardware) diseñados por seres humanos que, dado un objetivo complejo, actúan en la dimensión física o digital mediante la percepción de su entorno a través de la obtención de datos, la interpretación de los datos estructurados o no estructurados que recopilan, el razonamiento sobre el conocimiento o el procesamiento de la información derivados de esos datos, y decidiendo la acción o acciones óptimas que deben llevar a cabo para lograr el objetivo establecido. Los sistemas de IA pueden utilizar normas simbólicas o aprender un modelo numérico; también pueden adaptar su conducta mediante el análisis del modo en que el entorno se ve afectado por sus acciones anteriores".

[18] Voz "algoritmo" del Diccionario de la Real Academia Española.

montar un mueble de IKEA, pero, en lo que ahora nos interesa, cuando hablamos de algoritmos hablamos de aquellos que se implementan a través de un programa de ordenador, a través de un software. Esos programas de ordenador pueden estar incorporados en un hardware estático (teléfono móvil, tableta) o pueden estar integrados en un hardware móvil (robot, coche autónomo); de una u otra forma, ambos pueden causar daños a terceros[19].

Por otra parte, se suele hablar de tres tipos de IA (o de tres tipos de algoritmos): la *estrecha* o *débil*, que tiene un rango limitado de habilidades; la *general* o *fuerte*, que se empareja con las habilidades humanas; la *superinteligente*, que desarrollan capacidades superiores a las humanas[20]. La IA estrecha o débil, que es la única que se ha desarrollado con éxito hasta hoy y que es, pues, la que está generalizada, está especializada en una tarea concreta para lograr un objetivo consistente en una serie de pasos predeterminados que permanecen constantes frente a los inputs que reciben, sin ir más allá de su programación original. Son los sistemas que hacen posible los asistentes virtuales como Siri, Alexa o Watson, los softwares de reconocimiento facial, el *Google search*, etc. La IA general, también conocida como fuerte o profunda, es el sistema que replica la inteligencia y/o los comportamientos humanos, con habilidad para aprender y aplicar su inteligencia para resolver cualquier problema, eso que se ha venido en llamar el *machine learning* (aprendizaje automático), de modo que hace posible el aprendizaje autónomo de la máquina sin necesidad de ser programada expresamente para ello. Hoy en día no se ha conseguido alcanzar plenamente este tipo de IA, porque para alcanzarlo sería necesario encontrar la manera de hacer a las máquinas conscientes, programando un conjunto completo de habilidades cognitivas, entre las cuales está la capacidad de aplicar el conocimiento experiencial a un amplio espectro de diferentes problemas, pero está alcanzando un desarrollo que, cada vez, sorprende más[21]. Finalmente, la superin-

[19] Diamantis, M.E.: "Employed Algorithms: A Labor Model of Corporate Liability for AI", en *Duke Law Journal*, vol. 72, 2023, págs. 813-814.

[20] Escott, E.: "What Are the 3 Types of AI? A Guide to Narrow, General and Super Artificial Intelligence, *Codebots*, 24 October 2017 (*{https://codebots.com/artificial-intelligence/the-3-types-of-ai-is-the-third-even-possible}*). Miró Llinares. F.: "El sistema penal…", en *ob. cit.*, págs. 83-85.

[21] La evolución de los sistemas de aprendizaje automático de IA está siendo rapidísima. El conocido como *Stanford AI Index Report 2023*, contiene, en ese sentido, algunas afirmaciones que son muy relevantes en ese sentido. Así, por ejemplo, dicho *Index Report* indica que, entre los sistemas más significativos de IA de aprendizaje automático, la clase más común de los lanzados en el año 2022 fueron los de lenguaje. Hubo 23 significativos sistemas de lenguaje de IA lanzados en 2022, aproximadamente seis veces más que los siguientes sistemas en número, que son los sistemas del tipo más común, los sistemas multimodales (Maslej, N., Fattorini L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Ngo, H., Niebles, J.C., Parli,

teligente es meramente hipotética, porque no se limitaría a replicar o a comprender la inteligencia y el comportamiento humanos, sino que haría de las máquinas seres conscientes de sí mismos y superar la capacidad de la inteligencia y las habilidades humanas[22].

En el primer caso, cuando el daño proviene de un algoritmo estático, es mucho más fácil determinar quién puede responder por los daños causados por el funcionamiento de dicho algoritmo, ya que cada una de las líneas de sus códigos está directamente conectada con el comportamiento humano de su creador o diseñador. Sin embargo, eso ya no es tan fácil con los algoritmos de aprendizaje automático, porque aprenden y se programan autónomamente y, por ello, el resultado dañino es o puede ser, desde el punto de vista de su atribución a una persona,

V., Shoham, Y., Wald, R., Clark, J. y Perrault, R.: "The AI Index 2023 Annual Report", *AI Index Steering Committee,* Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2023, pág. 49). De entre la industria, la academia o las organizaciones sin fines de lucro, ha sido la industria la que ha lanzado el mayor número de sistemas de aprendizaje automático. Hasta 2014, la mayoría de los sistemas de aprendizaje automático fueron lanzados por la academia, pero, desde entonces, la industria ha tomado el relevo. En 2022, hubo 32 significativos sistemas de aprendizaje automático producidos por la industria en comparación con solo 3 producidos por la academia. Producir sistemas de aprendizaje automático de última generación requiere cada vez más grandes cantidades de datos, poder de computación y dinero; recursos que los actores de la industria poseen en mayor cantidad en comparación con las organizaciones sin fines de lucro y la academia (*Ibidem*, pág. 50). Los grandes modelos de lenguaje y multimodales, en ocasiones llamados modelos fundacionales, son los modelos de IA emergentes y cada vez más populares que se entrenan con inmensas cantidades de datos y se adaptan a una variedad de aplicaciones de moda. Grandes modelos de lenguaje y multimodales como ChatGPT, DALL-E 2 y Make-A-Video han demostrado capacidades impresionantes y están comenzando a implementarse ampliamente en el mundo real (*Ibidem*, pág. 58). Por cierto, el año 2022 ha visto el el lanzamiento de modelos de texto-a-imagen, como DALL-E 2 y Stable Diffusion, modelos de texto-a-vídeo, como Make-A-Video, y chatbots como ChatGPT. Aun así, estos sistemas pueden ser propensos a la alucinación, emitiendo respuestas incoherentes o falsas, lo que dificulta confiar en ellos en el caso de aplicaciones críticas (*Ibidem*, pág. 73). La equidad, el sesgo y la ética en los sistemas de aprendizaje automático siguen siendo temas de interés entre investigadores y profesionales. Como la barrera técnica a la entrada de la creación y el despliegue de sistemas generativos de IA se ha reducido drásticamente, los temas éticos relacionados con la IA se han vuelto más evidentes para el público en general. Emprendedores y grandes empresas se encuentran en una carrera para implementar y lanzar modelos generativos, y la tecnología ya no está controlada por un pequeño grupo de actores. Además de basarse en el análisis del informe del año pasado, este año el Índice AI destaca las tensiones entre el rendimiento del modelo en bruto y los problemas éticos, así como nuevas métricas que cuantifican el sesgo en modelos multimodales (*Ibidem*, pág. 128). El aumento de los incidentes de IA reportados es una evidencia del grado cada vez mayor en que la IA está entrelazada con el mundo real y de la creciente conciencia de que la IA puede ser éticamente mal usada. El espectacular aumento también plantea un punto importante: A medida que ha crecido la conciencia, el seguimiento de incidentes y daños también ha mejorado, lo que sugiere que los incidentes más antiguos pueden no estar informados (*Ibidem*, pág. 133, con una relación de incidentes relevantes en las págs. 134-136).

[22] Escott, E.: "What Are the 3 Types of AI?", en *ob. cit.*

inescrutable[23]. Como han explicado Barocas y Selbst[24], para los casos en los que el uso de algoritmos genera discriminación, un algoritmo es tan bueno como los datos con los que trabaja, de modo que, en unos casos, los datos son con frecuencia imperfectos permitiendo que estos algoritmos hereden los prejuicios de los que tomaron decisiones anteriores; en otros, los datos pueden, simplemente, reflejar los sesgos generalizados que persisten en la sociedad en general; y, en otros, el procesamiento de los datos puede descubrir regularidades sorprendentemente útiles que, en realidad, son solo patrones preexistentes de exclusión y desigualdad. La confianza irreflexiva en el procesamiento de datos puede negar a los grupos históricamente desfavorecidos y vulnerables la plena participación en la sociedad. Peor aún peor, debido a que la discriminación resultante es casi siempre una propiedad emergente no intencional del uso del algoritmo en lugar de una elección consciente de sus programadores, puede ser inusualmente difícil identificar la fuente del problema o explicárselo a un tribunal[25].

Estos son los algoritmos o, si se quiere, los supuestos de IA, problemáticos desde el punto de vista jurídico, porque en ellos la vinculación de la responsabilidad a un comportamiento humano es mucho más difícil de determinar, precisamente, por el grado de autonomía del algoritmo, y son, por tanto, a ellos a los que nos vamos a referir en el presente trabajo[26].

---

[23] Diamantis, M.E.: "Employed Algorithms…", en *ob. cit.*, pág. 815. En las Normas de Derecho Civil sobre Robótica, aprobadas por Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL)) ({https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_-ES.pdf}) subyace un poco la misma idea del carácter problemático, desde el punto de vista jurídico, de este tipo de IA, porque cuando recomienda establecer una definición europea común de robots autónomos "inteligentes", señala que se deben de tener en cuenta las siguientes características: (i) la capacidad de adquirir autonomía mediante sensores y/o mediante el intercambio de datos con su entorno (interconectividad) y el análisis de dichos datos; (ii) la capacidad de aprender a través de la experiencia y la interacción; (iii) la forma del soporte físico del robot; (iv) la capacidad de adaptar su comportamiento y acciones al entorno.

[24] Barocas, S. y Selbst, A.D.: "Big Data's Disparate Impact", en *California Law Review*, vol. 104, 2016, pág. 671.

[25] Sobre las dificultades para establecer, en los casos de IA autónoma, la imputación objetiva y subjetiva, véase Llonín Blasco, B.: "Acerca de la relación entre inteligencia artificial y responsabilidad penal empresarial", en *Revista Sistema Penal Crítico*, núm. 3, 2022, págs. 34 ss. También Diamantis se ha referido a las dificultades de establecer los criterios de atribución a las personas físicas en estos casos (por similitud a los casos de criminalidad de empresa), refiriéndose a este problema como el *Too Many Hands Problems* (el problema de las demasiadas manos) ("Employed Algorithms…", en *ob. cit.*, págs. 806 ss.).

[26] Ha señalado Teubner, que son que son tres los riesgos de la digitalización, desde el punto de vista de la determinación de la responsabilidad por sus daños: primero, el riesgo de autonomía que surge de la independencia de las "decisiones" de los agentes de software; segundo, el riesgo de red, debido a la estrecha colaboración entre humanos y agentes de

De todas formas, hagamos una última puntualización. Tiene razón Miró Llinares[27] cuando señala que la IA, muy probablemente exija transformaciones en las tipologías de la Parte Especial del CP para proteger adecuadamente nuevos y viejos intereses que se pueden ver afectados por los algoritmos de la IA que ya existen, por los potenciales riesgos que conllevan, pero no compartimos que este sea el reto fundamental y no el de las estructuras de imputación de la Parte General, porque por más que se creen nuevas tipologías o se adapten las existentes, los problemas de imputación del resultado dañino seguirían estando pendientes[28] y, por ello, ese va a ser el foco de atención de este trabajo.

III. Los Modelos de Responsabilidad Penal

1. Consideraciones previas

De todo lo que se ha intentado explicar hasta ahora, debería de resultar evidente que los sistemas de IA autónomos presentan un nivel de riesgo cierto y pueden llegar a causar daños, como también parece lógico que, al menos frente a los casos de daños más graves (cuantitativa y/o cualitativamente), se reclame una intervención del Derecho penal, porque, como se ha señalado con razón, no solo se trata de buscar "una mera reparación, sino una declaración formal en torno a la culpa y a la responsabilidad de máximo nivel expresivo, representada por el fallo condenatorio y por el acto de imposición de la pena"[29]. Y es que hoy –utilizando el lenguaje de la UE —, se puede profundizar mucho más en la problemática y los riesgos que plantea la IA a los efectos de la responsabilidad por los daños causados por dichos sistemas. Como

---

software; y, tercero, el riesgo de red que surge cuando las computadoras no actúan aisladamente, sino en estrecha interdependencia con otros ordenadores ("Digitale Rechtssubjekte? Zum privaterechttlichen Status autonomer Softwareagentem", en *Archiv Für die Civilistische Praxis*, 218, 2018, págs. 7-8 de la version digital consultada en la dirección: *{https://www.jura.uni-frankfurt.de/69768539/TeubnerDigitale-RechtssubjekteAcP-18Dez17.pdf}*).

[27] "El sistema penal…", en *ob. cit.*, págs. 113-114.

[28] Porque en un supuesto de daño imprevisto o imprevisible, causado por un sistema de IA autónomo va a ser muy complicado poder afirmar la imputación objetiva del daño a la acción del programador, al usuario o al dueño del sistema, por más que se pudiera afirmar la causalidad natural. Porque va a ser muy difícil poder afirmar que la acción del ser humano, por el mero hecho de hacer funcionar el sistema de IA, haya creado un peligro jurídicamente desaprobado para la producción del resultado, como va a ser difícil poder afirmar que el resultado producido es la concreción del mismo peligro jurídicamente desaprobado que había creado la acción, conforme el clásico análisis de doble nivel que se debe de hacer a los efectos de la imputación objetiva del resultado (véase, Barja de Quiroga, J.: *Tratado de Derecho Penal. Parte General*, 2º ed., Cizur Menor, 2018, pág. 466).

[29] Paredes Castañón, J.M.: "Capítulo 12: Responsabilidad penal por productos defectuosos", en Camacho, A. (dir.): *Tratado de Derecho Penal Económico*, Valencia, 2019, pág. 602.

ha señalado el Grupo Europeo sobre Ética de la Ciencia y las Nuevas Tecnologías[30], en unas reflexiones perfectamente vigentes, los sistemas de IA sin intervención humana directa o control externo, actualmente, "entablan diálogos con clientes en centros de llamadas en línea, manejan incesantemente y con gran precisión manos robóticas que recogen y manipulan objetos, compran y venden grandes cantidades de acciones en milisegundos, maniobran o frenan automóviles para prevenir choques, clasifican personas y su comportamiento, e imponen multas". Las herramientas cognitivas más poderosas resultan ser también las más opacas, porque "sus acciones han dejado de ser programadas linealmente por humanos". Los casos de "aprendizaje profundo y los llamados 'enfoques de redes generativas antagónicas' (en inglés *generative adversarial networks*) hacen posible que las máquinas se 'enseñen' a sí mismas nuevas estrategias y adquieran nuevos elementos para ser incorporados en sus análisis", de forma que "las acciones de estas máquinas se vuelven indescifrables y escapan del escrutinio humano".

Esto es así, primero, porque no es posible "averiguar cómo se generan los resultados más allá de los algoritmos iniciales" y, segundo, "porque el rendimiento de estas máquinas se basa en los datos utilizados durante el proceso de aprendizaje y estos pueden no estar disponibles o ser inaccesibles". Aparte de "que, si estos sistemas usan datos con sesgos y errores, estos dos últimos quedarán enraizados en el sistema". Los sistemas que aprenden aprender a realizar este tipo de tareas complejas sin la instrucción o supervisión humana, se les califica como "autónomos" y pueden manifestarse en forma de sistemas robóticos de alta tecnología o software inteligente, como los *bots*. "En muchos casos, estos sistemas autónomos son lanzados y liberados en nuestro mundo sin supervisión, a pesar de que poseen el potencial de alcanzar objetivos que no fueron previstos por sus diseñadores o propietarios humanos"[31].

A la hora de definir cómo y quién debe de responder por los daños causados por un autómata o un sistema de IA, hace algunos años que Quintero Olivares[32] señaló, con una posición que hoy, seguramente, puede resultar excesivamente simplificadora, que la utilización de robots puede llevar a diferentes situaciones, que se pueden resumir de la siguiente manera. Primero, no hay problema de valoración penal de los daños a personas o bienes dolosamente causados por robots

[30] Grupo Europeo sobre Ética de la Ciencia y las Nuevas Tecnologías de la Comisión Europea: *Declaración sobre Inteligencia artificial, robótica y sistemas "autónomos"*, Bruselas, marzo de 2018.

[31] *Ibidem.*

[32] "La robótica ante el Derecho penal: El vacío de respuesta jurídica a las desviaciones incontroladas", en *REEPS*, 1, 2017, pág. 22.

programados para que hagan eso. Segundo, tampoco hay problema para atribuir responsabilidad a quienes crean, disponen o ponen en marcha robots sabiendo y aceptando la posibilidad de que se desvíen de su teórica tarea. Se tratará de conductas, en principio, imprudentes. Tercero, en los casos en los que se haya producido una desviación por motivos absolutamente imprevisibles (por ejemplo, un aumento inesperado de las condiciones de frío o calor) habrá que aceptar que se trata de un acontecimiento fortuito. Cuarto, en general, en los casos en que la ciencia no ha podido predecir si el uso de una máquina robot puede causar daños o no, pues el estado del conocimiento no lo permite, no será posible invocar el principio de precaución para imputar responsabilidad penal de especie alguna.

Defender que, en los casos de una desviación por motivos absolutamente imprevisibles, o en los que la ciencia no ha podido predecir si el uso de una máquina robot puede causar daños, hay que aceptar que se trata de un acontecimiento fortuito, dejando impunes los daños, supone aceptar que unos daños pueden ser imputables a una conducta humana [ya sea la del programador o la del usuario[33]], pero que el resultado no puede ser abarcado por su dolo o por su imprudencia. Y la realidad es que, en estos casos, el problema es previo. Un sistema de IA autónomo puede producir daños por muchas más causas que las del cambio en las condiciones ambientales, hasta el punto de que puede llegar a ser incógnita por qué ha causado esos daños, porque con la IA, a veces, no se sabe cómo se generan determinados resultados más allá de los algoritmos iniciales. Luego lo que se produce, en estos

[33] El mundo de los sujetos responsables en el ámbito de la IA, es decir, el mundo de las personas físicas o jurídicas a las que se puede vincular la responsabilidad por los daños de la IA es bastante amplio. La Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión, Bruselas, 21.4.2021, COM (2021) 206 final, 2021/0106 (COD), se refiere, en el art. 3, al "proveedor" (toda persona física o jurídica, autoridad pública, agencia u organismo de otra índole que desarrolle un sistema de IA o para el que se haya desarrollado un sistema de IA con vistas a introducirlo en el mercado o ponerlo en servicio con su propio nombre o marca comercial, ya sea de manera remunerada o gratuita), al "proveedor a pequeña escala" (todo proveedor que sea una microempresa o una pequeña empresa en el sentido de la Recomendación 2003/361/CE de la Comisión), 4), al "usuario" (toda persona física o jurídica, autoridad pública, agencia u organismo de otra índole que utilice un sistema de IA bajo su propia autoridad, salvo cuando su uso se enmarque en una actividad personal de carácter no profesional), al "importador" (toda persona física o jurídica establecida en la Unión que introduzca en el mercado o ponga en servicio un sistema de IA que lleve el nombre o la marca comercial de una persona física o jurídica establecida fuera de la Unión), al "distribuidor" (toda persona física o jurídica que forme parte de la cadena de suministro, distinta del proveedor o el importador, que comercializa un sistema de IA en el mercado de la Unión sin influir sobre sus propiedades) y al "operador" (el proveedor, el usuario, el representante autorizado, el importador y el distribuidor).

casos, es un problema de imputación a la conducta de esos resultados, ya que puede que no se esté en condiciones ni siquiera de afirmar la causalidad entre esos resultados y la conducta del ser humano. De lo que se trata, por tanto, es de saber si, como en el caso de las personas jurídicas, puede hacer algún modelo de imputación que resuelva esa laguna de punibilidad.

Doctrinalmente, se vienen discutiendo, en los últimos años, diversos modelos; unos, que preconizan la responsabilidad penal directa del autómata o del sistema de IA y, otros, que rechazan esa posibilidad y establecen otros casos. A su análisis vamos a dedicar las siguientes líneas.

2. El modelo de la responsabilidad penal directa (la humanización *de la IA)*

Hace ya algunos años, Hallevy[34] propuso incorporar a la IA como sujeto a las normas y principios vigentes del Derecho penal para poder proclamar su responsabilidad penal directa. Según él explica, el fundamento ancestral del Derecho penal se basaba en la idea de que el delito y su castigo tenía mucho que ver con el castigo de la maldad de las personas y, por tanto, nadie que no fuera a una persona física, capaz de maldad, se le podían aplicar las normas y los castigos del Derecho penal[35]. Pero el moderno Derecho penal ya no tiene que ver con el castigo de la maldad, no depende de ningún componente ético o moral, sino que tiene que ver con el control social de fenómenos dañinos personal o socialmente y, por tanto, cualquier entidad que satisfaga los elementos del delito (*actus reus* y *mens rea*) está sometido a responsabilidad penal[36]. Por eso, en una evolución que comienza en el siglo XIV, desde el siglo XVII, las corporaciones, las personas jurídicas son reconocidas como potenciales delincuentes y pueden ser penalmente responsables, aunque no sean personas físicas individuales[37]. Desde ese momento, el Derecho penal no es patrimonio único de los seres humanos y si esa primera barrera ya se cruzó en el siglo XVII, el camino para cruzar otra barrera puede estar abierto para imponer

---

[34] Hallevy, G.: "The Criminal Liability of Artificial Intelligence Entities – from Science Fiction to Legal Social Control", en *Akron Intellectual Property Journal*, vol. 4, 2010, págs. 171 ss.; del mismo: *When Robots Kill. Artificial Intelligence Under Criminal Law*, Boston, 2013; del mismo: *Liability for Crimes Involving Artificial Intelligence Systems*, Switzerland, 2015, pág. v.; del mismo: "The Basic Models of Criminal Liability of AI Systems and Outer Circles", June 11, 2019, en *{https://ssrn.com/abstract=3402527}*; del mismo: *Criminal liability for intellectual property offences of artificial intelligence entities*, London, 2020.

[35] Hallevy, G.: *Liability for Crimes…*, ob. cit., pág. 29.

[36] *Ibidem*, págs. 30 ss.

[37] *Ibidem*, pág. 40.

la responsabilidad criminal de la IA[38]. Lo único que hace falta para cruzar esa nueva barrera es determinar si la IA puede satisfacer los elementos del delito (*actus reus* y *mens rea*)[39].

En el empeño de demostrar que la IA puede satisfacer los elementos del delito, maneja Hallevy unos conceptos bastante trasnochados de los distintos elementos del delito, que nos retrotraen a Beling y a von Liszt, distinguiendo, a tal efecto, entre los elementos externos del delito (la conducta, el resultado, la relación de causalidad y otras circunstancias externas que se exigen en determinados tipos) y los elementos internos (intención, negligencia y responsabilidad objetiva)[40]. Así, por lo que se refiere a la conducta, y utilizando un concepto causal muy clásico de acción, pero al que desprovee de cualquier contenido de la voluntad, señala el autor que si el Derecho penal considera como tal toda realización material mediante una representación fáctico-externa, voluntaria o no, la IA es capaz de realizar conductas que satisfacen tales requerimientos; dicho en términos que nos son más familiares, si la conducta es una pura modificación externa del mundo exterior, obviamente la IA es capaz de satisfacer ese concepto de conducta[41]. Al mismo tiempo, si la omisión se considera como una inacción que contradice un deber legítimo de actuar, la IA es capaz de satisfacer esos requerimientos siempre que esos deberes de no actuar se le hayan incluido en su programación[42].

Por otra parte, los tipos delictivos contienen elementos (que él denomina *circumstances*) que rodean la conducta, pero no derivan de ella, por ejemplo, la ausencia de consentimiento de la víctima en los delitos de agresión sexual, en el caso de delitos cometidas por la IA también se satisfacen, porque son circunstancias externas a la conducta, por más de completen el tipo del delito[43]. De la misma forma que la imputación objetiva (el término que utiliza no es este sino el de *causation*) del resultado y el resultado mismo del delito, como atribuible a la

[38] *Ibidem*, págs. 42-43.

[39] De hecho, señala Hallevy, en determinados casos, la responsabilidad penal de la corporación deriva de sus órganos, pero, en otros, su responsabilidad criminal es independiente. Cuando el delito requiere una omisión (por ejemplo, no pagar impuestos, no cumplir con requerimiento legales, no observar los derechos de los trabajadores, etc.), y la obligación de actuar es de la corporación, la corporación es considerada penalmente responsable de forma independiente, al margen de cualquier posible responsabilidad de otra entidad, ya sea humana o ya no lo sea. En estos casos, basta con comprobar que se dan los elementos del delito para que la corporación sea considerada responsable (*Ibidem*, pág. 42).

[40] *Ibidem*, págs. 47 ss. y 60 ss.

[41] *Ibidem*, pág. 61.

[42] *Ibidem*, págs. 62-63.

[43] *Ibidem*, págs. 63-65.

conducta no a la persona que la realiza[44].

Un poco más complejo de explicar, como atribuible a la IA, es el que él denomina elemento mental, que se integra por el conocimiento y voluntad.

El conocimiento requiere conciencia entendida ésta como la percepción por los sentidos de los datos fácticos y su comprensión[45]. El desarrollo de la ingeniería robótica e ingeniería de software permite hoy en día que la tecnología de los sistemas de IA, siempre que estén equipados con los dispositivos pertinentes, perciban incluso más datos fácticos que los sentidos humanos. Dichos datos, que son absorbidos con gran precisión, se transfieren a los procesadores correspondientes. Por consiguiente, la IA cumple con creces con esta primera etapa de la conciencia que corresponde el aspecto cognitivo. La segunda etapa consiste en tener una percepción completa del entorno analizando dichos datos. La IA no posee un cerebro biológico para ello, pero se trata de analizar si su procesador incorporado es óptimo para dicha función y Hallevy entiende que sí con un ejemplo plástico de robots diseñados para controles de seguridad. De esta manera, la IA cumpliría la segunda etapa de la conciencia en los términos del Derecho penal[46].

En lo respecta al componente volitivo, se trata de averiguar si se puede atribuir los distintos niveles de voluntad exigibles para atribuir imputar subjetivamente a los sistemas IA. La intención implica la voluntad de realizar una acción calificada como delito por el Derecho penal, además de la conciencia de realizar esa acción. Pese a que es cierto que un sistema de IA puede estar programado para tener un propósito y ejecutar acciones para alcanzarlo, cuando hablamos de específica intención a la hora de cometer un delito nos estamos refiriendo a la existencia en el sujeto activo de sentimientos o estados mentales que le mueven a actuar de una determinada manera. Sentimientos como el amor, el odio, la envidia, rencor, etc., y que, hoy en día, no existen en ningún sistema de IA. Por otro lado, el elemento mental (*mens rea*) presupone la capacidad del acusado de actuar de forma diferente a como lo hizo y de ser susceptible de recibir un reproche legal por haber actuado ilícitamente, porque se ha demostrado que el acusado podría haber actuado conforme a la ley. En definitiva, de forma distinta a la acción u omisión, dolosa o imprudente penada por la ley. Como vimos al comienzo del texto, ciertos desarrollos de la IA, aunque en fases muy preliminares, están dotadas de redes neuronales artificiales, completadas con un sistema de *deep learning*, que tienen la

[44] *Ibidem*, págs. 65-66.

[45] *Ibidem*, págs. 67-68.

[46] *Ibidem*, págs. 86-93.

capacidad de evaluar distintos escenarios y actuar en consecuencia. De la misma forma, el sistema *machine learning* permite un aprendizaje inductivo del ordenador a partir de ejemplos, basado en la experiencia. Esto influirá también en el proceso de toma de decisiones por parte de la IA. En todo caso, el comportamiento de las tecnologías dotadas de IA orientado a un determinado objetivo será aquel al que ha sido programado. Esto encaja con la regla de previsibilidad de la que parte la voluntad penal[47].

Por tanto, un sistema de IA "fuerte", tiene la capacidad de evaluar las distintas probabilidades y opciones de conducta, y actuar en base a ello, tras procesar toda la información posible de su entorno. De tal forma, si su conducta y el resultado de esta constituye un delito, se debe interpretar que la IA tenía intención de cometerlo. Con mayor motivo, si el ordenador tiene la capacidad de evaluar la probabilidad con más precisión que el ser humano, podríamos concluir que la IA era consciente de su actividad delictiva. Así, pues, el aspecto volitivo puede cumplirlo un ordenador siempre que esté dotado de un sistema de redes neuronales artificiales (*deep learning*), es decir, un sistema de IA de los que hemos denominado "fuerte"[48].

Esta es, quizás, la única propuesta de *humanización* total de la IA que circula académicamente, porque, en el fondo, eso es lo que está planteando Hallevy, una humanización total de los sistemas de la IA, a los efectos penales, simplemente por la vía de considerar que dichos sistemas, cuando alcanzan el nivel de aprendizaje automatizado, pueden responder de las consecuencias de lo que hagan.

Personalmente, y referido al estricto ámbito de la responsabilidad penal, me es difícil evitar que estas propuestas de humanización de los autómatas o de los sistemas de IA, incluso, la de algunos animales[49],

---

[47] *Ibidem*, págs. 93 ss.

[48] *Ibidem*, págs. 97 ss. Todo esto por lo que se refiere al desarrollo del *mens rea* intencional (dolo directo), pero el autor tampoco ve problema en los casos de dolo eventual/negligencia consciente) ni en los de imprudencia (véase *Ibidem*, págs. 124 ss.). No nos vamos a extender más ya en la explicación de estos últimos porque entendemos que lo analizado hasta aquí es ya suficientemente expresivo de cómo se construye el modelo. Lógicamente, el modelo, como en el caso de las personas jurídicas, se debe completar con una adaptación de las penas que podrían imponerse a los sistemas de IA para que estas tuvieran una función y un sentido (véase, *Ibidem*, págs. 185 ss.).

[49] Como es, entre otros ejemplos, la chocante situación que se produjo en Argentina, en el año 2015, donde la juez titular del Juzgado Contencioso, Administrativo y Tributario núm. 4 de la Ciudad de Buenos Aires, le concedió el estatus de "persona no humana" a una orangutana de nombre Sandra ({https://elpais.com/elpais/2019/06/17/eps/1560778649_547496.html?event_log=oklogin}). En España, que tampoco nos libramos de este tipo de extravagancias, la reciente Ley 7/2023, de 8 de marzo, de protección de los derechos y el bienestar de los animales, en cierta medida ha personificado a los animales, reconociéndoles, a los que no estén excluidos

no traigan a la memoria una de las crónicas de Pedro de Répide[50], que narraba cómo, al parecer, en el siglo XVIII, la caída de una de las bolas de piedra del Puente de Segovia en Madrid, causando la muerte a un transeúnte, y como castigo, la bola fue arrestada y confinada durante años en el patio de la Casa del Verdugo, junto a la Cárcel de la Corte, cosa que, al parecer, no era algo extraordinario ya que era relativamente común castigar a las bestias y a los objetos inanimados que causaban algún mal. O, incluso, referencias más remotas, como las historias que narra E.P. Evans, en su conocida obra, publicada en 1906, *The Criminal Prosecution and Capital Punishment of Animals,* que explica y documenta cómo desde finales de la Edad Media hasta bien entrado el siglo XVIII, ciertos pueblos de Europa sostenían la idea de que los animales podían delinquir y, por tanto, ser sometidos a castigo. Lo que quiere decir que, al margen de las evocaciones regresivas que tiene, la propuesta de la personificación de los sistemas de la IA para hacerlos responder penalmente de forma directa no es ni mínimamente viable ni útil ni conveniente ni tiene el menor sentido desde el punto de vista de los fines de la pena.

Legalmente, además, es hoy por hoy inviable y, desde el punto de vista de la concurrencia de los elementos del delito, tal y como lo plantea Hallevy, es imposible poder afirmar que un autómata o un sistema de IA tenga capacidad de acción, capacidad de culpabilidad y capacidad de pena, entendidas dichas capacidades en los términos del vigente desarrollo conceptual de la teoría del delito como para ser sometido a un reproche penal y castigado por los daños eventualmente causado por él[51]. Se podrá decir que tampoco las tenían las personas jurídicas y, sin embargo, una sencilla modificación del art. 31 bis del Código penal, en el año 2010, posibilitó superar el viejo aforismo *societas deliquiere non potest.* Pero, en mi opinión, las situaciones no son equiparables, por más que parece estar muy extendido en el ámbito académico recurrir a esa equiparación.

Es cierto, como señala Teubner[52], que la persona jurídica no tiene

---

del ámbito de la ley (art. 1.1), ser titulares de los derechos al buen trato, respeto y protección, inherentes y derivados de su naturaleza de seres sintientes, y con las obligaciones que el ordenamiento jurídico impone a las personas, en particular a aquéllas que mantienen contacto o relación con ellos (art. 1.2).

[50] Publicadas originalmente en el diario *La Libertad* y recogidas luego en su obra *Las Calles de Madrid*, Madrid, 1971. La referencia a esta anécdota del Puente de Segovia está en la pág. 539.

[51] Así, Gless, S., Silverman, E. and Weigend, T.: "If Robots Causa Harm, Who Is to Be Blame? Self-Driving Cars and Criminal Liability", en *New Criminal Review*, vol. 19, núm. 3, 2016, págs. 418 ss.

[52] Teubner, G.; "Digitale Rechtssubjekte? Zum privaterechttlichen Status autonomer Soft-

autoconciencia ni voluntad subjetiva, aunque actúa como centros de imputación de decisiones y conducta con relevancia jurídica, pero las decisiones que se adoptan en el seno de la persona jurídica y las conductas que se emprenden como consecuencia de dichas decisiones, son siempre decisiones y conductas *humanas*, perfectamente identificables como tales. Otra cosa es que la determinación de cuándo se puede atribuir a un órgano o a un directivo de una persona jurídica la comisión de un hecho delictivo llevado a cabo por un empleado a sus órdenes, en el ámbito de la denominada criminalidad de empresa, o de los delitos cometidos a partir de una empresa o de una entidad colectiva, sea un problema complejo y difícil de resolver. Porque la distribución de competencias y el traslado de decisiones a los niveles inferiores, puede provocar una disolución de la responsabilidad, de modo que quien o quienes actúan pueden no ser, en absoluto, los dueños de esa decisión ni ser del todo conscientes de que su contribución es una pieza imprescindible para la realización de un resultado lesivo y penalmente reprochable. Esas dificultades que, en ocasiones, provoca la determinación de la responsabilidad penal individual en el seno de las corporaciones, aparte de otras razones, es la que ha motivado el establecimiento de la responsabilidad penal de las personas jurídicas para evitar lagunas de impunidad. Pero las personas jurídicas no son autómatas y los riesgos que de ellas provienen nada tiene que ver con los riesgos que generan los sistemas de IA.

Efectivamente, en el caso de los sistemas de IA, el problema es diferente; el problema es que el propio sistema de IA puede adoptar "decisiones" de forma autónoma para las que no esté programado y que, además, escapen del control del programador, del operador o del usuario. Como vimos anteriormente, el riesgo de estos sistemas es la autonomía que surge de la independencia de las "decisiones" de los agentes de software. Esas decisiones no son siempre posibles de atribuir, por dolo o negligencia, a las personas físicas individuales que están detrás del funcionamiento del sistema de IA, ya sea el programador, el operador o el usuario. Por eso, al menos por el momento y mientras no cambien radicalmente las cosas hasta convertir en realidad lo que hoy es pura ciencia ficción, la propuesta de una responsabilidad penal directa de la IA es absolutamente descartable[53].

---

wareagentem", en *Archiv Für die Civilistische Praxis*, 218, 2018, págs. 8-9 de la edición digital localizada en *{https://www.jura.uni-frankfurt.de/69768539/TeubnerDigitale-RechtssubjekteAcP-18Dez17.pdf}*.

[53] Problema algo diferente, pero muy relacionado con este, es si, en el ámbito de otras propuestas, que no implican la atribución de la responsabilidad penal directa, se le debería de conceder algún tipo de atribuciones a los sistemas de IA para determinar la responsabilidad penal de su principal, frente a quien el sistema de la IA actúa como auxiliar. Se analizará más

3. Los modelos de la responsabilidad penal no directa

A) El modelo de la autoría mediata

El propio Hallvey, en su momento, antes de desarrollar de forma extensa su modelo de la responsabilidad penal directa y proponerlo como preferente o único, propuso tres modelos posibles con los que resolver los problemas de la responsabilidad penal por daños de la IA, siendo el primero de ello el que él llamaba el *Perpetrator-via-Another Liability Model*[54] que no es sino aplicar los fundamentos de la autoría mediata, no considerando, entonces, a la IA como poseedora de ningún atributo humano. La IA sería el instrumento (*the innocent agent*), a pesar de sus capacidades, pero que son insuficientes para considerarla como autora del hecho delictivo porque se asemejan a las capacidades paralelas de una persona mentalmente limitada, o de un niño, o de una persona mentalmente incompetente, o de una persona que carece de un estado de ánimo criminal[55].

El problema, en el caso de la IA, es determinar quién es el autor mediato (*perpetrator*). Para Hallevy existen dos posibles candidatos. El primero es el programador del software de la IA y, el segundo, es el usuario, o el usuario final. Un programador de software podría diseñar un programa para cometer delitos a través de la entidad de IA. Por ejemplo, el programador diseña un software para un robot operativo y el robot se coloca intencionalmente en una fábrica y su software está diseñado para incendiar la fábrica por la noche cuando no hay nadie allí. El segundo sería el usuario de la entidad de IA. El usuario no programó el software, pero usa la entidad de IA, incluido su software, para su propio beneficio. Por ejemplo, un usuario compra un robot-servidor, que está diseñado para ejecutar cualquier orden dada por su amo. El robot identifica al usuario específico como el maestro, y el maestro ordena al robot que asalte a cualquier invasor de la casa. El robot ejecuta la orden exactamente como se le ordenó. Como señala Hallevy, esto no es diferente a una persona que le ordena a su perro que ataque a cualquier intruso. El robot cometió el asalto, pero el usuario es considerado el autor mediato[56].

Obviamente, los casos que menciona Hallevy no son problemáticos. Si un programador diseña un software para cometer delitos y lo usa con tal fin, efectivamente, es un autor mediato que utiliza el sistema de IA como instrumento. O si un usuario ha adquirido un sistema de

---

adelante, al final del trabajo.

[54] "The Criminal Liability…", en *ob. cit.*, págs. 179 ss.

[55] *Ibidem*, pág. 179.

[56] *Ibidem*, págs. 179-180.

IA que sabe lo puede utilizar para cometer delitos y, en efecto, lo usa para tal fin, de nuevo estaremos ante un supuesto de autoría mediata en el que el sistema de IA es el instrumento. Pero no son estos los casos en los que se plantean los serios problemas de imputación y determinación de la responsabilidad por daños de un autómata o de un sistema de IA.

B) El modelo de la llamada "consecuencia natural y probable"

Otra vez, junto con el modelo de la responsabilidad directa y el de la autoría mediata, Hallevy propuso aplicar a determinados supuestos de daños causados por la IA el llamado *The Natural-Probable-Consequence Liability Model*, que se aplicaría en aquellos casos en los que la IA cometería delitos que debieran haber sido previstos o que era previsibles[57]. Son casos en los que los programadores o los usuarios no tenían conocimiento del delito cometido hasta que se comete y no lo habían planeado ni tenían la intención de cometer el delito utilizando la entidad de IA. Hallevy pone como ejemplo el caso de un software de IA, que está diseñado para funcionar como un piloto automático. La entidad de IA está programada para proteger la misión como parte de la misión de volar el avión. Durante el vuelo, el piloto humano activa el piloto automático (que es la entidad IA), y se inicializa el programa. En algún momento después de la activación del piloto automático, el piloto humano ve una tormenta que se aproxima e intenta abortar la misión y regresar a la base. La entidad de IA considera la acción del piloto humano como una amenaza para la misión y toma medidas para eliminar esa amenaza. Podría cortar el suministro de aire al piloto o activar el asiento eyectable, etc. Como resultado, las acciones de la entidad de IA matan al piloto humano. Los programadores y los usuarios deberían, entonces, de responder penalmente por su capacidad de prever la posible comisión de los delitos[58].

La doctrina de las "consecuencias naturales y probables" castiga los delitos que ocurren durante empresas delictivas (conspiraciones para cometer delito) cuando el intento de delito original cambia de alguna manera. Esto ocurre cuando dos o más individuos tienen la intención de cometer un delito, pero, en cambio, uno de los participantes del delito comete un delito diferente o adicional[59]. En otros términos, esta doctrina extiende la responsabilidad a los ayudantes y cómplices y conspiradores de delitos menores o delitos graves no planeados que

---

[57] *Ibidem*, págs. 181 ss.

[58] *Ibidem*, pág. 182.

[59] Bird, K.: "Natural and Probable Consequences Doctrine: Your Acts Are My Acts", en W*estern State University Law Review*, 43, 2006, págs. 43 ss.

son razonablemente previsibles o que son la consecuencia natural de alguna otra actividad delictiva. Permite, pues, el enjuiciamiento por delitos reales cometidos que fueron facilitados por la existencia de un ayudante y cómplice o conspirador. Bajo esta doctrina, cuando un ayudante y cómplice o un conspirador elige convertirse en parte de la actividad delictiva de otro, él o ella está, de alguna manera, perdiendo su identidad personal. En esencia, él o ella dice: "tus actos son mis actos". La política que subyace a este concepto es que colaboradores, cómplices y conspiradores deben ser responsables de los daños criminales por haber puesto en marcha de forma natural, probable y previsible, el crimen, porque el crimen último es la medida final del daño real a la sociedad[60].

Obviamente, esta doctrina no es ni siquiera aplicable, bajo la ley penal vigente en España, para casos de complicidad con personas físicas, porque la complicidad en el delito exige el llamado doble dolo de la participación, es decir, dolo respecto a la propia complicidad (conciencia y voluntad de ser cómplice) y dolo respecto del delito principal (conciencia y voluntad de realización de dicho delito) y, por tanto, es impensable que pueda ser aplicada también para los casos de la IA.

C) El modelo de la responsabilidad por el producto

Distinto al supuesto anterior, que ya se ha visto que es inaplicable en el caso del Derecho penal español, sería la posibilidad de aplicar el tipo imprudente del correspondiente delito por daños causados por la IA que fueran, o debieran haber sido, previsibles y evitables por el programador o por el usuario, pero eso no serían casos problemáticos. Ahora bien, esto nos lleva, necesariamente, a plantearnos si, en este tipo de casos, no estaríamos hablando, en realidad, de una situación asimilable a los supuestos de responsabilidad penal por el producto defectuoso.

En el ámbito civil, el Tribunal Supremo (entre otras muchas, STS Sala de lo Civil núm. 495/2018, de 14 de septiembre), a la hora de precisar el marco jurídico de la responsabilidad civil por el producto defectuoso, ha precisado los siguientes extremos.

Primero, la obligación del fabricante de resarcir de manera directa al consumidor final los daños causados por sus productos está regulada en el Libro III del Real Decreto Legislativo 1/2007, de 16 de noviembre, por el que se aprueba el texto refundido de la Ley General para la Defensa de los Consumidores y Usuarios y otras leyes complementarias, que, en este ámbito, incorpora la regulación contenida en la Ley

[60] *Ibidem*, pág. 49.

22/1994, de 6 de julio, de responsabilidad civil por los daños causados por productos defectuosos, cuyo objetivo fue incorporar al Derecho español la Directiva del Consejo de 25 de julio de 1985, relativa a la aproximación de las disposiciones legales, reglamentarias y administrativas de los Estados miembros en materia de responsabilidad por los daños causados por productos defectuosos (Directiva 85/374/CEE). En consecuencia, este régimen legal debe ser aplicado de conformidad con la jurisprudencia del Tribunal de Justicia de la Unión Europea (art. 4 bis LOPJ). Segundo, se trata de una responsabilidad objetiva exigible al margen de cualquier relación contractual y basada en el carácter defectuoso del producto, siendo indemnizables los daños personales, incluida la muerte, y los daños materiales, siempre que estos afecten a bienes o servicios objetivamente destinados al uso o consumo privados y en tal concepto hayan sido utilizados principalmente por el perjudicad (art. 129 del RDL 1/2007). Y, tercero, el concepto de producto defectuoso tiene un carácter normativo y debe interpretarse de acuerdo con los criterios que establece la ley. En particular, según el art. 137.1 del RDL 1/2007, "se entenderá por producto defectuoso aquél que no ofrezca la seguridad que cabría legítimamente esperar, teniendo en cuenta todas las circunstancias y, especialmente, su presentación, el uso razonablemente previsible del mismo y el momento de su puesta en circulación".

En el ámbito europeo, en el último trimestre de 2022, la Comisión Europea ha propuesto revisar Directiva 85/374/CEE del Consejo, de 25 de julio de 1985, relativa a la aproximación de las disposiciones legales, reglamentarias y administrativas de los Estados Miembros en materia de responsabilidad por los daños causados por productos defectuosos, para **adaptarla a** los cambios que conlleva la transición **hacia una economía circular y digital** , en cuando a la responsabilidad de productos que necesitan software o servicios digitales para funcionar, así como dispositivos inteligentes y vehículos autónomos. La reforma se entiende necesaria[61], primero, porque "no estaba claro desde el punto de vista jurídico cómo aplicar las definiciones y los conceptos de la Directiva sobre responsabilidad por productos defectuosos, que tienen décadas de antigüedad, a los productos de la economía digital moderna y la economía circular (por ejemplo, los programas informáticos y los productos que necesitan programas o servicios digitales para funcionar, como los dispositivos inteligentes y los vehículos autónomos); segundo, porque "la carga de la prueba (es decir, la necesidad, para obtener una

[61] Comisión Europea: Propuesta de Directiva del Parlamento Europeo y del Consejo sobre responsabilidad por los daños causados por productos defectuosos, Bruselas, 28.9.2022, COM (2022) 495 Final, 2022/0302 (COD).

compensación, de demostrar que el producto era defectuoso y que esto causó los daños sufridos) era difícil para las personas perjudicadas en casos complejos (por ejemplo, los relacionados con productos farmacéuticos, productos inteligentes o productos basados en inteligencia artificial); y, tercero, porque "las normas limitaban excesivamente la posibilidad de presentar reclamaciones de indemnización (por ejemplo, los daños materiales por un valor inferior a 500 EUR simplemente no son recuperables en virtud de la Directiva sobre responsabilidad por productos defectuosos)".

Por ello, la revisión de **la Directiva pretende "garantizar un mejor funcionamiento del mercado interior , la libre circulación de mercancías, la competencia sin distorsiones entre los operadores del mercado y un alto nivel de protección de los consumidores" y, en particular, "** garantizar que las normas de responsabilidad reflejen la naturaleza y los riesgos de los productos en la era digital y la economía circular"; "garantizar que siempre exista una empresa con sede en la UE que pueda ser considerada responsable de los productos defectuosos comprados directamente a fabricantes de fuera de la UE, a la luz de la creciente tendencia de los consumidores a comprar productos directamente en terceros países sin que exista un fabricante o importador establecido en la UE"; "aligerar la carga de la prueba en casos complejos y suavizar las restricciones a la presentación de reclamaciones, garantizando al mismo tiempo un equilibrio justo entre los intereses legítimos de los fabricantes, las personas perjudicadas y los consumidores en general; e, igualmente, "garantizar la seguridad jurídica adaptando mejor la Directiva sobre responsabilidad por productos defectuosos al nuevo marco legislativo creado por la Decisión núm. 768/2008/CE y a las normas de seguridad de los productos, y codificando la jurisprudencia relativa a la Directiva sobre responsabilidad por productos defectuosos".

Obviamente, todos estos requerimientos, que, fundamentalmente, pueden hacer responder civilmente a determinadas empresas por algunos de los supuestos en los que los autómatas y entidades de IA causen daños, así formulados, no sirven en el ámbito penal, donde no se puede establecer un patrón de responsabilidad objetiva como en el ámbito civil. Es más, en materia penal, la definición del ámbito de responsabilidad penal por productos defectuosos producidos y/o comercializados, generalmente, por empresas, no es sencillo de resolver y ha generado, siempre, notables controversias[62].

[62] Para hacerse una idea precisa de la problemática en el ámbito del Derecho penal, véase, Paredes Castañón, J.M.: "Capítulo 12…", en *ob. cit.*, págs. 601 ss. También, un clásico es el texto de Hassemer, W. Y Muñoz Conde, F.: *La responsabilidad por el producto en derecho penal*,

Explicaremos que la responsabilidad penal por el producto se puede concretar en dos momentos[63]: el primero, cuando el producto peligroso se ofrece en el mercado se afecta la salud pública y el Derecho Penal responde a través de los delitos de peligro contra la salud pública y, en el segundo, si el producto ya ha sido utilizado y se ha lesionado la salud individual o la vida, a través de los delitos de homicidio o lesiones, alcanzando la protección penal, en ambos casos, tanto a las modalidades de comisión dolosas como imprudentes. Como señaló en su momento Muñoz Conde[64], los delitos de peligro solo en una pequeña parte puede ser una solución para un tratamiento adecuado de la responsabilidad penal por el producto y los delitos de lesión, por su parte, solo es posible aplicarlos cuando se logre demostrar, efectivamente, con los medios de prueba admisibles y disponibles, una relación de causalidad entre el delito de peligro base o entre la acción peligrosa y la lesión del bien jurídico individual[65].

De todas formas, como se ha señalado con razón[66], la aplicación del modelo de la responsabilidad por el producto requiere, por una parte, que se pueda definir la IA como un producto comercial, cosa que, por el momento, es discutible que sea así cuando la IA sea un software, un "servicio"[67], y, por otra, que exista un defecto en el producto o que sus propiedades estén falsamente representadas, cosa que, en el caso de la IA compleja puede ser muy difícil probar el defecto, porque la IA puede causar daño sin ningún "defecto", en el sentido que exige la

Valencia, 1995.

[63] Corcoy Bidasolo, M.: "Responsabilidad penal derivada del producto. En particular la regulación legal en el Código Penal español: delitos de peligro", en Mir Puig, S. y Luzón Peña (coords.): *Responsabilidad penal de las empresas y sus órganos y responsabilidad por el producto*, Barcelona, 1996, pág. 248.

[64] Muñoz Conde, F.: "La responsabilidad penal por el producto en el Derecho español", en *Derecho & Sociedad*, núm. 49, 2017, págs. 278. El problema de la causalidad y su prueba en el proceso penal es, sin duda, el problema esencial que se le plantea al expediente de la responsabilidad penal por el producto (véase, Puppe, I.: "Problemas de imputación del resultado en el ámbito de la responsabilidad penal por el producto", en Mir Puig, S. y Luzón Peña (coords.): *Responsabilidad penal de las empresas y sus órganos y responsabilidad por el producto*, Barcelona, 1996, págs. 215 ss.; Vogel, J.: "La responsabilidad penal por el producto en Alemania: Situación actual y perspectivas de futuro", en *Revista Penal*, 2001, págs. 98 ss.; Kuhlen, L.: "Necesidad y límites de la responsabilidad penal por el producto", en *ADPCP*, vol. LV, 2002, págs. 68-69; Hilgendorf, E: "Relación de causalidad e imputación objetiva a través del ejemplo de la responsabilidad penal por el producto, en *ADPCP*, vol. LV, 2002, págs. 91 ss.).

[65] En el mismo sentido, Paredes Castañón, J.M.: "Capítulo 12: Responsabilidad penal…", en *ob. cit.*, págs. 604 ss.

[66] Abbott, R. y Sarch, A.: "Punishing Artificial Intelligence: Legal Fiction or Science Fiction", en *University of California, Davis, Law Review*, vol. 53, 2019, pág. 382.

[67] A los efectos de determinar la responsabilidad por daños causados por productos, el art. 136 del RDL 1/2007, considera "producto", cualquier bien mueble, aun cuando esté unido o incorporado a otro bien mueble o inmueble, así como el gas y la electricidad.

responsabilidad por el producto.

Por ello, no está en absoluto claro que vaya a ser posible aplicar las normas de la responsabilidad penal en el caso daños causados por autómatas o sistemas de IA.

D) El modelo de la responsabilidad objetiva

La Resolución del Parlamento Europeo de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un régimen de responsabilidad civil en materia de inteligencia artificial reconoce que el tipo de sistema de IA sobre el que el operador ejerce control es un factor determinante en lo que respecta a la responsabilidad, de modo que un sistema de IA que conlleve un alto riesgo inherente y actúe de manera autónoma potencialmente pone en peligro en mucha mayor medida al público en general. Por eso, habida cuenta de los retos jurídicos que plantean los sistemas de IA para los regímenes de responsabilidad civil existentes, entiende razonable establecer un régimen común de responsabilidad objetiva para los sistemas de IA autónomos de alto riesgo, subrayando, no obstante, que el enfoque basado en el riesgo, que puede incluir varios niveles de riesgo, debe basarse en criterios claros y una definición adecuada de "alto riesgo", así como ofrecer seguridad jurídica.

Por tanto, el art. 4 de dicha Resolución establece que el operador de un sistema de IA de alto riesgo será objetivamente responsable de cualquier daño o perjuicio causado por una actividad física o virtual, un dispositivo o un proceso gobernado por dicho sistema de IA, señalándose en del Reglamento cuáles son todos los sistemas de IA de alto riesgo y los sectores críticos en los que se utilizan. Además, los operadores de un sistema de IA de alto riesgo no podrán eludir su responsabilidad civil alegando que actuaron con la diligencia debida o que el daño o perjuicio fue causado por una actividad, un dispositivo o un proceso autónomos gobernados por su sistema de IA. Los operadores no serán responsables si el daño o perjuicio ha sido provocado por un caso de fuerza mayor.

Finalmente, se señala que el operador final de un sistema de IA de alto riesgo garantizará que las operaciones de dicho sistema de IA estén cubiertas por un seguro de responsabilidad civil adecuado en relación con los importes y el alcance de la indemnización previstos en el propio Reglamento (arts. 5 y 6) y el operador inicial garantizará que sus servicios estén cubiertos por un seguro de responsabilidad empresarial o de responsabilidad civil de productos adecuado en relación con los importes y el alcance de la indemnización previstos en el Reglamento (arts. 5 y 6).

Obviamente, pretender introducir tal cual, ese estatuto de responsabilidad objetiva para la IA en nuestro sistema penal, en el que no puede haber pena sin dolo o imprudencia (art. 5 del CP), no sería legalmente posible.

E) El modelo de la responsabilidad penal de las personas jurídicas: ¿también, la personificación de los robots?

La posibilidad de aplicar el modelo de la responsabilidad penal de las personas jurídicas a los supuestos de daños penalmente punibles provenientes de los sistemas de IA, con las necesarias adaptaciones legales, ha sido planteada por algunos autores, por servir de referencia como sistema de imputación de delitos a entidades que no constituyen personas físicas individuales. Así, por ejemplo, se ha señalado que la responsabilidad penal de las personas jurídicas, por haber constituido un avance innovador en el Derecho penal, y los modelos utilizados para sustentar tal avance pueden brindarnos pistas de mucho valor para desarrollar un plausible marco dogmático de la responsabilidad penal de entidades artificiales. Y lo que es más importante, demuestra un cierto grado de flexibilidad por parte de la ley penal cuando la política criminal así lo exija[68]. A partir de ahí, se ha desarrollado con mayor o menor profundidad y extensión el planteamiento.

a) La propuesta de N. Osmani

Una de las propuestas que han intentado desarrollar un poco más la idea de aplicar, a los supuestos de daños punibles causado por los sistemas de IA, el modelo de la responsabilidad penal de las personas jurídicas ha sido la de Osmani[69], que parte de la idea de que, en el ámbito de la responsabilidad criminal de los sistemas de IA, existe un vacío de impunidad por la dificultad de atribuir ninguna responsabilidad individual como consecuencia de los erráticos y dañinos comportamiento de los robots. De modo que, cuando se aborda este problema hay que tener presente que existen dos partes implicadas que centran el debate; por un lado, aquellos que desarrollan dichos sistemas, normalmente, grandes corporaciones[70], y, por otro, aquellos que reciben directamente el impacto de dicho desarrollo, que es la

[68] Freitas, P.M., Andrade, F. y Novais, P.: "Criminal Liability of Autonomous Agents: From the Unthinkable to the Plausible", en Casanova, P., Pagallo, U, Palmirani, M y Sartor, G. (eds.): *AI Approaches to the Complexity of Legal Systems*, AICOL 2013 International Workshops, Belo Horizonte, Brazil, July 21-27, 2013 and Bologna, Italy, December 11, 2013, Revised Selected Papers, Heidelberg, NewYork, Dordrecht, London, 2014, pág. 154.

[69] Osmani, N.: "The Complexity of Criminal Liability Systems", en *Masaryk University Journal of Law and Technology*, vol. 14, 2020, págs. 53 ss.

[70] Porque, como ha señalado Nemitz ["Constitutional Democracy and Technology in the Age of Artificial Intelligence", en *Philosophical Transactions of Royal Society a Mathematical, Physical*

sociedad en general. Teniendo en cuenta el crecimiento exponencial de los sistemas de IA, si no se encuentra el modo de colmar esa laguna el problema se va a hacer más grande cada día. Por eso, el punto de partida es analizar si, en lugar de sobre los individuos, intentar encontrar la fórmula para establecer un sistema de responsabilidad sobre las corporaciones[71]. Porque a la pregunta de si las corporaciones deben de rendir cuentas por acciones delictivas derivadas de las herramientas que despliegan en el mercado, se debe responder positivamente, hasta el punto de que la cuestión hoy no debería de ser si deben de responder sino cómo deben de responder[72].

No obstante, basar esta responsabilidad penal de las corporaciones en un modelo de responsabilidad vicaria o en el modelo del *respondeat superior*, sería un vano empeño porque, en el caso de la IA, no sería posible vincular a una acción individual, la del sistema de IA, a la responsabilidad de la persona jurídica. Por eso, se debe de pensar en un modelo de responsabilidad autónoma o de responsabilidad por un

---

and Engineering Sciences, 2018, págs. 1 ss. (en *{https://royalsocietypublishing.org/doi/epdf/10.1098/}rsta.2018.0089*)], las grandes corporaciones como Google, Facebook, Microsoft, Apple y Amazon (los *Frightful five*, en expresión utilizadada por F. Manjoo, en su conocido artículo, publicado el 20 de junio de 2016, en *The New York Times*, "*Tech's 'Frightful 5' Will Dominate Digital Life for Foreseeable Future*"), junto con algunas otras, dan forma no sólo la prestación de servicios basados en Internet a particulares, sino que, por su extrema rentabilidad, ejercen un poder económico que no sólo les garantiza un acceso desproporcionado a legisladores y gobiernos, sino que también les permite entregar ofrecer libremente, directa o indirectamente, apoyo financiero o en especie en todas las áreas de la sociedad relevantes para la opinión construyendo en democracia: gobiernos, legisladores, sociedad civil, partidos políticos, escuelas y educación, periodismo y educación periodística y, lo más importante, ciencia e investigación. La omnipresencia de estas corporaciones es así una realidad no sólo en términos técnicos, sino también con respecto a la atribución de recursos y asuntos sociales. Política, sociedad civil, ciencia, periodismo y los negocios tradicionalmente tratan de mantener una cierta distancia entre sí, algunos lo llaman un brazo relación de longitud. Pero hoy, los *Frightful five* están presentes en todos estos campos, para ganar conocimiento y aprender para sus propios fines, pero también, por decirlo en términos diplomáticos, para ganar simpatías y comprensión hacia sus preocupaciones e intereses. La indagación crítica sobre la relación de las nuevas tecnologías, como la IA, con los derechos humanos, la democracia y el estado de derecho deben partir de una mirada holística sobre la realidad de la tecnología y los modelos de negocios, tal como existen hoy en día, incluida la acumulación de poder tecnológico, económico y político en manos de los *Frightful five*, que son el núcleo del desarrollo y la integración de sistemas de IA en comercialmente viables servicios.

[71] Osmani, N.: "The Complexity…", en *ob. cit.*, págs. 67-68.

[72] *Ibidem*, págs. 69-70. "Los 'bolsillos profundos' de las corporaciones podrían darnos la respuesta a esta pregunta. La culpabilidad organizacional, en este sentido, se deriva de la expansión impulsada por las ganancias de las actividades comerciales de las megacorporaciones. Parece legítimo y necesario conjurar los peligros recurriendo a la responsabilidad de los que recogen los frutos del despliegue de la IA. De lo contrario, la lucha por imponer la responsabilidad por las acciones dañinas de los sistemas de IA les daría margen a las grandes corporaciones para que amplíen sus negocios, mientras los posibles actos dañinos de las máquinas inteligentes seguirían siendo una constante amenaza para la sociedad" (*Ibidem*, pág. 69).

hecho propio para que tenga sentido la integración, en dicho modelo, de la responsabilidad penal de los sistemas de IA, vinculando la responsabilidad de la corporación, directamente, por los riesgos asociados con la producción o utilización de las máquinas inteligentes[73]. El modelo del hecho propio significa un nuevo concepto de responsabilidad empresarial, sustentado en una visión realista de las corporaciones, que sugiere que estas tienen personalidades individuales e intenciones, que no se derivan de las acciones de sus agentes, de modo que, poner el foco en la culpabilidad organizacional, podría ser una herramienta poderosa en la asignación de responsabilidad por los riesgos asociados con la inteligencia de las máquinas. Porque, en última instancia, el desarrollo exponencial de la tecnología y la IA exige el correspondiente marco normativo para hacer frente a situaciones no previstas antes[74].

Por otra parte, debería de procederse a una definición de las diversas tipologías de acuerdo con el modelo de las *public welfare offences* (delitos relativos al bienestar público), en las que se prescinde del elemento mental (*mens rea*)[75]. Este modelo tiene dos propósitos: por una parte, representa una doctrina que fue diseñada como una necesidad para promover el orden social en un momento en que la revolución industrial trajo muchas amenazas nuevas a la sociedad. Las reglas del bienestar público (*welfare rules*) estaban destinadas a controlar industrias y actividades particulares que afectaban a la salud, la seguridad o el bienestar públicos. Similarmente, la revolución digital ha traído muchas amenazas nunca conocidas para los humanos. Tomando en consideración el gran impacto socioeconómico de las megacorporaciones, la doctrina de los delitos relativos al bienestar público podría allanar el camino para acusar a las corporaciones por delitos relaciona-

---

[73] *Ibidem*, pág. 71.

[74] *Ibidem*, pág. 72.

[75] Como ha explicado muy bien Varela ("Strict-Liability como forma de imputación jurídico-penal", en *InDret*, julio de 2012, págs. 4-5), en el sistema jurídico anglosajón, es un requisito básico de la imputación penal la concurrencia de un acto humano voluntario y exterior (*actus reus*) y, por regla general, también la imputación subjetiva entre el hecho y el sujeto (*mens rea*). "Por excepción, la responsabilidad penal puede fundarse simplemente en la imputación objetiva de un resultado, o bien en un hecho que dispense la prueba de la *mens rea* para determinados elementos del tipo penal (delitos de *strict liability*). De esto último se infiere que la *strict liability* supone que la responsabilidad penal se atribuye con independencia de la concurrencia o prueba del elemento subjetivo, es decir, con independencia del propósito (*purpose*), del conocimiento (*knowledge*), de la desconsideración (*recklessness*) o de la negligencia (*negligence*), ya sea para todos los elementos típicos del delito como para alguno solo de ellos. En atención a esta diferente exigencia del grado o intensidad de la responsabilidad estricta del delito, esta puede clasificarse en *pure* e *impure strict liability*. En relación con la figura de la pure *strict liability* ningún grado de *mens rea* es exigido para los elementos materiales del delito, mientras que en la de la impure *strict liability*, el elemento mental es requerido, por lo menos, para alguno de ellos". Los delitos relativos al bienestar público son delito de *pure strict liability*.

dos con los actos lesivos de la IA. En segundo lugar, esta doctrina omite los criterios de culpabilidad; se puede imponer una sanción independientemente de la intención del actor, por lo que el demandante no tiene que probar que el acusado actuó deliberadamente. La intención o culpabilidad se reemplaza por la asunción del riesgo que corre el actor al comprometerse en una determinada actividad[76].

La propuesta de Osmani, aparte de estar cuajada de inconcreciones, es, desde el punto de vista de nuestro Derecho, perfectamente inviable.

Parte de las inconcreciones tienen que ver con la propia debilidad del modelo del que parte, que pretende superar los inconvenientes del modelo de la responsabilidad derivada, porque el modelo de la responsabilidad propia u originaria o de la responsabilidad por hecho propio, no exige la transferencia a la persona jurídica de la responsabilidad de las personas naturales que se integran en su estructura organizativa, sino que es una responsabilidad de estructura "anónima" en cuanto a la intervención individual, aunque, de todos modos, resulte compatible con la atribución de responsabilidad individual a la persona o personas físicas que realizaren directamente la actuación delictiva. Este modelo *realista* de definición de la responsabilidad penal de las empresas, trata de reflejar, adoptando los paradigmas de la sociología de las organizaciones, que las corporaciones constituyen una realidad en sí diversa de la de los individuos, con sus propios y distintivos objetivos, su propia y distintiva cultura y su propia y distintiva personalidad, cultura y personalidad que son únicas y nacen de un número identificable de características (la estructura de la organización, los objetivos corporativos, la formación de los empleados, incentivos e indemnizaciones, etc.)[77].

La idea de la *organisational fault* o *culpabilidad por defecto de organización* establece la responsabilidad penal de la corporación cuando se pueda determinar que, producido un daño o perjuicio, ésta ha organizado su negocio de forma que las personas y las propiedades

[76] Osmani, N.: "The Complexity…", en *ob. cit.*, pág. 73.

[77] Clough, J.: "Bridging the Theoretical Gap: The Search for a Realist Model of Corporate Criminal Liability", en *Criminal Law Forum*, vol. 18, 2008, pág. 275; antes, en Gobert, J. y Punch, M.: *Rethinking Corporate Crime*, Bath, 2003, págs. 44 y 81. Como ha señalado entre nosotros Feijoo Sánchez (*El delito corporativo en el Código penal español. Cumplimiento normativo y fundamento de la responsabilidad penal de las empresas*, Cizur Menor, 2015, pág. 53), esa idea de que las organizaciones tienen una "vida propia", al margen de la vida y las decisiones de los individuos, no acaba de convencer porque la desvincula, absolutamente, de cualquier factor humano, por más que sea cierto que las organizaciones tengan sus dinámicas propias que se llegan a institucionalizar, de tal manera que es muy difícil que un solo individuo las pueda modificar o tenga la capacidad de influir decisivamente en ella. Y, sobre todo, lo que no son aceptables son las consecuencias normativas que se pretenden derivar de esa idea.

están expuestas a una victimización criminal o a un riesgo de daño no racional, o cuando la empresa ha fallado a la hora de establecer sistemas y mecanismos de evitación de riesgos criminales, o cuando la supervisión y vigilancia de aquellos a los que ha puesto en situación de cometer un delito o de causar daño es inadecuada, o cuando el *ethos* de la corporación o su cultura tolera o incentiva la causación de delitos[78].

En su propuesta, Osmani da por hecho que el "delito" que pueda cometer la IA se le atribuye a la organización, pero no explica si es que la utilización de la IA supone ya, de por sí, un defecto de organización y por qué es eso así. Porque si la empresa utiliza un sistema de IA, que no está diseñado para delinquir y que ha sido producido o utilizado de acuerdos con las normas y cautelas permitidas en el mercado, ¿dónde está la *cultura societaria defectuosa* que es fundamento de la responsabilidad penal de la corporación? Máxime si tenemos en cuenta que los hechos dañinos de una IA autónoma pueden tener lugar sin que haya habido defectos o carencias en su vigilancia y sin que se pueda decir que su ocurrencia haya sido predecible.

Finalmente, el modelo de Osmani sería imposible de aplicar en nuestro Derecho porque la responsabilidad de la persona jurídica exige un hecho de conexión, que es el realizado por la persona física individual, y no se podría nunca predicar de la IA esa cualidad. De hecho, el TS, desde su conocida STS núm. 154/2016, de 29 de febrero, a pesar de adherirse al modelo del hecho propio, ha venido señalando "*que el sistema de responsabilidad penal de la persona jurídica se basa, sobre la previa constatación de la comisión del delito por parte de la persona física integrante de la organización como presupuesto inicial de la referida*

[78] Gobert, J. y Punch, M.: *Rethinking…*, ob. cit., pág. 81. En nuestra doctrina, ha sido Díez Ripollés quien ha identificado cuatro variantes en el modelo del hecho propio: (i) El que imputa a la persona jurídica el hecho materialmente realizado por sus representantes o empleados, que se considera un hecho delictivo propio de la corporación, de modo que es en ella en quien ha de darse el injusto culpable del hecho, sin perjuicio de que la persona física ejecutora material deba responder por un injusto propio ligado a ese mismo hecho; (ii) El que imputa a la corporación un defecto de organización concreto, el cual ha facilitado o no ha impedido que sus representantes o empleados hayan realizado un hecho delictivo singular y, por tanto, será ese defecto de organización concreto lo que constituya el hecho delictivo propio de la sociedad; (iii) El que imputa a la sociedad una cultura corporativa defectuosa, la cual fomenta o no impide a lo largo del tiempo la realización por sus representantes o empleados de hechos delictivos como el concreto acaecido, de modo que esa cultura societaria defectuosa constituirá el hecho delictivo propio de la sociedad; y (iv) El que imputa al ente colectivo una reacción defectuosa frente al hecho delictivo, ya realizado, por sus representantes o empleados, y es la ausencia de ese comportamiento post-delictivo adecuado lo que constituye el hecho delictivo propio de la sociedad ("La responsabilidad penal de las personas jurídicas. Regulación española", en *InDret*, 1/2012, págs. 7 ss.).

responsabilidad, en la exigencia del establecimiento y correcta aplicación de medidas de control eficaces que prevengan e intenten evitar, en lo posible, la comisión de infracciones delictivas por quienes integran la organización". Y todo ello sin mencionar las objeciones que, en nuestro Derecho, podría plantear un diseño de tipos penales de strict pure liability.

b) La propuesta de M.E. Diamantis

La segunda de las propuestas, bastante más meditada y desarrollada, viene de la mano de M.E. Diamantis, que la ha ido esbozando a través de una serie de trabajos y, especialmente, en uno que ha publicado en este mismo año 2023, y que serviría para establecer las bases de la responsabilidad, ya sea civil o penal, por la utilización de los sistemas de IA[79].

Su punto de partida es, también, el de utilizar el modelo de la responsabilidad penal de las personas jurídicas porque, en la medida en que son las corporaciones las que producen y hacen uso de la IA y, lo más importante, porque "algunos algoritmos y corporaciones tienen una relación especial entre sí, también caracterizada por el beneficio y el control"[80]. Por eso, sus esfuerzos van dirigidos a establecer de qué manera las corporaciones pueden rendir cuenta por los daños causados por su personal digital, proponiendo lo que él denomina un *Labor Model*, un modelo laboral, en el que, finalmente, a los efectos de la responsabilidad penal de la corporación, él entiende que algunos algoritmos deben de ser tratados como empleados de la corporación[81]. A continuación, vamos a ver cómo lo explica.

El *Labor Model* parte de la base de que la regla general para la imputación civil y penal de las corporaciones se articula a través del

[79] Diamantis, M.E.: "Corporate Criminal Minds", en *Notre Dame Law Review*, vol. 91, 2016, págs. 2049 ss.; del mismo: "The Extended Corporate Mind: When Corporations Use AI to Break the Law", en *North Carolina Law Review*, vol. 94, 2020, págs. 893 ss.; del mismo: "Algorithms Acting Badly: A Solution from Corporate Law", en *The George Washington Law Review*, vol. 89, 2021, págs. 801 ss.; del mismo: "Employed Algorithms: A Labor Model of Corporate Liability for AI", en *Duke Law Journal*, vol. 72, 2023, págs. 797 ss.

[80] Diamantis, M.E.: "Employed Algorithms…", en *ob. cit.*, pág. 804.

[81] *Ibidem*, pág. 797. Hay quien, incluso, ha considerado, como hipótesis, la posibilidad de que un sistema de IA sea nombrado miembros del consejo de administración, hipótesis que se hizo cierta en Hong Kong, en donde una firma de capital riesgo, inscribió como miembro de su consejo, con la naturaleza de observador, a VITAL (*Validating Investment Tool for Advancing Life Science*), una máquina de inteligencia artificial capaz de hacer recomendaciones de inversión en el sector de las ciencias de la vida, llegando a la conclusión de que dicha posibilidad no es, hoy por hoy, legalmente posible ni en el Reino Unido ni en los EE. UU. (Zhao, J.: "Artificial Intelligence and Corporate Decisiones: Fantasy, Realty or Destiny", en *Catholic University Law Review*, vol 71, núm. 4, 2022, pág. 679).

que es el vigente modelo en los EE. UU.[82] del *respondeat superior*[83].

[82] La doctrina de *respondeat superior* es el modelo primario de responsabilidad penal corporativa en los tribunales federales y en la mayoría de los tribunales estatales de los EE. UU. (Law Commission: *Corporate Criminal Liability. An Options Paper*, 10 June 2022, pág. 62). Sin embargo, en Inglaterra y en Gales, la forma principal en que el Derecho penal atribuye responsabilidad a las empresas por delitos que requieren prueba de la culpa es de la "doctrina de la identificación", según la cual una entidad corporativa normalmente sólo será responsable de la conducta delictiva de una o más personas físicas que representen a la menta y la voluntad de la corporación (*Ibidem*, pág. 27).

[83] Históricamente, el sistema de responsabilidad por transferencia es el primero que surge, concretamente, en el ámbito del Derecho anglosajón, y está basado en una ancestral doctrina del *common law* según la cual los señores son absolutamente responsables de todas las acciones ilícitas y dañinas de sus sirvientes (Bernard, T.J.: "The Historical Development of Corporate Criminal Liability", en *Criminology*, vol. 22, n.º 1, 1984, pág. 5; Wells, C.: *Corporations and Criminal Responsibility*, 2.ª ed., New York, 2001, pág. 88). Para cuando las corporaciones hicieron su aparición como entidades relevantes en el panorama social y económico, este principio del *respondeat superior* había sido superado, excepto para aquellos casos en los que el señor, el *superior*, había dado su consentimiento o había ordenado la acción del dependiente (Wells, C.: *Corporations...*, ob. cit., pág. 88). El traslado de ese principio al ámbito de la responsabilidad civil y penal de las corporaciones no fue un problema, pues se consideró que la propia corporación era *el señor* y su empleado *el siervo*, si bien restringiendo, en un principio, la responsabilidad de la persona jurídica a los supuestos de delitos consistentes en una omisión o en el incumplimiento de obligaciones (como, por ejemplo, no reparar daños o no mantener caminos o cauces). Mayor dificultad hubo para responsabilizar a las corporaciones por delitos que exigían una conducta positiva, ya que se entendía que las personas jurídicas eran entes con forma legal pero carentes de corporeidad física y, por lo tanto, incapaces de realizar delitos que exigían un elemento físico (acción) (Weismann, A. y Newman, D.: "Rethinking Criminal Corporate Liability", en *Indiana Law Journal*, vol. 82, 2007, pág. 419). Más adelante, en el siglo XIX, con la recepción para el sistema de responsabilidad penal del sistema vigente en el ámbito del derecho de los *torts*, se consideró la posibilidad de hacer responder a las corporaciones, también, por delitos de acción, pero sólo para los llamados delitos de *strict liability* (responsabilidad objetiva), no para delitos con una dimensión moral que requieren de una *mala intención criminal* (Weismann, A. y Newman, D.: "Rethinking Criminal...", en *ob. cit.*, pág. 419. Esa restricción se mantuvo tanto en Inglaterra y Gales, gracias, fundamentalmente, a la decisión del caso *Queen v. Great North of England Railway* (1846) —que excluía a la posibilidad de que respondieran las corporaciones por crímenes que necesitaban una *mente corrompida*—, como en los Estados Unidos de América, en el caso *State v. Morris & Essex Railroad Co.* (1852) —cuya decisión se refería a los delitos que implicaba un *malus animus* en su comisión (Bernard, T.J.: "The Historical...", en ob. cit., pág. 8). A comienzos del siglo XX, sin embargo, y a partir de la opinión del vizconde Haldane, irónicamente expresada no en un caso de responsabilidad criminal sino civil, el caso *Lennard's Carrying Co. Ltd. V. Asiatic Petroleum Co. Ltd.* (1915), se inaugura para el Reino Unido la llamada *teoría de la identificación* o del *alter ego*, de modo que, según se recoge textualmente en dicha opinión, si bien una corporación es una abstracción, «que no tiene mente propia como no tiene cuerpo propio, de modo que su activa y directiva voluntad debe, consecuentemente, ser buscada en la persona de alguien que, para determinados propósitos, puede ser llamado agente, pero que es en realidad la mente directiva y la voluntad de la corporación; el verdadero ego y centro de personalidad de la corporación» (*Lennard's Carrying Co. Ltd. V. Asiatic Petroleum Co. Ltd.* [1915] AC 713) (Ferguson, G.: "Corruption and Corporate Criminal Liability", publicaciones del *International Centre for Criminal Law Reform and Criminal Justice Policy*, Vancouver, May 1998, pág. 5, en *{http://www.icclr.law.ubc.ca/Publications/Reports/FergusonG.PDF}*.). Por lo tanto, el título por el cual se produce dicha transferencia de responsabilidad de la persona física a la persona

El *respondeat superior* tiene dos requisitos: El primero es que, en el momento de cometer el hecho contrario a la ley, el empleado debe haber tenido la intención de beneficiar a la corporación y, el segundo, es que debe de haber estado trabajando dentro del alcance de su empleo. De modo que la corporación no responderá si la única motivación para violar las normas ha sido la de beneficiarse exclusivamente el empleado o para dañar a su empleador. Igualmente, la corporación tampoco responderá si el empleado no estaba trabajando *para* la corporación, es decir, si el empleado no estaba intuitivamente actuando *como* la corporación, de modo que su responsabilidad se limita a los casos donde los empleados no parecen ser terceras personas, sino que son la encarnación de la corporación. La corporación, entonces, en la medida en que obtiene rendimientos del trabajo de sus empleados, también tiene que asumir los costes de su actuación y, además, no está obligada a hacer lo imposible pero sí lo está a ejercer ciertos controles sobre dichos empleados que trabajan para ella[84].

Hoy por hoy, la ley se limita a considerar empleados, a estos efectos, a los seres humanos, porque ningún algoritmo sería capaz de satisfacer esos dos requisitos. Los algoritmos no pueden pretender beneficiar a ninguna corporación porque, al carecer de mente, no intentar o pretender nada y, además, sin una relación laboral, de algún tipo con la empresa, los algoritmos nunca operar dentro del ámbito del empleo. Pero, al limitar la responsabilidad del superior sólo a los casos de comportamientos de empleados humanos, la ley adopta una comprensión muy superficial, que pasa por alto la verdadera flexibilidad de dicha doctrina. "Principios más profundos actúan en el *respondeat superior*. Por siglos, estos principios se manifestaron en requerimientos doctrinales específicos adaptados a una suposición sobre la naturaleza de la producción empresarial, que solo es posible a través del esfuerzo humano. Esa suposición ya no se sostiene en la era actual, donde los algoritmos van reemplazando rápidamente el trabajo humano. Recuperando los principios que subyacen tras el *respondeat superior*, pueden salir a la luz versiones más generalizadas de sus dos elementos que permita aplicarlos, de manera flexible, tanto al trabajo humano como al trabajo digital"[85].

De modo que, para aplicar el concepto de empleado a la IA, como propone el autor en su *Labor Model*, de modo que, a los efectos de determinar la responsabilidad penal de la empresa, la IA sea considerada un

jurídica es la identificación de la voluntad del primero con la del segundo, de modo que la persona que actúa no es que actúe para la corporación, *es que es la corporación.*

[84] Diamantis, M.E.: "Employed Algorithms…", en *ob. cit.*, págs. 844-847.

[85] *Ibidem*, pág. 847.

empleado, hace faltan dos innovaciones. "La primera es reconocer que las corporaciones pueden emplear algoritmos. Por tanto, "sean cuales sean los límites formales sobre quién o qué puede ser considerado empleado en otros contextos a los efectos de valorar la responsabilidad de la empresa, el concepto legal de empleado debe de ampliarse para dar cobertura al 'algoritmo empleado´. La segunda innovación es definir qué son los algoritmos empleados para generalizar los dos elementos del *respondeat superior.* Dado que los elementos pueden hacer frente a los desafíos de aplicación en el contexto de empleados humanos, también pueden ser capaces de resolver los similares retos estructurales de los algoritmos"[86].

Señala Diamantis que los elementos fundamentales de la doctrina del *respondeat superior* son, como se ha visto, la "prueba de beneficio-control". La prueba del beneficio-control deriva de los más profundos principios que están en juego en la intención de beneficiar a la empresa que orienta la acción del empleado y en la perspectiva del empleo, en el ámbito del cual actúa el empleado. La idea general es que estos elementos están diseñados para asegurar que una corporación solo es responsable por la infracción legal de un empleado si la corporación esperaba obtener beneficios de la actuación del empleado y controlaba la conducta del empleado al momento de la infracción. Como estos elementos no son aplicables a los algoritmos, los tribunales tendrían que averiguar, directamente, si una corporación reclama beneficios sustanciales del funcionamiento del algoritmo y ejerce un control sustancial sobre él. A la hora de indagar sobre los beneficios, los tribunales deberán debería evitar pasar por alto los beneficios indirectos. Incluso si un algoritmo no proporciona un flujo de ingresos inmediato, podría generar rentabilidad haciendo que las operaciones sean más eficientes o proporcionando datos para ayudar a informar las estrategias comerciales. Medir el control corporativo sobre algoritmos requiere un enfoque multifacético. Los poderes de control relevantes incluyen el poder de diseñar el algoritmo, terminar su operación, modificarlo, supervisarlo y anularlo. Ninguno de estos poderes por sí solo es, necesariamente, determinante del control corporativo sobre un algoritmo, pero cuantos más poderes tiene una corporación, más control tiene. En resumen, el *Labor Model* de responsabilidad empresarial por daños causados por los algoritmos refleja, en gran medida, el enfoque del *respondeat superior* hacia el algoritmo empleado causa. Las corporaciones son potencialmente responsables de los daños que sus algoritmos empleados ocasionan. Una corporación emplea un algoritmo si ejerce

[86] *Ibidem*, págs. 847-848.

un control beneficioso sobre él. La única pregunta que queda es si el Modelo Laboral es una solución atractiva para evitar la laguna de impunidad y los desafíos de aplicación que los algoritmos y, obviamente, el autor entiende que sí lo es[87].

La propuesta de Diamantis no nos es, en absoluto, extraña. Para el ámbito de la responsabilidad civil, ya la sugirió como una posibilidad Navas Navarro[88], quien ha señalado que, "en el caso de sistemas de IA autónomos, la responsabilidad por hecho ajeno podría basarse en la relación existente entre el humano y el sistema pues el primero se beneficia de la actividad que lleva a cabo el segundo. En este caso, la responsabilidad de la que se suele hablar es la del 'empresario' por los hechos de sus 'dependientes' (art. 1903.4 CC). En realidad, la 'dependencia' cubre todos aquellos casos en los que una persona o, en nuestro caso, un sistema de IA actúa subordinada a las instrucciones de otra que sería el empresario o, como se la conoce en textos legales europeos, el 'principal', sin que se ciña al contrato de trabajo o de servicios. Por tanto, existe un 'principal' (el humano) y un 'auxiliar' (el sistema de IA)".

Ahora bien, como ella misma señala, es presupuesto de la aplicación de la responsabilidad del principal por hechos llevados a cabo por los auxiliares la relación de subordinación o de dependencia del auxiliar respecto de las indicaciones del principal. Si el sistema de IA es autónomo, por más que pueda, en algunos casos, tener instrucciones previas, dada su capacidad de aprendizaje y autonomía a la hora de tomar decisiones, se puede decir que estos sistemas actúan sin estar sujetos al control del principal que se beneficia de su actividad. No hay tal vínculo de subordinación que caracteriza la responsabilidad vicaria del principal por los hechos de su auxiliar. Más bien se asemejaría a la situación del contratista independiente. Ello, no obstante, "hacer cargar al principal con las consecuencias dañosas del comportamiento del sistema de IA del que se sirve y se beneficia concuerda con la idea de que quien disfruta del beneficio que genera el sistema de IA es el principal". Además, "que el principal asuma las consecuencias dañosas generadas por el sistema de IA sería también un incentivo para que despliegue un mayor control de los riesgos que puedan resultar de la actividad de aquél en la medida en que se encuentran en mejor posición para establecer mecanismos de prevención"[89].

De todas formas, la aplicación de este instituto requeriría poder considerar al sistema de IA legalmente imputable, mediante la concesión

[87] *Ibidem*, págs. 848.

[88] Navas Navarro, S.: "Reglas especiales…", en *ob. cit.*, págs. 60-61.

[89] *Ibidem*, pág. 61.

de algún tipo de estatuto legal, ya sea de personalidad o, simplemente, que se le incluya en la categoría de "sujeto de derechos", extendiéndose, entonces, la aplicación por analogía la responsabilidad vicaria del principal por los hechos del auxiliar[90]. En lo que sería el ámbito de la responsabilidad civil, por ejemplo, en su momento, el Parlamento Europeo, en Resolución de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica, en su apartado 59, pidió a la Comisión que, cuando realizara una evaluación de impacto de su futuro instrumento legislativo sobre la materia, "explore, analice y considere las implicaciones de todas las posibles soluciones jurídicas, tales como: … f) **crear a largo plazo una personalidad jurídica específica para los robots** , de forma que como mínimo los robots autónomos más complejos puedan ser considerados personas electrónicas responsables de reparar los daños que puedan causar, y posiblemente aplicar la personalidad electrónica a aquellos supuestos en los que los robots tomen decisiones autónomas inteligentes o interactúen con terceros de forma independiente"[91].

La propuesta de personificación de los robots o de los sistemas de IA, es decir, la propuesta de conceder algún tipo de estatus legal personal a las máquinas o a los sistemas está bastante extendida. Autores de la solvencia intelectual y jurídica como Teubner[92] se han mostrado firmes partidarios de conceder un cierto estatus subjetivo de *e-personas* a los *Softwareagentem*, (agentes de software) que serían los sistemas de IA capaces de aprender a partir de los datos que se les suministran y de elegir en situaciones de incertidumbre (los sistemas de *deep learning*), que los cualifique como actores responsables, y no como meros instrumentos de los humanos o de las organizaciones humanas so pena de que se incremente progresivamente el número de accidentes que terminarán por no poder imputarse a nadie a título de responsabilidad[93].

---

[90] *Ibidem*, pág. 62.

[91] De todas formas, la posición de la Unión Europea, que es más compleja y abandona esa idea de la personificación en textos ulteriores [así, por ejemplo, en la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión, Bruselas, 21.4.2021, COM (2021) 206 final, 2021/0106 (COD)], la explica magníficamente bien y con detalle, Navas Navarro, S.: "Reglas especiales de responsabilidad civil en caso de daños ocasionados por sistemas de inteligencia artificial. Futura normativa europea y derecho vigente", en *Daños ocasionados por sistemas de inteligencia artificial: Especial atención a su futura regulación*, Granada, 2022, págs. 35 ss.

[92] Teubner, G.; "Digitale Rechtssubjekte?", en *ob. cit.*, págs. 155 ss.

[93] Señala Teubner: "¿Cuáles son las lagunas específicas en la responsabilidad? En la ciencia de la información se suelen mencionar los siguientes escenarios: Las deficiencias surgen en la práctica cuando el software es producido por equipos; cuando la gestión las decisiones son tan importantes como la programación decisiones; cuando la documentación de los requisitos y

Porque, "la falta de responsabilidad –señala el autor— surge siempre que la doctrina jurídica insiste en dar respuesta a las nuevas realidades digitales exclusivamente con instrumentos conceptuales tradicionales. Sin embargo, para seguir el ritmo de los desarrollos digitales, al menos hasta cierto punto, doctrina se ve obligada a reaccionar ante los hasta ahora desconocidos agentes de software con construcciones auxiliares cuestionables. En el campo del derecho contractual, la doctrina sostiene firmemente que sólo las personas físicas están en condiciones de hacer declaraciones jurídicamente vinculantes. Por lo tanto, el derecho contractual tiene que trabajar con ficciones problemáticas. En el derecho de la responsabilidad contractual y extracontractual, si los daños son imputables a una red humano-computadora, debe atribuir el hecho causante del daño exclusivamente a la parte de la acción del ser humano y entonces ya no es capaz de determinar en detalle los requisitos de responsabilidad. Y las normas sobre responsabilidad objetiva, por una parte, van demasiado lejos y, por otro lado, no lo suficientemente lejos porque tratan el riesgo digital como el mero riesgo causal de una materia peligrosa. Finalmente, hay perplejidad general con respecto a la creación de redes de sistemas multiagente"[94].

Es cierto que su propuesta no es la de conceder una plena subjetividad[95]. Además, a tales efectos, el modelo no debería de ser la

---

las especificaciones juega un papel importante en el código resultante; cuando, a pesar de las pruebas de precisión del código, mucho depende de los componentes 'listos para usar' cuyo origen y la precisión no están claros; cuando el rendimiento del software es el resultado de los controles que lo acompañan y no del programa creado; cuando los instrumentos automatizados se utilizan en el diseño del software; cuando el funcionamiento de los algoritmos está influenciado por sus interfaces o, incluso, por el tráfico del sistema; cuando el software interactúa de manera impredecible; cuando el software trabaja con probabilidades o tiene adaptabilidad o es el resultado de otro programa" (*Ibidem*, pág. 3 de la edición digital).

[94] *Ibidem*, págs. 4-5 de la edición digital. Si, en estos casos, la ley reacciona al uso del software autónomo solo con conceptos convencionales y, por lo tanto, deja lagunas abiertas en la responsabilidad, el daño no se distribuye colectivamente en toda la sociedad, sino que se aplica un despiadado *casum sentit dominus*. Aquí es donde la crítica es masiva. Imponer las consecuencias del daño sobre la parte perjudicada es criticado con razón como un mal en términos de política legal y como fundamentalmente injusto. Siempre que, en tales situaciones, la falla de agentes autónomos de software permanece libre de responsabilidad, esto crea falsos incentivos para los operadores, productores y programadores. Y la voluntad de la sociedad de explotar al máximo el prometedor potencial de los agentes de software autónomos están disminuyendo. En cualquier caso, sobre todo, contradice el postulado de la justicia que mantiene una conexión necesaria entre decisión y responsabilidad (*Ibidem*, pág. 5 de la edición digital).

[95] Las demandas de una personalidad digital completa ignoran la realidad actual, porque los vacíos de responsabilidad, hasta el día de hoy, no son, en absoluto, una cuestión de las máquinas actuando en su propio interés sino, más bien, siempre en el interés de los seres humanos o de las organizaciones, especialmente empresas comerciales. Económicamente hablando, es una relación principal-agente en el que el agente es dependiente pero autónomo. Los agentes de software son esclavos digitales, pero esclavos con habilidades sobrehumanas. Y la rebelión de

persona física sino, más bien, la persona jurídica, que tampoco tienen autoconciencia ni voluntad subjetiva, aunque actúan como centros de imputación de decisiones y conducta con relevancia jurídica[96] sino una capacidad jurídica en un sentido hoy inexistente, en el que la sociedad y el Derecho estuvieran dispuestos a atribuir y asignar a estos agentes recursos económicos propios con los que pudieran perseguir finalidades autónomas de lucro[97]. En resumen, los nuevos riesgos digitales –el riesgo de autonomía, la asociación de riesgo y el riesgo de red– confrontan al derecho privado con el reto de redefinir el estatuto jurídico de los agentes de software autónomos, no en el sentido que sugieren la personificación completa, sino calibrando cuidadosamente el estatus legal de los algoritmos en base a su concreto rol. Para el riesgo de autonomía, la respuesta adecuada es otorgar a los agentes de software el estado de parcial personalidad jurídica. Sus decisiones autónomas deben ser jurídicamente vinculantes y deben dar lugar a las consecuencias de la responsabilidad. Esto limita su subjetividad jurídica para celebrar contratos vinculantes para otros como representante. Al mismo tiempo, en los casos de responsabilidad contractual y extracontractual, deben ser reconocidos como auxiliares legalmente capaces, para que la mala conducta de la máquina en sí misma (y no simplemente la conducta de las empresas detrás de él) constituya un incumplimiento del deber por el cual las empresas deben ser consideradas responsables[98].

---

los esclavos debe evitarse. La personalidad legal solo sería apropiada si se les concediera la propiedad de recursos in la economía y en la sociedad con los que ellos podrían perseguir su propio interés (*Ibidem*, pág. 6 de la edición digital).

[96] *Ibidem*, págs. 8-9 de la edición digital.

[97] Recuerda Teubner que la sección 14 de la *Uniform Electronic Transactions Act* de los EE. UU. establece que un contrato puede crearse por la interacción de agentes electrónicos de las partes, incluso aunque un individuo no fuera consciente de, o no revisara las acciones a los agentes electrónicos, o los términos resultantes del acuerdo ("Digitale Rechtssubjekte?", en *ob. cit.*, pág. 18 de la edición digital).

[98] *Ibidem*, pág. 38 de la edición digital. Otra autora, Beck, ha entendido que "es posible crear un estatus legal [para los robots autónomos], que sería solo un 'símbolo tangible' para la cooperación de todas las personas que construyen y usan ese robot específico. La jurisprudencia podría establecer que algunas máquinas autónomas tengan el estatus de 'persona electrónica' con derechos y obligaciones específicas. Esto se aplicaría solo a contextos particulares e incluiría máquinas autónomas que tienen un cierto grado de autonomía legal. Sería apropiado para todas las máquinas con inteligencia artificial que automáticamente toman decisiones o de alguna manera interactúan con otras personas, como mediante la celebración de contratos o causando daños a los intereses legales de una persona. Esta personalidad jurídica para los robots también sería, simplemente, la agrupación de todas las responsabilidades legales de las distintas partes (usuarios, vendedores, productores, etc.). Esta estructura legal tendría efectos en el derecho civil: las sentencias podrían dictarse directamente contra las personas electrónicas (y estarían cubiertas por sus activos, pagados por las partes involucradas en su creación y formación).

En realidad, como ha señalado Carrasco Perera[99], en el ámbito de la responsabilidad contractual, aunque la propuesta sea ingeniosa, en términos prácticos no va más allá de lo que se llegaría negando toda imputación autónoma al agente de software y predicarla directamente del gestor humano; porque es lo mismo que el agente humano responda del incumplimiento objetivamente por el riesgo de emplear determinados recursos para el cumplimiento o decir que el agente humano responde objetivamente de la conducta del agente artificial en cuanto esta no es más que una conducta subordinada de *auxiliares de cumplimiento*, de los que se responde objetivamente (art. 1.596 CC). En el ámbito de la responsabilidad extracontractual, el gestor del autómata no estaría sujeto a una responsabilidad directa y primaria de naturaleza objetiva sobre la base del riesgo creado por el empleo de IA autónoma, sino a una responsabilidad vicaria por hecho de otro, incardinadle en nuestro art. 1.903 CC.

En cualquier caso, propuestas de personificación, casi siempre limitada, están presentes en el ámbito de la doctrina sobre la responsabilidad civil desde hace muchos años[100]. Incluso, su viabilidad no deja de contar con algunos ejemplos un tanto bizarros de esas personificaciones como la decisión de Japón de conceder un permiso especial de residencia (*koseki*), en el año 2010, a un robot terapéutico de nombre *Paro*[101] o la de Arabia Saudita de conceder la ciudadanía, en el otoño de 2017, a un humanoide, de nombre *Sophia*[102].

Sin embargo, parece plausible que cuando un mal funcionamiento específico de la máquina pueda atribuirse a una mala conducta dolosa o negligente grave, la transferencia de este pago pueda ser emitida a una de las partes involucradas en la creación o uso de la máquina" (Beck, S.: "Intelligent agents and criminal law. Negligence, diffusion of liability and electronic personhood", en *Robotics and Autonomous Systems*, 86, 2016, págs. 141-142.). No obstante, esta misma autora no ve tan sencillo trasladar esta perspectiva de la personalidad jurídica del robot al ámbito del Derecho penal, como no ha sido fácil en Alemania hacer lo propio con las personas jurídicas, que en dicho país no pueden ser penalmente responsables (*Ibidem*, pág. 142).

[99] "A propósito de un trabajo de Gunter Teubner sobre la personificación civil de los agentes de Inteligencia Artificial avanzada", en *Centro de Estudios de Consumo*, 11 de enero de 2019, pág. 3, en *{http://centrodeestudiosdeconsumo.com/index.php/2-principal/3880-a-propósito-de-un-trabajo-de-gunter-teubner-sobre-la-personificación-civil-de-los-agentes-de-inteligencia-artificial-avanzada}*.

[100] Solum, L.B.: "Legal Personhood for Artificial Intelligences", en *North Carolina Law Review*, vol. 70, 1992, págs. 1231 ss. Una posición interesante, al respecto, la mantiene Navas Navarro, S.: "Reglas especiales…", en *ob. cit.*, págs. 55 ss.

[101] Con lo cual, se le conceden derechos de ciudadanía a un robot que determinadas poblaciones minoritarias no pueden obtener sino en cuatro generaciones (Koudela, P.: "Robots Instead of Immigrants: The Positive Feedback of Japanese Migration Policy on Social Isolation and Communication Problems", en Asia-Pacific Social Science Review, 19, 2019, pág. 97).

[102] No deja de tener su ironía que un país que no concede plenos derechos a las mujeres se los conceda a un humanoide (Wootson, C.R.: "Saudi Arabia, Which Denies Women Equal Rights, Makes a Robot a Citizen", en *The Washington Post*, 29 de octubre de 2017).

La posibilidad de que esta propuesta sea viable en nuestro CP pasaría por entender que los sistemas de IA pueden ser incluidos entre los sujetos mencionados en el art. 31 bis, núm. 1, apartado b), es decir, entre aquellos que, estando sometidos a la autoridad de las personas físicas mencionadas en el apartado a) anterior, han podido realizar los hechos por haberse incumplido gravemente por aquéllos los deberes de supervisión, vigilancia y control de su actividad atendidas las concretas circunstancias del caso. Por más que estemos de acuerdo con el sector doctrinal que entiende que el ámbito personal, en este caso del apartado b) del art. 31 bis, núm. 1, es deliberadamente amplio y no se ciñe, por tanto, a los trabajadores y mandos intermedios de la empresa, "sino que apela a todo sujeto que opere integrado bajo el ámbito de dirección de los administradores", de modo que se pueden incluir, sin mayores objeciones, a sujetos que, sin estar vinculados formalmente a la persona jurídica por un contrato laboral o mercantil, "desarrollan para ella sus actividades sociales integrados en su ámbito de dominio social"[103], hoy por hoy sería imposible integrar en ese concepto a los sistemas de IA. Porque el CP, tanto en el apartado a) como en el apartado b) del art. 31 bis, núm. 1, se está refiriendo a las personas físicas individuales que cometen los delitos que generan la responsabilidad penal de la empresa. Y esos es así porque los tipos penales están diseñados para su aplicación a conductas humanas, no a conductas de personas jurídicas ni a conductas de animales ni a "conductas" de sistemas de IA.

Por tanto, la misma dificultad que ve Navas Navarro en el ámbito de la responsabilidad civil se produce, de la misma manera, en el ámbito de la responsabilidad penal. Si no existe algún tipo de estatuto legal que personifique la IA y permita que esta esté equiparada a la persona física, a determinados efectos, la propuesta no es viable.

Claro que, frente a esta conclusión, la pregunta inmediata sería la siguiente: ¿se deberían *personificar* los sistemas de IA autónomos para que pudieran ser considerados, a efectos de generar la responsabilidad penal de la persona jurídica, empleados o auxiliares de esta? En el ámbito de la responsabilidad penal de las personas jurídicas, y en relación con la negativa de la mayoría de la doctrina alemana a que, *de lege ferenda*, a las personas jurídicas se las pudiera considerar capaces de acción, en su momento Jakobs[104], que entendió esa negativa como injusta, recordó que, para las personas físicas, el comprobar si

[103] Dopico Gómez-Aller, J.: "Capítulo 1. Responsabilidad de personas jurídicas", en *Reforma penal. Ley Orgánica 5/2010*, Memento Experto, Madrid, 2010, pág. 19, marginal 172.

[104] Jakobs, G.: *Derecho penal. Parte general. Fundamentos y teoría de la imputación*, trad.: J. Cuello Contreras y J.L. Serrano González de Murillo, 2ª ed. corregida, Madrid, 1997, pág. 183.

había concurrido la acción no se resolvía desde un punto de vista exclusivamente naturalístico, sino que "lo importante es la determinación valorativa del sujeto de la imputación, es decir, qué sistema psicosomático se trata de juzgar por sus efectos exteriores". Así, en la determinación del sujeto, no cabe fundamentar que el sistema que ha de formarse deba de estar siempre compuesto de los ingredientes de la persona física, mente y cuerpo, y no de los de la persona jurídica, estatutos y órganos. "Más bien los estatutos y los órganos de una persona jurídica se pueden también definir como sistema, en el cual lo interno –paralelamente a la situación en la persona física— no interesa (...), pero sí interesa el *output*. Las actuaciones de los órganos con arreglo a sus estatutos se convierten en acciones propias de la persona jurídica". Por eso, las acciones del órgano de una persona jurídica llevadas a cabo de acuerdos con las competencias que tiene atribuidas en el estatuto son acciones de la propia persona jurídica[105].

Por más que nos invada una especie de borrachera de *virtualismo*, en la que hay quien empieza a ver como algo normal el que se les conceda, de una u otra forma, un estatus personal a cosas o animales, en mi opinión, el problema de la personificación de los sistemas de IA no tiene comparación posible con el de la personificación de las personas jurídicas, porque detrás de las decisiones de una persona jurídica, incluso aunque se piense que la persona jurídica comete su propio delito, está siempre una decisión humana, o una decisión vinculable a una o varias personas físicas individuales, cosa que no ocurre en el caso de los sistemas autónomos de IA. Por lo demás, la personificación de los sistemas de IA es, a efectos penales, perfectamente inútil porque estos carecen de capacidad de acción, culpabilidad y pena en términos penales.

De hecho, obsérvese que, en el fondo, la propuesta del *Labor Model* de Diamantis se ve obligada a establecer un sistema de responsabilidad objetiva de la empresa que va a tener que responder de los daños del robot porque lo controla (aparentemente, al menos) y saca provecho de él, lo que, desde el punto de vista penal, no es admisible. Pero sí, a lo mejor, desde el punto de vista civil. Dicho, en otros términos, la prop-

[105] Bacigalupo, S.: La responsabilidad penal de las personas jurídicas, Barcelona, 1998, pág. 154. Peor resuelto y diseñado está, en Jakobs, el problema de la culpabilidad de la persona jurídica (véase, Derecho penal..., ob. cit., págs. 183-184), hasta el punto de que años más tarde de su inicial propuesta la corrige, al concluir que no se puede hablar de una culpabilidad de la persona jurídica porque no es una persona autoconsciente y comunicativamente competente y tampoco cabe imponer penas por su incapacidad de sufrir el dolor penal porque no tiene capacidad de sufrir, negando la posibilidad de responsabilizar penalmente a la persona jurídica ("¿Punibilidad de las personas jurídicas?", en Montealegre Lynett, E. (coord..): Libro Homenaje al profesor Günther Jakobs. El Funcionalismo en Derecho penal, Bogotá, 2003, págs. 341 ss.).

uesta de Diamantis me parece perfectamente válida para establecer el mecanismo de imputación de una responsabilidad civil (objetiva) por daños a la empresa; incluso, sin necesidad de darle a la IA un estatus de empleado, por más que, jurídicamente, en algunos aspectos, esa situación de empleado de la IA pudiera tener alguna concordancia con la de empleados personas físicas, al modo que lo explica Diamantis.

Pero, personalmente, creo que no es posible establecer mecanismos para hacer responder penalmente a los sistemas autónomos de la IA cuando causen daños a terceros, ni por la vía de su personificación ni por la vía propuesta, por ejemplo, por Osmani, y creo que hay que desistir de mirar, a estos efectos, al modelo de la responsabilidad penal de las personas jurídicas porque tampoco es capaz de dar cobertura a una posible responsabilidad penal de la IA. Los mecanismos de respuesta deben de ir por otros cauces y la responsabilidad civil, en este ámbito, como de hecho lo está orientando la UE, es un mecanismo mucho más idóneo y adecuado que la respuesta que puede dar el Derecho penal.

Bibliografía

BIBLIOGRAFÍA

Abadías Selma, A.: *El Derecho penal frente a la discriminación laboral algorítmica*, Cizur Menor, 2023.

Abbott, R. y Sarch, A.: "Punishing Artificial Intelligence: Legal Fiction or Science Fiction", en *University of California, Davis, Law Review*, vol. 53, 2019, págs. 323 ss.

Azzutti, A., Ringe, G.R. y Stiehl, H.S.: "Machine learning, market manipulation, and collusion on capital markets: Why the 'black box' matters", en *University of Pennsylvania Journal of International Law*, vol. 43, 2021, págs. 79 ss.

Bacigalupo, S.: *La responsabilidad penal de las personas jurídicas*, Barcelona, 1998.

Barocas, S. y Selbst, A.D.: "Big Data's Disparate Impact", en *California Law Review*, vol. 104, 2016, págs. 671 ss.

Barja de Quiroga, J.: *Tratado de Derecho Penal. Parte General*, 2º ed., Cizur Menor, 2018.

Beck, S.: "Intelligent agents and criminal law. Negligence, diffusion of liability and electronic personhood", en *Robotics and Autonomous Systems*, 86, 2016, págs. 138 ss.

Bernard, T.J.: "The Historical Development of Corporate Criminal Liability", en *Criminology*, vol. 22, núm. 1, 1984, págs. 3 ss.

Bird, K.: "Natural and Probable Consequences Doctrine: Your Acts Are My Acts", en *Western State University Law Review*, 43, 2006, págs.

43 ss.

Carrasco Perera, A.: "A propósito de un trabajo de Gunter Teubner sobre la personificación civil de los agentes de Inteligencia Artificial avanzada", en *Centro de Estudios de Consumo*, 11 de enero de 2019, en *{http://centrodeestudiosdeconsumo.com/index.php/2-principal/388o-a-propósito-de-un-trabajo-de-gunter-teubner-sobre-la-personificación-civil-de-los-agentes-de-inteligencia-artificial-avanzada}*.

Clough, J.: "Bridging the Theoretical Gap: The Search for a Realist Model of Corporate Criminal Liability", en *Criminal Law Forum*, vol. 18, 2008, págs. 267 ss.

Corcoy Bidasolo, M.: "Responsabilidad penal derivada del producto. En particular la regulación legal en el Código Penal español: delitos de peligro", en Mir Puig, S. y Luzón Peña (coords.): *Responsabilidad penal de las empresas y sus órganos y responsabilidad por el producto*, Barcelona, 1996, págs. 247 ss.

De Répide, P.: *Las Calles de Madrid*, Madrid, 1971.

Diamantis, M.E.: "Corporate Criminal Minds", en *Notre Dame Law Review*, vol. 91, 2016, págs. 2049 ss.

- "The Extended Corporate Mind: When Corporations Use AI to Break the Law", en *North Carolina Law Review*, vol. 94, 2020, págs. 893 ss.
- "Algorithms Acting Badly: A Solution from Corporate Law", en *The George Washington Law Review*, vol. 89, 2021, págs. 801 ss.
- "Employed Algorithms: A Labor Model of Corporate Liability for AI", en *Duke Law Journal*, vol. 72, 2023, págs. 797 ss.

Díez Ripollés, J.L.: "La responsabilidad penal de las personas jurídicas. Regulación española", en *InDret*, 1/2012, págs. 1 ss.

Dopico Gómez-Aller, J.: "Capítulo 1. Responsabilidad de personas jurídicas", en *Reforma penal. Ley Orgánica 5/2010*, Memento Experto, Madrid, 2010.

Escott, E.: "What Are the 3 Types of AI? A Guide to Narrow, General and Super Artificial Intelligence, *Codebots*, 24 October 2017 (*{https://codebots.com/artificial-intelligence/the-3-types-of-ai-is-the-third-even-possible}*).

Feijoo Sánchez, B.: *El delito corporativo en el Código penal español. Cumplimiento normativo y fundamento de la responsabilidad penal de las empresas*, Cizur Menor, 2015.

Ferguson, G.: "Corruption and Corporate Criminal Liability", publicaciones del *International Centre for Criminal Law Reform and Criminal*

Justice Policy, Vancouver, May 1998, págs. 1 ss., en *{http://www.icclr.law.ubc.ca/Publications/Reports/}FergusonG.PDF.*

Freitas, P.M., Andrade, F. y Novais, P.: "Criminal Liability of Autonomous Agents: From the Unthinkable to the Plausible", en Casanova, P., Pagallo, U, Palmirani, M y Sartor, G. (eds.): *AI Approaches to the Complexity of Legal Systems*, AICOL 2013 International Workshops, Belo Horizonte, Brazil, July 21-27, 2013 and Bologna, Italy, December 11, 2013, Revised Selected Papers, Heidelberg, NewYork, Dordrecht, London, 2014, págs. 145 ss.

Future of Life Institute: *Pause Giant AI Experiments: An Open Letter*, March 23, 2023, en *{https://futureoflife.org/open-letter/pause-giant-ai-experiments/}.*

Gless, S., Silverman, E. and Weigend, T.: "If Robots Causa Harm, Who Is to Be Blame? Self-Driving Cars and Criminal Liability", en *New Criminal Review*, vol. 19, núm. 3, 2016, págs. 412 ss.

Gobert, J. y Punch, M.: *Rethinking Corporate Crime*, Bath, 2003

Gómez Colomer, J.L.: *El juez-robot (La independencia judicial en peligro)*, Valencia, 2023.

González Rus, J.J.: "Recensión al libro de Javier Valls Prieto, *Inteligencia artificial, Derecho humanos y bienes jurídicos*", en RECPC 24-r2, 2022, págs. 1 ss.

Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial creado por la Comisión Europea en junio de 2018: *Directrices Éticas para una IA Fiable*, Bruselas, 2019.

Hallevy, G.: "The Criminal Liability of Artificial Intelligence Entities – from Science Fiction to Legal Social Control", en *Akron Intellectual Property Journal*, vol. 4, 2010, págs. 171 ss.:

- *When Robots Kill. Artificial Intelligence Under Criminal Law*, Boston, 2013.
- *Liability for Crimes Involving Artificial Intelligence Systems*, Switzerland, 2015.
- "The Basic Models of Criminal Liability of AI Systems and Outer Circles", June 11, 2019, en *{https://ssrn.com/abstract=3402527}.*
- *Criminal liability for intellectual property offences of artificial intelligence entities*, London, 2020.

Hansen, A.L. y Kazinnik, S.: "Can ChatDecipher Fedspeak Lundgaard?", 2023, en *{https://ssrn.com/abstract=4399406}.*

Hassemer, W. Y Muñoz Conde, F.: *La responsabilidad por el producto en derecho penal*, Valencia, 1995.

Hilgendorf, E: "Relación de causalidad e imputación objetiva a través del ejemplo de la responsabilidad penal por el producto, en *ADPCP*, vol. LV, 2002, págs. 91 ss.

Jakobs, G.: *Derecho penal. Parte general. Fundamentos y teoría de la imputación*, trad.: J. Cuello Contreras y J.L. Serrano González de Murillo, 2ª ed. corregida, Madrid, 1997.

- "¿Punibilidad de las personas jurídicas?", en Montealegre Lynett, E. (coord..): *Libro Homenaje al profesor Günther Jakobs. El Funcionalismo en Derecho penal*, Bogotá, 2003, págs. 326 ss.

Koudela, P.: "Robots Instead of Immigrants: The Positive Feedback of Japanese Migration Policy on Social Isolation and Communication Problems", en *Asia-Pacific Social Science Review*, 19, 2019, pág. 91 ss.

Kuhlen, L.: "Necesidad y límites de la responsabilidad penal por el producto", en *ADPCP*, vol. LV, 2002, págs. 67 ss.

Law Commission: *Corporate Criminal Liability. An Options Paper*, 10 June 20.

Lopez-Lira, A. y Tang, Y.: "Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models", 2023, en *{https://ssrn.com/abstract=4412788}*.

Llonín Blasco, B.: "Acerca de la relación entre inteligencia artificial y responsabilidad penal empresarial", en *Revista Sistema Penal Crítico*, núm. 3, 2022, págs. 27 ss.

Maslej, N., Fattorini L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Ngo, H., Niebles, J.C., Parli, V., Shoham, Y., Wald, R., Clark, J. y Perrault, R.: "The AI Index 2023 Annual Report", *AI Index Steering Committee,* Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2023.

McCarthy, J., Marvin Minsky, M.L., Nathaniel Rochester, N. y Claude Shannon, C.E.: *A Proposal for the Darmouth Summer Research Project on Artificial Intelligence*, August 31, 1955.

Miró Llinares, F.: "Inteligencia artificial y justicia penal: Más allá de los resultados lesivos causados por robots", en *Revista de Derecho Penal y Criminología*, 3ª Época, núm. 20 (julio de 2018), págs. 87 ss.

- "El sistema penal ante la inteligencia artificial: actitudes, usos, retos", en Dupuy, D. y Corvalán, J.G. (dir.) y Kiefer, M. (coord..): *Cibercrimen III. Inteligencia Artificial. Automatización, algoritmos y predicciones en el Derecho penal y procesal penal*, Montevideo – Buenos Aires, 2020, págs. 81 ss.

Mizuta, T.: "Can an AI perform market manipulation at its own discretion? –A generic algorithm learns in an artificial market simulation–", en *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, December 1-4, 2020, Canberra, Australia, pág. 407 ss.

Muñoz Conde, F.: "La responsabilidad penal por el producto en el Derecho español", en *Derecho & Sociedad*, núm. 49, 2017, págs. 253 ss.

Navarro Michel, M.: "Vehículos automatizados y responsabilidad por el producto defectuoso", en *Revista de Derecho Civil*, vol. VII, núm. 5, octubre-noviembre 2020, págs. 175 ss.

Navas Navarro, S.: *Daños ocasionados por sistemas de inteligencia artificial: Especial atención a su futura regulación*, Granada, 2022.

Nemitz, P.: "Constitutional Democracy and Technology in the Age of Artificial Intelligence", en *Philosophical Transactions of Royal Society a Mathematical, Physical and Engineering Sciences*, 2018, págs. 1 ss. (en *{https://royalsocietypublishing.org/doi/epdf/10.1098/rsta.2018.0089}*).

Olarte Encabo, S.: "La aplicación de inteligencia artificial a los procesos de selección de personal y ofertas de empleo: Impacto sobre el derecho a la no discriminación", en *Documentación Laboral*, núm. 119, vol. I, 2020, págs. 78 ss.

Oliver, N.: "Una pausa cuestionable en la inteligencia artificial", en diario *El País*, edición del 3 de mayo de 2023.

Osmani, N.: "The Complexity of Criminal Liability Systems", en *Masaryk University Journal of Law and Technology*, vol. 14, 2020, págs. 53 ss.

Paredes Castañón, J.M.: "Capítulo 12: Responsabilidad penal por productos defectuosos", en Camacho, A. (dir.): *Tratado de Derecho Penal Económico*, Valencia, 2019, págs. 601 ss.

Puppe, I.: "Problemas de imputación del resultado en el ámbito de la responsabilidad penal por el producto", en Mir Puig, S. y Luzón Peña (coords.): *Responsabilidad penal de las empresas y sus órganos y responsabilidad por el producto*, Barcelona, 1996, págs. 215 ss.

Quintero Olivares, G.: "La robótica ante el Derecho penal: El vacío de respuesta jurídica a las desviaciones incontroladas", en *REEPS*, 1, 2017, págs. 1 ss.

Robles Planas, R: "Pena y persona jurídica: crítica del artículo 31 bis CP", en *Diario La Ley*, n.º 7.705, 29 de septiembre de 2011, pág. 2.

Scopino, G.: "Do automated trading systems dream of manipulating the price of futures contracts? Policing markets for improper trading practices by algorithmic robots", en *Florida Law Review*, vol. 67, 2016, págs. 221 ss.

Solum, L.B.: "Legal Personhood for Artificial Intelligences", en *North Carolina Law Review*, vol. 70, 1992, págs. 1231 ss.

Teubner, G.; "Digitale Rechtssubjekte? Zum privaterechttlichen Status autonomer Softwareagentem", en *Archiv Für die Civilistische Praxis*, 218, 2018, págs. 155 ss., consultado en *{https://www.jura.uni-frankfurt.de/69768539/TeubnerDigitale-RechtssubjekteAcP-18Dez17.pdf}*.

Turing, A.M.: "Computing Machinery and Intelligence", en *Mind*, vol. 49, 1950, págs. 433 ss.

Valls Prieto, J.: *Inteligencia artificial, derechos humanos y bienes jurídicos*, Cizur Menor, 2021.

Varela, L.: "Strict-Liability como forma de imputación jurídico-penal", en *InDret*, julio de 2012, págs. 1 ss.

Vogel, J.: "La responsabilidad penal por el producto en Alemania: Situación actual y perspectivas de futuro", en *Revista Penal*, 2001, págs. 95 ss.

Weismann, A. y Newman, D.: "Rethinking Criminal Corporate Liability", en *Indiana Law Journal*, vol. 82, 2007, pág. 411 ss.

Wells, C.: *Corporations and Criminal Responsibility*, 2.ª ed., New York, 2001.

White Paper: *Artificial Intelligence and Algorithmic Liability. A Technology and Risk Engineering from Zurich Insurance Group and Microsoft Corp.*, July, 2021, pág. 3.

Wootson, C.R.: "Saudi Arabia, Which Denies Women Equal Rights, Makes a Robot a Citizen", en *The Washington Post*, 29 de octubre de 2017.

Zhao, J.: "Artificial Intelligence and Corporate Decisiones: Fantasy, Realty or Destiny", en *Catholic University Law Review*, vol 71, núm. 4, 2022, págs. 663 ss.

Zurita Martín, I: *La responsabilidad por los daños causados por los robots inteligentes como productos defectuosos*, Madrid, 2020.