

CAPÍTULO V

Deshumanizando la culpabilidad: los sistemas inanimados en la teoría del delito

Angelo Giraldi

SUMARIO: 1. Consideraciones introductorias – 2. Aproximaciones a la vulnerabilidad del concepto de culpabilidad – 3. La incidencia de los sistemas de inteligencia artificial en la culpabilidad como presupuesto de la punibilidad – 3.1. El tentativo de reconocer una nueva subjetividad pasiva: la inteligencia artificial como víctima de delitos – 3.2. Hacia una subjetividad activa instrumental: la inteligencia artificial como autora de delitos – 3.3. La inteligencia artificial como instrumento de apoyo para el ejercicio de la justicia – 4. La incidencia de los sistemas de inteligencia artificial en la culpabilidad como criterio para la determinación de la pena en concreto – 5. Conclusiones – 6. Bibliografía.

1. *Consideraciones introductorias*

La estructura del Derecho penal contemporáneo se ha plasmado en los siglos a través de los reflejos sociales de la ciencia y de la técnica. No es ninguna novedad que la ciencia jurídica no pueda (o no deba) prescindir de los avances tecnológicos y sociales, a su vez influenciados por los acontecimientos históricos y el desarrollo de los demás ámbitos del conocimiento¹, pues, de hacerlo, se arriesgaría a que la regulación de toda conducta humana –y, a esta altura, también inhumana²– quedase sin

¹ El propio concepto de culpabilidad, así como las teorías de la pena, representan el resultado de reflexiones doctrinales elaboradas en el marco de determinadas realidades históricas y sociales. Al respecto, v. D. SANTAMARIA, voz *Colpevolezza*, en «Enciclopedia del diritto», VII, 1960, pp. 646 ss.; G. MARINI, voz *Colpevolezza*, en «Digesto delle Discipline Penalistiche», II, 1988, pp. 314 ss.; A. VALLINI, *Antiche e nuove tensioni tra colpevolezza e diritto penale artificiale*, Giappichelli, Torino, 2003, pp. 9 ss.; H.H. JESCHECK, *Evolución del concepto jurídico penal de culpabilidad en Alemania y Austria*, en «Revista Electrónica de Ciencia Penal y Criminología», 2003, pp. 2 ss.; J.M. NAVARRETE URIETA, *El nacimiento de la idea de culpabilidad*, en «Aldaba», 18/1992, pp. 31 ss.

² Con las precisiones debidas, que se detallarán *infra*. Hasta la fecha se utiliza el adjetivo 'artificial' en la locución 'inteligencia artificial' para subrayar que, aunque adquieren

efecto alguno. Con ello, sólo se pretende insistir en que una regulación que carece de efectividad refleja un estado anacrónico del ordenamiento jurídico que conlleva riesgos para la seguridad colectiva, objeto último de tutela por el Derecho, especialmente desde una perspectiva penal³. En otras palabras, lo que se pretende es proteger al individuo a través de un «margen de *seguridad cognitiva* frente a la posibilidad de ataques a sus intereses por parte de otros individuos»⁴.

No puede omitirse que las reflexiones que siguen están encaminadas al análisis de cuestiones futuras desde una perspectiva limitada, abordada desde la visión peculiar del Derecho penal. Como es lógico, el análisis ‘integrado’ de las cuestiones afectadas por la inteligencia artificial supone un conocimiento global de los asuntos científicos que, sin duda alguna, exceden del perímetro de estas consideraciones. Así que, sin alguna pretensión de exhaustividad, este escrito persigue a la vez un doble objetivo: por un lado, intentar averiguar el *an* y el *quomodo* de la influencia de los sistemas de inteligencia artificial en la teoría del delito y, en particular, en los elementos que –con fundadas razones de discrepancia en la doctrina⁵–

autonomía en el aprendizaje y en la orientación de los procesos causales, estos sistemas son el fruto de la creación por parte del ser humano.

³ En las palabras de G. JAKOBS, *El principio de culpabilidad*, en «Anuario de Derecho Penal y Ciencias Penales», 3/1992, p. 1074, la finalidad de prevención general positiva de la pena consta de «prevención general, porque pretende dirigirse un efecto en *todos* los ciudadanos; [y se denomina] positiva, porque este efecto no se pretende en que consista en miedo ante la pena, sino en una tranquilización en el sentido de que la norma esta vigente, de que la vigencia de la norma, que se ha visto afectada por el hecho, ha vuelto a ser fortalecida por la pena». Al respecto, v. también C. ROXIN, *Derecho Penal. Parte General. Tomo I. Fundamentos. La estructura de la teoría del delito*, Civitas, Madrid, 1997, p. 97, traducción de D.M. Luzón Peña, M. Díaz y García Conlledo, J. de Vicente Remesal: «[e]n primer lugar, el fin de la conminación penal es de pura prevención general».

⁴ R. ALCÁCER GUIRAO, *Los fines del Derecho Penal. Una aproximación desde la filosofía política*, en «Anuario de Derecho Penal y Ciencias Penales», 1/1998, p. 425: «En palabras de Vehling, el Derecho sirve al fin de “allanar los momentos de inseguridad respecto a la conducta de los otros en los contactos sociales a través de un mínimo de orientación jurídica”». No debe extrañar que la regulación de las conductas humanas, y en particular la represión de algunas de estas a través del Derecho penal, estén (y tiendan a modificarse para estar) ancladas a lo que la razón humana percibe en un determinado momento histórico. Con las palabras *ante litteram* de B. SPINOZA, *Tratado político*, Alianza Editorial, Madrid, 1986, p. 96, traducción de Atilano Domínguez: «no resulta tan inadecuado que los hombres que están habituados a vivir en el Estado llamen pecado a lo que contradice el dictamen de la razón, puesto que los derechos del mejor Estado [...] deben estar fundados en ese dictamen».

⁵ Sobre la evolución del concepto de culpabilidad, v. *amplius* Á. TORIO LÓPEZ, *El concepto individual de culpabilidad*, en «Anuario de Derecho Penal y Ciencias Penales», 2/1985, pp. 285 ss.

conformarían el núcleo de la llamada culpabilidad; por otro, hacer hincapié en la influencia y en los cambios que la introducción masiva de sistemas inanimados autónomos en la realidad social tendrá en el Derecho penal.

Tradicionalmente, cuando se empezó a hablar de inteligencia artificial, en términos muy amplios y cada vez más acogedores, se hizo especial referencia al ámbito de la privacidad y de los datos personales. La masificación de los datos facilitados o introducidos en programas informáticos, bases de datos, aplicaciones y servidores representaba un peligro para la intimidad personal, protegida en España por el propio art. 18 de la Constitución⁶ y, como reflejo, por la Ley penal. No es casualidad que, hasta el día de hoy, los órganos jurisdiccionales se vean implicados en el pronunciamiento de resoluciones que de alguna manera limiten el empleo de la inteligencia artificial⁷.

A título ejemplificativo, cabe mencionar la reciente sentencia de 21 de junio de 2022, dictada por el Tribunal de Justicia de la Unión Europea en al asunto C-817/19. El asunto está relacionado con la Directiva *Passenger Name Record* (PNR)⁸ y su incorporación al ordenamiento interno de Bélgica. Dicha Directiva impone el tratamiento sistemático de un número considerable de datos de los pasajeros aéreos de los vuelos que entran y salen de la Unión Europea, y excepcionalmente también de los vuelos ‘intra-UE’,

⁶ No se prevé análoga expresa previsión normativa en la Constitución italiana, esencialmente por razones históricas. Antes de la promulgación del Convenio Europeo de Derechos Humanos en 1950, la Constitución italiana tutelaba la intimidad personal desde una perspectiva entonces estática, principalmente obstativa, que se desprende de sus arts. 14 («El domicilio es inviolable»), 15 («La libertad y el secreto de la correspondencia y de cualquier otra forma de comunicación son inviolables») y 21 («Todos tienen derecho a manifestar libremente su pensamiento con la palabra, el escrito y cualquier otra forma de difusión»). Con la entrada en vigor del Convenio Europeo de Derechos Humanos, entre los derechos fundamentales mencionados en el art. 2 de la Constitución italiana se asoma el contenido del art. 8 CEDH: «1. Toda persona tiene derecho al respeto de su vida privada y familiar, de su domicilio y de su correspondencia. 2. No podrá haber injerencia de la autoridad pública en el ejercicio de este derecho sino en tanto en cuanto esta injerencia esté prevista por la ley y constituya una medida que, en una sociedad democrática, sea necesaria para la seguridad nacional, la seguridad pública, el bienestar económico del país, la defensa del orden y la prevención de las infracciones penales, la protección de la salud o de la moral, o la protección de los derechos y las libertades de los demás».

⁷ Por razones metodológicas y de espacio, no procede examinar en esta sede el panorama de las resoluciones judiciales que mencionan, directa o indirectamente, la inteligencia artificial. Para un *excursus* de algunas resoluciones, v. D.L. MORILLAS FERNÁNDEZ, *Implicaciones de la inteligencia artificial en el ámbito del Derecho Penal*, en esta obra, § 3.

⁸ Directiva (UE) 2016/681 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativa a la utilización de datos del registro de nombres de los pasajeros (PNR) para la prevención, detección, investigación y enjuiciamiento de los delitos de terrorismo y de la delincuencia grave.

con el fin de luchar contra el terrorismo y demás delitos considerados ‘graves’. No es necesario detenerse sobre el asunto específico, pero resulta curioso mencionar la referencia (quizás anacrónica) que el Tribunal de Justicia hizo a los sistemas de inteligencia artificial.

Con arreglo a la Directiva, la Unidad de Información sobre Pasajeros (UIP) de cada Estado miembro, para realizar la evaluación de los pasajeros antes de su llegada o salida programada del Estado, puede tratar los *passenger name record* exclusivamente con arreglo a criterios predeterminados⁹. Sobre la base de esta disposición, en la resolución del Tribunal de Justicia de la Unión Europea se establece que la evaluación con arreglo a criterios predeterminados impide la utilización de tecnologías de inteligencia artificial en el ámbito de sistemas de *machine learning*, ya que estos serían capaces de modificar el proceso de evaluación sin que intervenga el control humano¹⁰. En particular, sería capaz de alterar los criterios de evaluación que, por expresa previsión normativa, han de ser predeterminados. Con otras palabras, el Tribunal hace hincapié en la opacidad del funcionamiento de los mecanismos de inteligencia artificial, cuya comprensión de cara a averiguar el alcance de un determinado resultado puede resultar hasta imposible¹¹.

⁹ Art. 6, apartado 3, letra b), de la Directiva (UE) 2016/681.

¹⁰ El § 194 de la resolución de 21 de junio de 2022 establece: «Con respecto a los criterios que el UIP puede utilizar para dicha finalidad, hay que destacar ante todo que, según el propio texto del artículo 6, párrafo 3, letra b), de la Directiva PNR, dichos criterios deben ser “preestablecidos”. Como se ha puesto de relieve por parte del abogado general en el párrafo 228 de sus conclusiones, este requisito impide la utilización de tecnologías de inteligencia artificial en el ámbito de sistemas de autoaprendizaje (*machine learning*), que pueden modificar, sin intervención ni control humanos, el proceso de evaluación y, en particular, los criterios de evaluación sobre los cuales se basa el resultado de la aplicación de dicho proceso, además de la ponderación de tales criterios» (traducción de la versión italiana realizada por el autor, al no existir traducción oficial al español).

¹¹ El Tribunal acoge las observaciones formuladas por el Abogado General en el párrafo 228: «En tercer lugar, tanto del tenor del artículo 6, apartado 3, letra b), de la Directiva PNR como del sistema de garantías que acompaña al tratamiento automatizado de los datos PNR establecido por la Directiva PNR se desprende que el funcionamiento de los algoritmos utilizados para realizar el análisis al que se refiere esta disposición debe ser transparente y permitir la trazabilidad del resultado de su aplicación. Evidentemente, este requisito de transparencia no implica que deban hacerse públicos los “perfiles” utilizados. En cambio, requiere que se garantice que pueda identificarse la toma de decisión algorítmica. En efecto, por una parte, el requisito de que los criterios con arreglo a los cuales debe realizarse este análisis sean “predeterminados” excluye la posibilidad de que puedan modificarse sin intervención humana y se opone, en consecuencia, a la utilización de las tecnologías de inteligencia artificial llamadas “machine learning”, las cuales, al mismo tiempo que pueden presentar un grado más elevado de precisión, son difíciles de interpretar, incluso por parte de los operadores que han efectuado el tratamiento automatizado. [...] La transparencia del funcionamiento de los algoritmos utilizados constituye también un requisito necesario para

En cierta medida, y esta es la razón por la cual se ha mencionado dicha resolución, sorprende la censura de la utilización de sistemas inteligentes inanimados en un sector que sólo afecta a la intimidad y la privacidad de los datos de los pasajeros aéreos, al mismo tiempo que las demás Instituciones europeas están demandando (o al menos autorizando) la implementación regulada de estos avances tecnológicos especialmente en el ámbito de la justicia penal.

Cabe recordar, por cierto, que en algunos Países de la Unión Europea ya se encuentran implementados mecanismos de decisión automatizada en procedimientos judiciales. Sin entrar en detalles, en el Reino Unido (entonces parte de la Unión), Francia y Holanda se han implementado modelos de algoritmos en el ámbito de la justicia civil¹², que posiblemente trascienden el alcance de la privacidad relativa a los datos de los pasajeros aéreos. Asimismo, en España se ha introducido la utilización de *VioGén*, una técnica de policía predictiva encaminada a la valoración policial del riesgo¹³, y de *VeriPol*, el polígrafo utilizado por las comisarías para detectar denuncias falsas¹⁴.

De todos modos, la rápida evolución del estado de la técnica ha intervenido en favor de una reconsideración de la inteligencia artificial y de sus efectos, de manera mucho más amplia y dinámica que la exclusivamente ligada a la intimidad personal. Por dicha razón, se ha ido introduciendo prolíficamente un *corpus* europeo de recomendaciones y normas especializado en la materia¹⁵, con el objetivo de fijar algunos límites

que los interesados puedan ejercer sus derechos de reclamación y su derecho a un recurso jurisdiccional efectivo».

¹² Cfr. L. NOTARO, *Intelligenza artificiale e giustizia penale*, en *Intelligenza artificiale e giustizia penale*, dirigido por A. Massaro, Paruzzo Editore, Caltanissetta, 2020, pp. 102 ss.; M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, en «Diritto penale contemporaneo», 29 de mayo de 2019, pp. 10 ss.

¹³ Permítase la referencia a A. GIRALDI, *Algorithms and Big Data Towards a Crime-Preventing Groupware*, en «Roma Tre Law Review», 2/2021, pp. 26 ss.

¹⁴ El polígrafo está siendo utilizado desde 2018. Sobre «la herramienta que más dudas nos genera», v. C. ALONSO SALGADO, *Acerca de la inteligencia artificial en el ámbito penal: especial referencia a la actividad de las fuerzas y cuerpos de seguridad*, en «Ius et Scientia», 1/2021, pp. 32-34.

¹⁵ Cfr. E. GAMERO CASADO, *El enfoque europeo de inteligencia artificial*, en «Revista de Derecho Administrativo», 2021, pp. 268 ss. Más recientemente, con particular referencia al ámbito de la justicia penal, se aprobó la Resolución del Parlamento Europeo, de 6 de octubre de 2021, sobre la inteligencia artificial en el Derecho penal y su utilización por las autoridades policiales y judiciales en asuntos penales. A este respecto, v. MORILLAS FERNÁNDEZ, *Implicaciones de la inteligencia artificial*, cit., § 1. También cabe mencionar, desde una perspectiva más amplia, la Resolución del Parlamento Europeo, de 3 de mayo de 2022, sobre la inteligencia artificial en la era digital, en cuya sección «La IA y el futuro de la democracia» señala algunas criticidades de naturaleza jurídica: «estas tecnologías de IA plantean cuestiones éticas y jurídicas cruciales;

a un mundo –el tecnológico– que de por sí no conocería restricciones.

2. *Aproximaciones a la vulnerabilidad del concepto de culpabilidad*

Como se adelantaba al principio, el alcance de los sistemas de inteligencia artificial se abordará en estas consideraciones desde la perspectiva peculiar del Derecho penal y, en particular, de aquel conjunto de elementos que, según la doctrina que se acoja, se denomina culpabilidad. De ahí, resulta imprescindible efectuar algunas premisas de carácter metodológico.

Los fundamentos de la culpabilidad –que encuentra su acogida en el Derecho penal español¹⁶ y, según la mayoría de la doctrina actual¹⁷, en el

señala que determinadas tecnologías de IA permiten automatizar el tratamiento de la información a una escala sin precedentes, lo que allana el camino para la vigilancia colectiva y la injerencia de manera ilegal y *representa una amenaza para los derechos fundamentales* [énfasis añadido], en especial el derecho a la privacidad y la protección de datos»; además recuerda esta resolución que «muchos regímenes autoritarios utilizan sistemas de IA para controlar, espiar, seguir y clasificar a sus ciudadanos, restringir su libertad de circulación y ejercer vigilancia colectiva; hace hincapié en que cualquier forma de marcaje normativo ciudadano por parte de las autoridades públicas, especialmente en el ámbito de las fuerzas de seguridad del Estado, el control de las fronteras y el poder judicial, así como su uso por parte de empresas privadas o particulares, conduce a la pérdida de autonomía y privacidad y no está en consonancia con los valores europeos; recuerda que tecnologías como la cibervigilancia y el reconocimiento biométrico, que pueden utilizarse para estos fines, están sujetas al Reglamento sobre el control de las exportaciones de la Unión; muestra su profunda preocupación por los casos anteriores de empresas de la Unión que han vendido a regímenes autoritarios de países no pertenecientes a la Unión sistemas biométricos cuyo uso sería ilegal en la Unión y condena estos casos».

¹⁶ L. MORILLAS CUEVA, *Sistema de Derecho Penal. Parte general*, Dykinson, Madrid, 2018, pp. 597 ss.; M. COBO DEL ROSAL, T.S. VIVES ANTÓN, *Derecho penal. Parte general*, Tirant lo Blanch, Valencia, 1999, p. 549; F. MUÑOZ CONDE, M. GARCÍA ARÁN, *Derecho Penal. Parte General*, Tirant lo Blanch, Valencia, 2019, pp. 341 ss. Con diferentes matices, varios autores le atribuyen valor en sí, prescindiendo de su función generalpreventiva. Cfr. S. MIR PUIG, *Función de la pena y teoría del delito en el Estado social y democrático de Derecho*, Bosch, Barcelona, 1982, pp. 91 ss.; M. PÉREZ MANZANO, *Culpabilidad y prevención. Las teorías de la prevención general positiva en la fundamentación de la imputación subjetiva y de la pena*, Universidad Autónoma de Madrid, Madrid, 1990, p. 213: «El argumento conforme al cual la culpabilidad debe ser mantenida porque tiene efectos general integradores, no puede ser esgrimido como fundamental, pues entonces la culpabilidad carece de valor en sí y el criterio determinante son las mudables exigencias sociales. La necesidad de mantener la culpabilidad es consecuencia de un argumento valorativo derivado de los principios del Estado de Derecho».

¹⁷ G. FIANDACA, E. MUSCO, *Diritto penale. Parte generale*, Zanichelli, Bolonia, 2019, pp. 329 ss.; G. MARINUCCI, E. DOLCINI, G.L. GATTA, *Manuale di Diritto Penale. Parte Generale*, Giuffrè, Milán, 2018, pp. 335 ss.

italiano— han ido evolucionando y se orientan hoy hacia la superación del modelo *punitur quia peccatum*. Desde el Derecho penal contemporáneo, constitucionalmente orientado, se ha ido huyendo de las teorías retributivas, en aras de la afflictividad del proceso sancionador y de su teleología reeducativa y resocializadora¹⁸. Con base en estas evoluciones del concepto de culpabilidad, y de su trascendencia en el sistema jurídico-penal vigente, es evidente que la introducción de sujetos inanimados en el entorno social tiene varias repercusiones en los elementos que lo conforman. Dicho de otro modo, desde la perspectiva psicológica o desde la más avanzada concepción normativa, en un mundo dominado (o transformado) por la actuación de sistemas de inteligencia artificial, el Derecho penal puede ver amenazados sus principios rectores.

Con el fin de analizar cuáles pueden ser las implicaciones de un cambio radical que supone la llegada de objetos inanimados en la esfera de los sujetos de derecho, se intentará esbozar un análisis de la culpabilidad penal distinguiendo entre sus dos vertientes. Por un lado, se examinará la influencia de los sistemas de inteligencia artificial en la culpabilidad en su condición de presupuesto de la punibilidad¹⁹, y por otro, se analizará la cara más empírica de la moneda, es decir, la culpabilidad como criterio de determinación de la sanción penal.

Antes de ceñir el enfoque de estos temas, parece oportuno acercarse al concepto de inteligencia artificial y de los correspondientes sistemas. En efecto, pese a que la expresión se halle incorporada al lenguaje común, se hace necesario aclarar su significado con vistas a estas reflexiones. El concepto de inteligencia artificial ha sido definido como el potencial biopsicológico de procesar informaciones para solventar problemas o crear productos que tienen valor en una determinada cultura social²⁰ y, asimismo, como el proceso de llevar una máquina a comportarse de una manera que

¹⁸ Cfr. los artículos 25.2 de la Constitución española («Las penas privativas de libertad y las medidas de seguridad estarán orientadas hacia la reeducación y reinserción social y no podrán consistir en trabajos forzados») y 27, apartado tercero, de la Constitución italiana («Las penas no pueden consistir en tratamientos contrarios al sentido de humanidad y deben tender a la reeducación del condenado»). Se llega a dudar que pueda defenderse la retribución como «desvinculada del fin». Así, W. HASSEMER, *¿Por qué castigar? Razones por las que merece la pena la pena*, Tirant lo Blanch, Valencia, 2016, p. 62, traducción de M. Cancio Meliá y F. Muñoz Conde.

¹⁹ Es decir, la culpabilidad como elemento constitutivo del delito, pese al escenario dogmático que se acoja con respecto a la estructura del delito.

²⁰ H. GARDNER, *Intelligence Reframed. Multiple Intelligences for the 21st Century*, Basic Books, New York, 2000, p. 33: «biopsychological potential to process information [...] to solve problems or create products that are of value in a culture».

se consideraría inteligente si un humano actuara así²¹.

No debe extrañar, pues, que al tratarse de un mundo en continua evolución, ya hayan aparecido distintas etapas en el desarrollo de dichos sistemas. En cambio, lo que sí debería impresionar –y quizás no lo hace por el justificable recelo hacia el Derecho positivo– es el consuetudinario aletargamiento del ordenamiento jurídico, que mientras sigue ocupándose de –y asolando la ley ante– presuntos incrementos de ‘alarma social’, aparta la adecuación de la legislación vigente a los retos tecnológicos que avanzan, y que ya han dejado atrás las primeras etapas de su desarrollo. Evidentemente, la primera generación de los sistemas de inteligencia artificial (*artificial narrow intelligence*, que permitió desde la creación de los asistentes de voz hasta el diseño de los primeros coches autónomos) está ahora al alcance de cualquiera y ha sido ya superada por la *artificial general intelligence*, la segunda generación capaz de razonar, plantear y solventar problemas de forma independiente e incluso inesperada (*machine learning*) por sus diseñadores. Está previsto para el futuro próximo el desarrollo de una tercera generación (*artificial super intelligence*), que algunos llaman la «verdadera inteligencia artificial», formada por sistemas conscientes y capacitados para relegar a un segundo plano al ser humano²², lo que asustaría, desde el punto de vista jurídico, al poder imaginar hasta la posibilidad de que alcance al mismo procedimiento de creación y aprobación de las leyes.

En cualquier caso, el concepto de sistema de inteligencia artificial que se utilizará a lo largo de este escrito incluye todo sistema inanimado, creado por el ser humano y capaz de alimentarse por sí solo, susceptible de causar autónomamente modificaciones del mundo exterior, relevantes para la sociedad y, por consiguiente, para el Derecho. En dicha definición, lo más amplia posible, caben máquinas, robots, algoritmos, y cualquier otro tipo de sujeto inhumano, hasta el robot aspirador de todos los días.

Regresando a las consideraciones sobre la culpabilidad penal, conviene distinguir el alcance de los sistemas de inteligencia artificial según impacten

²¹ La definición «making a machine behave in ways that would be called intelligent if a human were so behaving» se remonta a la Conferencia de Dartmouth de 1955, en la que probablemente tuvo lugar la primera aparición de la locución ‘inteligencia artificial’. Cfr. J. MCCARTHY, M.L. MINSKY, N. ROCHESTER, C.E. SHANNON, *A proposal for the Dartmouth summer research project on artificial intelligence*, 31 de agosto de 1955, disponible en <<http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>> (consultado el 18 de noviembre de 2022).

²² La delimitación de las tres etapas de desarrollo de la inteligencia artificial se encuentra en A. KAPLAN, M. HAENLEIN, *Siri, Siri, in my hand: Who’s the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence*, en «Business Horizons», 1/2019, p. 16.

en el elemento constitutivo del delito o en el criterio para la determinación de la pena en concreto.

Con respecto al primero de los casos, es decir, a la culpabilidad como presupuesto de la punibilidad, y entonces, como elemento del delito, el análisis estará encaminado a averiguar las implicaciones que conllevan los nuevos sujetos de derecho, dependiendo de que sean autores o víctimas de la comisión de un hecho abstractamente delictivo, o bien instrumentos de apoyo al ejercicio de la justicia. Es posible adelantar desde ahora que, sin perjuicio de las investigaciones desarrolladas en sentido contrario, todavía resulta difícil imaginar una concepción de culpabilidad que tenga carácter unitario, es decir, que permita armonizar en una categoría autónoma el conjunto de los distintos elementos que la concepción normativa moderna deposita en la propia culpabilidad²³. Evidentemente, a la luz de la identidad del efecto que producen las causas tipificadas, lejos de insistir en categorías *praeter legem* como la exigibilidad de una conducta distinta²⁴, parecería correcto descartar la idea de que, en el marco de la misma norma penal, existan causas que eximen de la responsabilidad penal sin llegar a justificar el hecho²⁵.

En segundo lugar, con respecto a la culpabilidad concebida como criterio para la determinación de la pena, es necesario advertir que el análisis tomará en consideración la medición de la pena en concreto, es decir, la que debe imponer el órgano judicial posteriormente a la averiguación de la existencia del hecho delictivo. Con ello, sólo se intenta restringir el examen

²³ M. GALLO, *Il concetto unitario di colpevolezza*, Giuffrè, Milán, 1951, pp. 95-96: «intesa come fatto giuridico, e precisamente come criterio di imputazione, la colpevolezza sfugge ad ogni concettualizzazione che non sia quella meramente estrinseca della specifica funzionalità assoluta dagli elementi dai quali si afferma costituirsi». Para el autor, el contenido de la culpabilidad coincide con el concepto de antijuridicidad, o más exactamente, se enmarca en el seno de esta, constituyendo una forma especialmente cualificada de antijuridicidad. Cfr. *ivi*, p. 144.

²⁴ Sucintamente, «la referencia a la inexigibilidad de una conducta conforme a la norma como fundamento de las denominadas causas de inculpabilidad o exculpación, no casa bien con el hecho de que al declarar el legislador una conducta como antijurídica, sí parece exigirla». Así, T. AGUADO CORREA, *Principio de inexigibilidad de otra conducta en las categorías del delito*, en «Revista Digital de la Maestría en Ciencias Penales de la Universidad de Costa Rica», 3/2011, p. 24.

²⁵ GALLO, *La teoria dell'azione «finalistica» nella più recente dottrina tedesca*, Giuffrè, Milán, 1950, p. 60. Cabe recordar que, en la perspectiva bipartita que apoya el autor, el hecho puede definirse típico al constar de la conducta, el nexo causal y el resultado, además de la ausencia de causas de justificación, que se encajan en el tipo como elementos negativos construidos positivamente.

al efecto último²⁶ del principio de culpabilidad, ya que, hasta la fecha, la determinación de la pena en abstracto le corresponde íntegramente al legislador, que actúa (o debería actuar) atendiendo a criterios jurídicos y a juicios de valor de carácter político-criminal. Se ha aludido antes a la posibilidad de que los sistemas de inteligencia artificial lleguen incluso a intervenir en el *iter legis*. Sin embargo, parece que de momento el control de dicho proceso queda exclusivamente en manos del legislador y no ha habido propuestas de incorporación de mecanismos artificiales en una actividad tan crucial, cuya justificación sigue rigiéndose por la población en aras de la insoslayable reserva de ley penal.

En este sentido, pese a los cambios de los marcos edictales realizados por los distintos mandatos parlamentarios, lo importante es que se respeten los principios rectores en atención al peligro o al daño causado al bien jurídico protegido por la norma penal. Ahora bien, en el ámbito de la determinación de la pena en concreto, el alcance de la inteligencia artificial se estudiará según se indica: por un lado, la inteligencia artificial pondría en peligro la operatividad de la proporcionalidad de la pena, al prever mecanismos que impedirían la explotación de todo segmento del marco edictal; por otro lado, puede darse el caso de la máquina que, al aplicar la pena en lugar del juez persona física, genera una condena injusta o ilícita, cuyo pronunciamiento puede adquirir a su vez relevancia penal.

3. La incidencia de los sistemas de inteligencia artificial en la culpabilidad como presupuesto de la punibilidad

Con respecto al conjunto de elementos que conformaría la categoría autónoma de la culpabilidad, ya se ha adelantado la existencia de discrepancias entre la doctrina contemporánea. En principio, varios han sido los esfuerzos para dotarla de independencia dogmática; tanto es así

²⁶ En realidad, la determinación judicial de la pena no sería el último de los efectos jurídicamente relevantes de la culpabilidad, al jugar sus elementos (especialmente los que están vinculados con las circunstancias personales del reo) otro papel en la fase de post-determinación de la pena. En cierta medida, es fundamental que el reo continúe a percibir justo, equitativo, y entonces proporcionado, el tratamiento sancionador que se le otorgó en su momento. Al respecto, véase la literatura citada en GIRALDI, *Política, discrecionalidad y Derecho en las implicaciones empíricas del principio de proporcionalidad de la pena*, en «Revista Internacional de Doctrina y Jurisprudencia», 22/2020, pp. 23 ss.

que la doctrina predominante, en una perspectiva pluripartita²⁷, incluye en la culpabilidad la imputabilidad, la conciencia de la antijuridicidad de la conducta realizada y la exigibilidad de una actuación distinta por parte del sujeto activo²⁸. Sin embargo, persiste la disyuntiva sobre la inclusión de las formas de culpabilidad (dolo, imprudencia y, según algunos autores, dolo mixto a imprudencia²⁹) en su propio ámbito³⁰ o en el tipo subjetivo

²⁷ La autonomía de la categoría dogmática se afirma con distintos matices desde ambas las teorías tripartitas y cuatripartitas del delito, actualmente mayoritarias.

²⁸ MUÑOZ CONDE, GARCÍA ARÁN, *Derecho Penal*, cit., pp. 341-342; COBO DEL ROSAL, VIVES ANTÓN, *Derecho penal*, cit., p. 567, incluye en la estructura del juicio de culpabilidad la imputabilidad, las formas de culpabilidad (dolo y culpa) y la exigibilidad; MIR PUIG, *Derecho penal. Parte general*, Editorial Reppertor, Barcelona, 2011, p. 139, compartiendo las dudas sobre la ambigüedad del término y las connotaciones moralizantes de la culpabilidad, más bien se refiere a ella incluyendo en la esfera de la imputación personal «tanto la indicada exigencia de infracción personal de la norma primaria, como la necesidad de un sujeto penalmente responsable».

²⁹ La teoría del dolo mixto a culpa fundamentaría la regulación expresa, en el ordenamiento italiano, de la preterintencionalidad y del homicidio preterintencional (arts. 43 y 584 del Código Penal italiano), sobre la base de que las conductas lesivas serían dolosas y la causación del resultado mortal se imputaría a título de imprudencia. *Contra*, la tesis de la unicidad del elemento psicológico de la preterintencionalidad, actualmente dominante en la jurisprudencia italiana, según la cual el elemento subjetivo del delito subsidiario (lesiones) absorbería la previsibilidad del resultado causado (muerte). Cfr., *amplius*, F. BASILE, *L'alternativa tra responsabilità oggettiva e colpa in attività illecita per l'imputazione della conseguenza ulteriore non voluta, alla luce della sentenza Ronci delle Sezioni Unite sull'art. 586 c.p.*, en «Rivista italiana di diritto e procedura penale», 3/2011, pp. 911 ss. En España, tras la derogación de la atenuante de preterintencionalidad y de los delitos cualificados por el resultado, con el fin de erradicar los supuestos de *versari in re illicita*, la mayoría de la jurisprudencia está de acuerdo en aplicar el concurso ideal del art. 77 CP entre las lesiones dolosas y el homicidio imprudente (cfr. recientemente STS 2833/2022, de 7 de julio). Dicha solución concursal ya había sido propuesta por MIR PUIG, *Preterintencionalidad y límites del art. 50 del Código Penal*, en «Revista Jurídica de Cataluña», 1979, p. 99. V. también G. QUINTERO OLIVARES, *Las vicisitudes del dolo y la subsistencia de la preterintencionalidad*, en *Constitución, derechos fundamentales y sistema penal. Semblanzas y estudios con motivo del setenta aniversario del profesor Tomás Salvador Vives Antón*, dirigido por J.C. Carbonell Mateu, J.L. González Cussac, E. Orts Berenguer, Tirant lo Blanch, Valencia, 2009, p. 1573 ss.

³⁰ COBO DEL ROSAL, VIVES ANTÓN, *Derecho penal*, cit., pp. 616-617. Aunque un determinado momento del ánimo pueda considerarse para determinar si concurre el injusto típico y, a la vez, para decidir sobre la reprochabilidad de dicho injusto, «en ningún caso, el dolo o la imprudencia, como tales, puedan pertenecer al injusto típico. [...] El dolo, aunque presuponga ciertas realidades psicológicas, no se identifica, desde luego, con ellas. Y no puede pertenecer al tipo de injusto desde el momento en que representa un modo o forma en que la realización del tipo puede reprocharse a su autor».

(*subjektive Tatbestand*)³¹, o bien en ambos con distintas funciones³².

En cualquier caso, prescindiendo de la validez de las observaciones formuladas acerca de la imposibilidad de construir una concepción unitaria de culpabilidad³³, al entrar en la esfera de las relaciones jurídicas un conjunto de sujetos inanimados, vuelven a adquirir pleno auge las polémicas sobre la cercanía entre culpabilidad y peligrosidad³⁴, y quizás resultará oportuno cuestionar la validez de la estructura actual del delito.

El análisis de la influencia de los sistemas de inteligencia artificial en la culpabilidad, entendida como elemento constitutivo del delito y, por ende, presupuesto de la punibilidad, se desarrollará con arreglo a las distintas formas en que la existencia de estos sujetos puede interactuar con las conductas delictivas.

En primer lugar, se plantearán las cuestiones relativas a la eventual atribución de subjetividad *pasiva* a los sistemas ‘inanimados’, que alcanzarían la condición de potenciales víctimas de un delito, con todos los problemas de índole procesal que tal reconocimiento pueda entrañar. En este sentido, también hay que pensar en la dificultad de determinar quién –y sobre todo cómo– asumiría su representación y defensa en el juicio, teniendo presente

³¹ MUÑOZ CONDE, GARCÍA ARÁN, *Derecho Penal*, cit., p. 249. «Volver a incluir estos elementos en la culpabilidad me parece una complicación innecesaria y una falta antisistemática. Y lo mismo creo, si se considera que el dolo pertenece a la culpabilidad –como hace la teoría tradicional–. Lo que no se puede hacer es querer incluir un mismo hecho en dos categorías sistemáticas distintas, pues entonces ¿para qué sirven las clasificaciones y distinciones sistemáticas?». Así, MUÑOZ CONDE, *Introducción*, en C. ROXIN, *Política criminal y sistema del derecho penal*, Hammurabi, Buenos Aires, 2002, p. 28, traducción de F. Muñoz Conde.

³² Las disyuntivas sobre la ubicación del elemento psicológico del delito ponen de manifiesto la «riqueza de pensamiento jurídico[que] crece en la medida en que aumentan las soluciones que se ofrecen». Así, QUINTERO OLIVARES, *Parte General del Derecho Penal. Adaptada al programa de ingreso en las carreras judicial y fiscal*, Aranzadi, Pamplona, 2015, p. 161. Véanse, al respecto, las posiciones que apoyan la ubicuidad del dolo en MORILLAS CUEVA, *Sistema de Derecho Penal*, cit., pp. 671 ss.: «como factor directivo del comportamiento es el núcleo central del injusto típico de la acción; como resultado final del proceso de formación de la voluntad que pertenece a la culpabilidad es también un integrante de esta». En la misma línea, ya se expresó ROXIN, *Derecho Penal*, cit., p. 228, al afirmar que «la tan discutida cuestión de si el dolo “pertenece” al injusto o a la culpabilidad [...] es para el sistema político criminal una alternativa planteada erróneamente, ya que el dolo puede ser significativo desde ambos puntos de vista en contextos jurídicos respectivamente diversos».

³³ Cfr. *supra*, *sub* nota 23.

³⁴ Piénsese en que el dolo y la imprudencia, en el marco de las teorías que los incorporan como elementos de la culpabilidad, fundamentarían tanto el reproche penal como el juicio de peligrosidad criminal.

que la norma penal «real» consta de segmentos tanto sustantivos como procesales, ambos irrenunciables³⁵.

En segundo lugar, a la inversa, se analizarán las cuestiones más controvertidas que lleva consigo la atribución de subjetividad *activa* a los sistemas de inteligencia artificial. En esta perspectiva, merece la pena cuestionar la oportunidad de reprochar a una máquina, un robot o incluso un algoritmo. De considerarlo oportuno, habrá que analizar las posibles medidas de adaptación del actual sistema penal a efectos del reconocimiento de autoría directa por parte de los sujetos de inteligencia artificial.

Por último, la utilización de los sistemas de inteligencia artificial como *instrumento* de apoyo en el ámbito de la justicia exige el desarrollo de algunas reflexiones sobre la posible afectación del criterio de libre valoración de la prueba y, de cara al futuro, sobre el posible fomento de una revolución (patológica) en la teoría del delito.

3.1. *El tentativo de reconocer una nueva subjetividad pasiva: la inteligencia artificial como víctima de delitos*

El reconocimiento de subjetividad pasiva a los sistemas de inteligencia artificial, entendidos con los matices susodichos, permite el planteamiento de dos distintas cuestiones que merecen ser analizadas por separado. Por un lado, hay que subrayar las dificultades ligadas a la vigencia de un Derecho penal necesariamente lesivo, en el que – pese al marco teórico elaborado por la doctrina del delito como *personales Unrecht*³⁶ – las exigencias de protección de un determinado bien jurídico han de primar sobre el desvalor personal de la acción. Por otro lado, la convergencia entre los conceptos de culpabilidad y peligrosidad hacen resurgir viejas preocupaciones a la hora de examinar los casos en que los sistemas de inteligencia artificial pudieran ser víctimas.

En primer lugar, debe advertirse que la doctrina penalista se ha interrogado extensamente sobre la oportunidad de introducir nuevos tipos

³⁵ GALLO, *Diritto penale italiano. Appunti di parte generale*, vol. I, Giappichelli, Torino, 2021, p. 15: «la norma penale reale, la proposizione cioè che esaurisce il significato che ad una determinata condotta umana è attribuito in quel settore dell'ordinamento definito penale e criminale, consta di segmenti che delineano la fattispecie condizionante e le conseguenze condizionate e di segmenti che dispongono come deve essere accertata detta fattispecie e come vanno pronunciate (irrogate) ed eseguite le conseguenze sanzionatorie».

³⁶ Sobre la dificultad de concebir un injusto no culpable desde la doctrina alemana del injusto personal, v. E.J. LAMPE, *Das personale Unrecht*, Duncker & Humblot, Berlín, 1967, p. 111; H. OTTO, *Personales Unrecht, Schuld und Strafe*, en «Zeitschrift für die gesamte Strafrechtswissenschaft», 3/1975, pp. 539 ss.

delictivos, o bien de modificar los ya existentes, de manera que adquieran relevancia penal los ataques dirigidos a los sistemas de inteligencia artificial³⁷. La cuestión se enmarca, desde una perspectiva más amplia, en la posibilidad de reconocer la titularidad de derechos y deberes a sujetos que, en principio, no son humanos.

A este respecto, ha evolucionado mucho la manera de entender la personalidad jurídica en relación con la inteligencia artificial. En un primer momento, parecía hasta complejo construir máquinas capaces de lo que hoy, en cambio, pudiera ser logro del futuro próximo, si no lo es ya del presente³⁸. El mismo concepto de *machina sapiens*³⁹, anteriormente sólo concebible en el ámbito de la ciencia ficción, parece ahora el objetivo tendencial de la sociedad, que exige su consideración por el ordenamiento jurídico.

Para que las entidades puedan definirse inteligentes, se ha señalado que deberían poseer al menos los siguientes atributos: en primer lugar, la capacidad de comunicar (*communication*), es decir, la posibilidad de que la máquina entienda de algún modo las interacciones que se le dirijan; en segundo lugar, el conocimiento de sí (*internal knowledge*), es decir, la capacidad de comprenderse a sí mismo; después, el conocimiento del mundo exterior (*world knowledge*), que implica la conciencia de los acontecimientos e informaciones externas a sí; la intencionalidad (*intentionality* o *goal-driven behavior*), es decir, la habilidad de realizar una o más acciones para alcanzar un determinado objetivo; por último, la creatividad (*creativity*), o sea la capacidad de buscar una solución alternativa

³⁷ BASILE, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, en «Diritto Penale e Uomo», 10/2019, pp. 32-33.

³⁸ Sobre el reconocimiento de la personalidad jurídica a máquinas y robots, y acerca de la posibilidad de crear una inteligencia artificial en un contexto temporal ya superado, v. L.B. SOLUM, *Legal Personhood for Artificial Intelligences*, en «North Carolina Law Review», 4/1992, pp. 1231 ss. Para formarse una idea sobre la vivacidad del debate ya en ese entonces, v. *ivi*, p. 1237, *sub nota* 24.

³⁹ Existen notables discrepancias sobre este concepto, a la luz de la perenne complejidad de proporcionar una definición válida incluso para la inteligencia, por así decirlo, natural. En este sentido, cfr. el pensamiento de Aristóteles en las palabras de H. VELÁZQUEZ FERNÁNDEZ, *¿Qué tan natural es la inteligencia artificial? Sobre los límites y alcances de la biomimesis computacional*, en «Naturaleza y Libertad. Revista de estudios interdisciplinarios», 12/2019, p. 238: «cuando intentamos imitar lo natural mediante lo artificial, es importante diferenciar las artes que reproducen la naturaleza, de otras que solo intentan perfeccionarla, completarla o modificarla para bien». Desde una perspectiva jurídico-penal, v. G. HALLEVY, *The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control*, en «Akron Intellectual Property Journal», 2/2010, pp. 175 ss.

en el caso de que la acción (intencional) no haya alcanzado su finalidad⁴⁰.

Ahora bien, no es esta la sede para evaluar las evidencias científicas que se han aportado a favor y en contra de la humanización de los sistemas inanimados. Pues, desde un punto de vista ontológico, la afirmación de que los artefactos no tienen –y quizás nunca podrán tener– sentimientos y no sienten emociones reales, es quizás la respuesta de una pregunta erróneamente planteada. En realidad, en un sentido similar al acertadamente brindado por KAPLAN, la pregunta básica de si los sistemas inanimados son humanizables no debería dirigirse a un público indefinido, sino que exigiría una respuesta personalizada sobre la base de los sujetos o las cosas a que cada uno decide extender la «cortesía de su empatía»⁴¹.

Sin embargo, las reflexiones ontológicas mencionadas exceden del objetivo que se propone el presente estudio, que a lo mejor debería ceñirse a la deontología de dicho fenómeno humanizador. En efecto, desde el punto de vista del Derecho penal, no se trata de establecer si se le concede la condición de ‘vida humana’ a un sistema de inteligencia artificial; más bien se trataría de averiguar si la protección de dicha ‘vida’ en el marco de los bienes jurídicos penalmente relevantes está permitida por la *extrema ratio* que trasluce del carácter necesariamente lesivo de la Ley penal.

Aun así, la cuestión formulada escaparía del alcance del concepto de culpabilidad, tanto en su vertiente psicológica como en la concepción normativa más moderna. La incorporación de un determinado bien jurídico entre los que merecen tutela penal va más allá de las problemáticas inherentes al principio de culpabilidad y, junto con todas las implicaciones de carácter político-criminal, se sitúa idealmente en la teoría del bien jurídico y de la legitimación del *ius puniendi* estatal⁴².

Podría objetarse que, para dotar al hecho delictivo del carácter de

⁴⁰ La enucleación de los atributos de las entidades inteligentes se desarrolló por R.C. SCHANK, *What Is AI, Anyway?*, en «AI Magazine», 4/1987, p. 60.

⁴¹ En sentido similar se expresa, con un sugestivo ejemplo, J. KAPLAN, *Artificial Intelligence. What Everyone Needs to Know*, Oxford University Press, Oxford, 2016, pp. 85 ss. «On the one hand, our intuitions lead us to believe that machines, no matter how sophisticated, raise no ethical concerns in their own right. On the other, we can't comfortably exclude certain entities from the community of living things based solely on what materials they are composed of. My personal opinion, not universally shared, is that what's at issue here is little more than a decision we get to make as to whom, or to what, we choose to extend the courtesy of our empathy» (*ivi*, p. 86).

⁴² Sobre estas teorías, v. R. HEFENDEHL, A.V. HIRSCH, W. WOHLERS (eds.), *La teoría del bien jurídico. ¿Fundamento de legitimación del Derecho penal o juego de abalorios dogmático?*, Marcial Pons, Madrid, 2016, *passim*, traducción de R.A. Guirao, M. Martín Lorenzo, Í. Ortiz de Urbina Gimeno.

‘culpable’, la lesión causada al bien jurídico sea subsidiaria frente al desvalor personal de la acción. Esta objeción fue impulsada, en su momento, por una parte de la doctrina alemana y obtuvo particular notoriedad en Italia, debido a la formulación del art. 27, apartado primero, de la Carta constitucional («la responsabilidad penal es *personalis*»). Sin embargo, es suficiente recordar la doctrina ampliamente consolidada, conforme a la cual dicha ofensa no representa una mera *quantité négligeable*, sino más bien un requisito necesario para que un hecho sea punible con arreglo a las Leyes penales⁴³. Desde luego se trata de una afirmación que se debe al posicionamiento, por parte de la Constitución, sobre la teleología de la pena y de las medidas de seguridad⁴⁴.

Ahora bien, por lo que respecta al presente estudio, se puede afirmar con suficiente seguridad que el Derecho penal, en última instancia, está encaminado a promover y proteger los intereses y los valores típicos de los seres humanos. En este sentido, se ha afirmado que el reconocimiento de derechos y obligaciones de entidades inhumanas siempre se lleva a cabo desde una perspectiva antropocéntrica, es decir, con arreglo al objetivo último de la protección de la persona humana⁴⁵. No obstante, mientras es

⁴³ F. BRICOLA, voz *Teoria generale del reato*, en «Novissimo digesto italiano», XIV, 1973, p. 63: «La concezione del reato come “*personales Unrecht*”, nella misura in cui riduce l’offesa al bene giuridico al rango di mera *quantité négligeable*, non si pone in esplicita antitesi con la Costituzione tedesca la quale, a differenza di quella italiana, non prende posizione sul teleologismo delle pene e delle misure di sicurezza e, quindi, non cristallizza, sulla base di quanto si dirà in seguito, la natura *necessariamente* offensiva dell’illecito penale».

⁴⁴ *Ivi*, pp. 62 ss. Cfr. también *supra*, *sub* nota 18. En España, la teleología de las medidas de seguridad se desprende expresamente del art. 27.3 de la Constitución. En Italia, en cambio, al no constar una expresa previsión de esa finalidad en el caso de las medidas de seguridad, sino solo para las penas, se ha debatido largamente sobre su naturaleza, inicialmente considerada de carácter administrativo (véase la propia rúbrica del Código penal italiano, «De las medidas administrativas de seguridad»), y sobre su finalidad, originalmente correspondiente a la mera prevención especial. Sin embargo, hoy la unanimidad de la doctrina está de acuerdo en apreciar su naturaleza de sanciones penales y en dotarlas de los rasgos finales reeducativos, al igual que las penas. Cfr., al respecto, A. CAVALIERE, *Misure di sicurezza e doppio binario. Considerazioni introduttive*, en «Rivista italiana di diritto e procedura penale», 1/2022, pp. 343 ss.; L. FURNARI, *Misure di sicurezza e doppio binario: un declino inarrestabile?*, en «Rivista italiana di diritto e procedura penale», 1993, 569 ss.; E. MUSCO, voz *Misure di sicurezza*, en «Enciclopedia del diritto», I, 1997, § 22; P. NUVOLONE, voz *Misure di prevenzione e misure di sicurezza*, en «Enciclopedia del diritto», XXVI, 1976, § 17.

⁴⁵ Con las palabras de S. RIONDATO, *Robotica e diritto penale (robots, ibridi, chimere e “animali tecnologici”)*, en *Genetics, robotics, law, punishment*, dirigido por D. Provolò, S. Riondato, F. Yenisey, Padova University Press, Padua, 2014, p. 603: «tali estensioni della nozione di “persona” e, più specificatamente, la qualificazione delle entità non-umane

cierto que tal reconocimiento se lleva a cabo desde una visión exquisitamente ligada a la percepción de dichas entidades por parte del ser humano en cada época, plantea problemas la tipificación de algunas conductas como, por ejemplo, los delitos contra los animales domésticos. Desde la perspectiva de la protección del ser humano, la tutela de la flora y de la fauna se habría incorporado al Código Penal de manera instrumental, de cara a proteger el equilibrio medioambiental a modo de delito obstáculo, con el objetivo último de tutelar al ser humano. Sin embargo, la introducción de los delitos contra los animales domésticos plantea alguna duda acerca de esa construcción, si bien podría considerarse que, de todos modos, la tutela a través del Derecho penal se ha otorgado de cara a proteger los sentimientos de las personas hacia aquellos.

Asimismo, se ha afirmado con acierto que no hay razones válidas para obstaculizar el reconocimiento de la subjetividad pasiva a los sistemas de inteligencia artificial, como sujetos dotados de derechos y obligaciones, siempre y cuando dicha atribución sea meramente ficticia y se justifique con la persecución de finalidades ‘humanas’, o al menos controladas por los seres humanos⁴⁶.

En este sentido, antes de acceder al otorgamiento de subjetividad pasiva a los sistemas de inteligencia artificial, merece la pena mencionar dos ejemplos sugestivos⁴⁷ que quizás proporcionen una idea sobre el alcance de los comportamientos de las máquinas en el marco de conductas que, si se dirigiesen hacia seres humanos, constituirían lesión de un bien jurídico penalmente protegido.

Es el caso, primero, de los programas de *doll therapy* o *pet therapy*,

come “soggetti” del reato, siano solitamente finalizzate a promuovere o proteggere gli interessi e i valori tipici dell’essere umano. Infatti, l’attribuzione di diritti, interessi ed obblighi a entità non-umane – siano esse animali o società – avviene sempre in base ai valori e al senso di giustizia del legislatore umano».

⁴⁶ *Ibidem*. Resulta oportuno destacar la discrepancia con la posición del autor, cuando sostiene que también los robots capaces de actuar de forma independiente e imprevisible están programados por los seres humanos, y por ello los propios seres humanos pueden impedir cualquier consecuencia. Hoy en día, la evolución de los sistemas de *machine learning* permite afirmar lo contrario: el aprendizaje autónomo, característico de los sistemas de inteligencia artificial, impide un control humano efectivo sobre las consecuencias de sus actuaciones independientes, constituyendo quizás el problema más nítido de la implementación de estas máquinas en la sociedad. Cfr., al respecto, J. PÉREZ ARIAS, *Algoritmos y big data en la responsabilidad penal: el reto de la cibercriminalidad en el Derecho Penal*, en esta obra, § 5; P.C. ARIAS, *Artificial Intelligence & Machine Learning: a model for a new judicial system?*, en «Revista Internacional Jurídica y Empresarial», 2020, pp. 81 ss.

⁴⁷ Se utilizan los ejemplos brindados por BASILE, *Intelligenza artificiale, loc. ult. cit.*

terapias dirigidas a sujetos enfermos (mentales, sobre todo) para que puedan desarrollar sentimientos y emociones, en las cuales las personas implicadas (no necesariamente inimputables) pueden llegar a causar la destrucción de estos robots, su maltrato y eventual lesión; y, en segundo lugar, del fenómeno denominado ‘violación robótica’⁴⁸, con particular referencia a las agresiones sexuales hacia robots que tienen rasgos parecidos a los de sujetos menores.

Los actos de ‘violencia’ perpetrados hacia las máquinas susodichas no constituyen delito con arreglo a la legislación vigente, a no ser que se enmarquen dentro de los tipos delictivos referidos exclusivamente a la protección de bienes materiales (*e.g.*, daños). Ahora bien, recuperando la sugestiva imagen retratada por BASILE⁴⁹, si la agresión sexual ‘robótica’ se llevara a cabo hacia un androide que está desarrollando la labor encomendada, es decir, un robot diseñado para ser recepcionista de un hotel o bien empleado en una tienda de ropa, ¿el hecho de que la conducta quede impune y su autor libre podría despertar un sentimiento de (legítima) inquietud por parte de la colectividad? En otras palabras, reformulando la pregunta con las lentes del jurista, *de iure condendo* ¿dicha conducta podría fundamentar un reproche penal o, al menos, un juicio de peligrosidad del autor?

Resulta claro que, finalmente, la alternativa conlleva distintas consecuencias jurídicas. Si se accede a fundamentar un reproche penal por las conductas llevadas a cabo en contra de un sistema inanimado, habría que reformar el Código penal con la introducción de la subjetividad pasiva en cuestión, que conllevaría la aplicación de una determinada pena. En cambio, si se decide fundamentar un juicio de peligrosidad del autor, las consecuencias en términos de medidas de seguridad o de prevención son distintas según la legislación vigente en cada ordenamiento, habida cuenta de la considerable diferencia que existe entre los diversos sistemas jurídicos positivos⁵⁰.

Con respecto a esta segunda opción, en España el principio de legalidad establecido en el art. 25 de la Constitución se ha interpretado de manera estricta, así que el Código penal de 1995 impide la aplicación de medidas

⁴⁸ Sobre este fenómeno, v. C. TORRAS GENÍS, L. DELICADO MORATALLA, *Blanco y Frio como El Polo Norte y El Polo Sur: Seres humanos y robots*, en «Sociología y tecnología: Revista digital de sociología del sistema tecnocientífico», 1/2022, pp. 263 ss. J. DANAHER, *Robotic Rape and Robotic Child Sexual Abuse: Should They be Criminalised?*, en «Criminal Law and Philosophy», 1/2017, pp. 71 ss.; M.H. MARAS, L.R. SHAPIRO, *Child Sex Dolls and Robots: More Than Just an Uncanny Valley*, en «Journal of Internet Law», 2017, pp. 3 ss.

⁴⁹ BASILE, *Intelligenza artificiale*, cit., p. 33.

⁵⁰ COBO DEL ROSAL, VIVES ANTÓN, *Derecho penal*, cit., p. 985 ss.

de seguridad a sujetos socialmente peligrosos, y restringe su imposición a quienes se enmarquen en el concepto de peligrosidad criminal, es decir, a aquellos individuos que, tras haber cometido algún delito, ostenten el riesgo de volver a violar la Ley penal. Pues, en este contexto, sólo cabría la posibilidad de fundamentar un juicio de peligrosidad si el autor de la conducta en contra del sistema inanimado haya realizado un hecho delictivo. Ahora bien, se podría sostener este fundamento del juicio de peligrosidad sobre la base de que el autor, en la mayoría de los casos, al menos cometería un delito de daños⁵¹, y la medida de seguridad se haría necesaria de cara a prevenir la reiteración futura del delito.

En Italia, en cambio, la problemática relativa a la viabilidad de un juicio de peligrosidad abre las puertas a debates todavía pendientes⁵². En el Código penal italiano, que acoge el sistema dualista, el concepto de peligrosidad está ligado a la peligrosidad social, que surte sus efectos tanto (i) después de la comisión de un delito, como (ii) antes de –y sin– que este se cometa, incluyendo en esta categoría la peligrosidad derivada de la realización de un cuasi-delito⁵³. De apreciarse la peligrosidad del sujeto, en el primer caso se podría aplicar una medida de seguridad, y en el segundo tanto una medida de seguridad⁵⁴ como una medida de prevención⁵⁵.

⁵¹ Por los problemas de causalidad que normalmente plantean, no se mencionan demás delitos que puedan apreciarse de forma colateral a la conducta realizada, como por ejemplo las lesiones psíquicas ocasionadas en los compañeros humanos del robot recepcionista violado.

⁵² Cfr., *amplius*, E. SQUILLACI, 'Ragioni' e 'costi' dell'attuale paradigma di prevenzione, en «Archivio penale web», 2/2020, pp. 1 ss.; A. MENCARELLI, *Le misure di sicurezza detentive al vaglio della Corte europea dei diritti dell'uomo*, en *La tutela della salute nei luoghi di detenzione. Un'indagine di diritto penale intorno a carcere, REMS e CPR*, dirigido por A. Massaro, RomaTrE-Press, Roma, 2017, pp. 333 ss.; A. MASSARO, *Pericolosità sociale e misure di sicurezza detentive nel processo di "definitivo superamento" degli ospedali psichiatrici giudiziari: la lettura della Corte costituzionale con la sentenza n. 186 del 2015*, en «Archivio penale», 2/2015, pp. 1 ss.; M. BERTOLINO, *Declinazioni attuali della pericolosità sociale: pene e misure di sicurezza a confronto*, en «Archivio penale», 2/2014, pp. 459 ss.

⁵³ Se trata de los casos previstos en los arts. 49, último apartado, y 115, último apartado, del Código penal italiano. Respectivamente, por un lado, se puede aplicar una medida de seguridad al sujeto que haya cometido un hecho creyendo erróneamente que este constituya un delito (mientras, en realidad, no lo constituye), y por otro, al sujeto que instigue a otro a cometer un delito, cuando la instigación no se haya acogido.

⁵⁴ De conformidad con el art. 202, apartado segundo, del Código penal italiano, cuando el sujeto no ha cometido un hecho delictivo, sólo se podrá aplicar una medida de seguridad si así lo prevé expresamente la Ley penal. Es el caso de los cuasi-delitos mencionados en la nota anterior, que están previstos específicamente en el Código.

⁵⁵ Con los requisitos y las formas previstas en el Código de las leyes antimafia y de las medidas de prevención (D.Lvo. italiano n. 159/2011, de 6 de septiembre). Para una

Por otro lado, con respecto al primer caso, es decir, a la fundamentación de un reproche penal, el ordenamiento vigente exigiría ser adaptado para incluir, entre los sujetos pasivos de cada delito que se considere relevante, los sistemas de inteligencia artificial. Esto se puede realizar, esencialmente, a través de la introducción de un título *ad hoc*, que incluyera delitos específicos cometidos en contra de dichos sistemas, o bien mediante el estudio de una cláusula general de aplicación automática en conjunción con las normas ya existentes, al hilo de la ‘cláusula de equivalencia’ establecida en el art. 40, apartado segundo, del Código penal italiano⁵⁶, si bien en este caso habría que manejar una pluralidad de desafíos interpretativos que quizás el Derecho penal –y con esta sinécdoque sus intérpretes– tendrían el placer de prevenir.

Pues bien, de considerar que el catálogo de los bienes jurídicos protegidos merece incorporar –aun bajo la máscara del *délit obstacle*– la integridad de las máquinas, aparecen las dudas teleológicas con respecto a la aplicación de la sanción correspondiente, que finalmente repercuten en la exigencia de culpabilidad del autor. De hecho, esta sanción quizás tendría sentido desde la rechazable visión orientada hacia el Derecho penal del enemigo impulsado por JAKOBS, que justificaría el «amplio adelantamiento de la punibilidad, es decir, [el] cambio de la perspectiva del hecho producido por la del hecho que se va a producir»⁵⁷. Sin embargo, para cumplir con las exigencias de prevención especial propiciadas por ROXIN, habría que averiguar «hasta qué punto, en los casos de circunstancias personales irregulares o condicionadas por la situación conviene una sanción penal a una conducta que, en principio, está amenazada con una pena»⁵⁸. Y es evidente que, en palabras del mismo autor, «una educación para la vida legal en la libertad del Estado de Derecho puede difícilmente llevarse a cabo con la privación de todas las libertades»⁵⁹.

imagen de los antiguos supuestos de peligrosidad *ante delictum*, v. F. TAGLIARINI, voz *Pericolosità*, en «Enciclopedia del diritto», XXXIII, 1983, § 12.

⁵⁶ «No impedir un evento que se tiene la obligación de impedir equivale a causarlo». Con esta fórmula, el Código penal italiano amplía la esfera de los delitos incorporando las omisiones impropias.

⁵⁷ JAKOBS, *La autocomprensión de la ciencia del Derecho penal ante los desafíos del presente*, en *La ciencia del Derecho penal ante el nuevo milenio*, dirigido por A. Eser, W. Hassemer, B. Burkhardt, Tirant lo Blanch, Valencia, 2004, p. 58.

⁵⁸ ROXIN, *Política criminal*, cit., p. 59.

⁵⁹ *Ivi*, p. 50. Hay determinadas conductas que «en el caso concreto devengan en un problema individual y deban tenerse en cuenta también dentro de lo posible y de los límites que permita la prevención general las finalidades preventivas especiales, p. ej., la resocialización del delincuente o, por lo menos, su no desocialización». Así,

Con ello, no sería conveniente instrumentalizar la intervención penal anticipándola hasta las conductas dirigidas a los sistemas inanimados, ya que tendría un carácter marcadamente 'educativo', cuyo objetivo difícilmente podría alcanzarse a través de la privación de las libertades esenciales. Más aún, no se puede olvidar que «cuanto mayor sea una intervención en un derecho fundamental, tanto mayor deberá ser la certeza de las premisas que fundamentan la intervención»⁶⁰.

Todavía está por analizar la posibilidad de utilizar la opción menos invasiva, es decir, aplicar una medida de seguridad al autor de las conductas en cuestión que se revele peligroso. En este sentido, de acuerdo con el modelo italiano, además de poder declarar la peligrosidad en su forma específica (habitualidad, profesionalidad y/o tendencia a delinquir⁶¹), también se puede valorar su modalidad genérica o simple⁶², para cuya apreciación es necesario considerar los criterios indicados en el mismo artículo que establece las reglas para medir la pena⁶³, con una formulación que todavía refleja el influjo de algunas exigencias (en parte vetustas) del contenido de la culpabilidad⁶⁴.

Pues bien, además de las complejidades debidas a la diferencia entre los distintos sistemas penales, que implicaría un tratamiento heterogéneo en los varios ordenamientos europeos, cabe insistir en las mismas razones formuladas anteriormente con respecto a la aplicación de la pena. En efecto, los elementos que conforman y de los que se desprende la peligrosidad social de un sujeto se acercan insidiosamente a las circunstancias valoradas en el ámbito de la culpabilidad. Pese a que peligrosidad y culpabilidad tengan

Id., *Culpabilidad y prevención en Derecho penal*, Editorial Reus, Madrid, 1981, p. 29, traducción de F. Muñoz Conde.

⁶⁰ R. ALEXY, *Teoría de la argumentación jurídica. La teoría del discurso racional como teoría de la fundamentación jurídica*, Centro de Estudios Políticos y Constitucionales, Madrid, 1989, p. 370, traducción de M. Atienza e I. Espejo.

⁶¹ Cfr. el art. 109, apartado primero, del Código penal italiano: «la declaración de habitualidad o de profesionalidad en el delito o bien la de tendencia a delinquir conlleva la aplicación de medidas de seguridad».

⁶² Se trata de la peligrosidad social de los sujetos que, tras haber realizado un hecho delictivo o un cuasi-delito (cfr. *supra*, *sub* notas 53 y 54), manifiesten el riesgo de reiteración delictiva. Sobre la diferenciación entre peligrosidad simple y específica, cfr. V. MANZINI, *Trattato di diritto penale italiano*, vol. III, Utet, Torino, 1981, pp. 249 ss.

⁶³ El art. 203, apartado segundo, del Código penal italiano establece que la condición de persona socialmente peligrosa se desprende de las circunstancias indicadas en el art. 133, que a su vez incluye los criterios para la medición de la pena (gravedad del hecho y capacidad para delinquir).

⁶⁴ Cfr. SANTAMARIA, *op. ult. cit.*, pp. 660 ss.

que mantenerse distintas por fundamentar dos consecuencias jurídicas diferentes⁶⁵, es evidente que las circunstancias personales del sujeto, que se valoran –desde una perspectiva distinta, intrafactual o extrafactual– terminan por ser las mismas.

Quizás las exigencias de prevención especial reeducativa, que conlleva la comisión de conductas abstractamente delictivas hacia los sistemas de inteligencia artificial, justificarían la anticipación de la tutela penal en el marco de un Derecho penal de la voluntad (*Willensstrafrecht*) o, tal vez mejor, de la actitud interna (*Gesinnungsstrafrecht*), que sin embargo son incompatibles con el principio del hecho, impuesto por los mandatos constitucionales⁶⁶.

Hoy en día, pues, y hasta que haya cambios relevantes en el ámbito de la ingeniería relacionada con de los sistemas de inteligencia artificial, el principio de subsidiariedad quizás respalda la protección de estos nuevos ‘sujetos pasivos’ a través de otras ramas del Derecho⁶⁷.

3.2. *Hacia una subjetividad activa instrumental: la inteligencia artificial como autora de delitos*

Después de haber analizado las cuestiones relativas a la inteligencia artificial como potencial víctima de un delito, cabe plantear las problemáticas inherentes a la otra cara de la moneda, es decir, a la posibilidad de que los sistemas inanimados puedan encajar en el concepto de autoría a efectos de la Ley penal. En otras palabras, para conseguir el objetivo que pretende alcanzar este escrito no se puede prescindir del análisis sobre la oportunidad de castigar un sistema inanimado y, de así estimarla, habrá que examinar sus implicaciones en la adecuación del ordenamiento jurídico-penal.

En primer lugar, pues, surge la cuestión de si es oportuno ‘reprochar’

⁶⁵ *Ivi*, p. 663 ss.

⁶⁶ BRICOLA, *op. ult. cit.*, pp. 65 ss.; J.M. PERIS RIERA, *Delitos de peligro y sociedad de riesgo: una constante discusión en la dogmática penal de la última década*, en *Estudios penales en homenaje al profesor Cobo del Rosal*, dirigido por J.C. Carbonell Mateu, B. del Rosal Blasco, L. Morillas Cueva, E. Orts Berenguer, M. Quintanar Díez, Dykinson, Madrid, 2005, pp. 687 ss. Con respecto a la proximidad entre el Derecho penal del autor y el Derecho penal del enemigo, v. G. JAKOBS, M. CANCIO MELIÁ, *Derecho penal del enemigo*, Civitas, Madrid, 2003, pp. 100 ss.

⁶⁷ Sobre las cuestiones de legitimidad de la intervención penal, v. M. CORCOY BIDASOLO, *Exigibilidad en el ámbito del conocimiento y control de riesgos: teorización*, en «Revista catalana de seguretat pública», 13/2003, pp. 31 ss; *amplius*, EAD., *Delitos de peligro y protección de bienes jurídico-penales supraindividuales*, Tirant lo Blanch, Valencia, 1999, *passim*.

a un sistema inanimado, en el sentido de que se le pueda conceder subjetividad activa. Sólo después, de considerar adecuada dicha concesión, se podrá analizar la medida en que podría adaptarse la legislación vigente. La doctrina se ha ido interrogando sobre la duda de si *machina delinquere potest*⁶⁸, al igual que ocurrió en su momento con las personas jurídicas⁶⁹. De hecho, la fórmula aquí referida se ha plasmado sobre la base de otro conocido brocardo que, en lugar de la máquina, alude a la sociedad⁷⁰.

Pues bien, ante la duda se puede intentar razonar *a contrario* y plantear la hipótesis de una respuesta negativa. Por razones de lógica (quizás simplista), y sobre todo en virtud del empirismo que informa hasta las quiebras más evidentes del Derecho penal, la habitualidad en concebir formas de autoría estrictamente ligadas a los seres humanos podría llevar a excluir perentoriamente que los sistemas inanimados adquieran subjetividad activa. No obstante, dicha denegación supondría la elusión de exigencias de política criminal que hoy en día están lanzando signos de alarma.

En efecto, no faltarían motivos para obstaculizar el reconocimiento de la subjetividad activa en cuestión. Es suficiente pensar en la 'personalidad' de la responsabilidad penal, mencionada arriba, o en las propias palabras de COBO DEL ROSAL sobre la persona jurídica⁷¹, para asombrarse ante una forma de autoría tan revolucionaria y, al mismo tiempo, temerosa. Sin embargo, al igual que en la historia se han sucedido modificaciones

⁶⁸ Cfr. J.M. MUÑOZ VELA, *Inteligencia Artificial y responsabilidad penal*, en «Derecho Digital e Innovación. Digital Law and Innovation Review», 11/2022, pp. 1 ss.; J. VALLS PRIETO, *Inteligencia artificial, Derechos humanos y bienes jurídicos*, Aranzadi, Pamplona, 2021, *passim*; P.M. DE LA CUESTA AGUADO, *Inteligencia artificial y responsabilidad penal*, en «Revista penal México», 2020, pp. 51 ss.; I. LLEDÓ BENITO, *El impacto de la robótica. La inteligencia artificial y la responsabilidad penal en los robots inteligentes*, en «Foro galego. Revista xurídica», 2020, pp. 173 ss.; D. LIMA, *Could AI Agents Be Held Criminally Liable? Artificial Intelligence and the Challenges for Criminal Law*, en «South Carolina Law Review», 3/2018, pp. 677 ss.

⁶⁹ *Amplius*, J. PÉREZ ARIAS, *Sistema de atribución de responsabilidad penal a las personas jurídicas*, Dykinson, Madrid, 2014, *passim*.

⁷⁰ BASILE, *Intelligenza artificiale*, cit., *sub nota* 100, p. 27.

⁷¹ «Desde otra perspectiva, puede concluirse que, en ninguna parte del vigente Código penal se afirma que se construya una autoría criminal de las personas jurídicas. Porque la autoría ni se puede improvisar ni se puede derivar de una prejuiciosa y errática interpretación, por muy buena y amplia voluntad represiva que conlleven mentalidades defensoras extremadamente de la sociedad o del Estado». Si se lee el Código penal «como debe ser, de forma no locoide e imprejujada, no hay posibilidad de afirmar una *societas* como realizadora, en régimen de autoría estricta o de complicidad, de infracción penal de clase alguna». Así, COBO DEL ROSAL, *Societas delinquere non potest*, en «Anales de derecho», 2012, pp. 8-9.

(incluso) en la teoría del delito, como resultado de la evolución de la realidad social en las distintas épocas⁷², quizás es posible plantear cambios también con respecto a conceptos más estrechos. Pues, las renovadas exigencias sociales necesitan, al menos, ser evaluadas por el Derecho, que en su caso se adaptará con las herramientas que posee.

Resulta incontrovertido que ya estamos en una época en la que el *machine learning* ha salido del control de los diseñadores de los sistemas de inteligencia artificial. Desde luego, si en el futuro se da la posibilidad de supervisar exhaustivamente este proceso de autoaprendizaje de las máquinas, las consideraciones que aquí se están desarrollando tendrían un alcance diferente y, por ello, habrían de reformularse. Sin embargo, en la actualidad, la ciencia parece viajar en sentido inverso. Pues, como ya se ha dicho, la evolución de la tecnología está llevando a la peligrosa aparición de sistemas que, aunque se hayan diseñado para alcanzar determinados fines, escapan al control de los programadores y se atreven no sólo a actuar de forma maquiavélica, de modo que cualquier medio (aun ilegal) se hace adecuado mientras se llegue al objetivo, sino también a buscar otros propósitos que ni se les habían ocurrido a los diseñadores.

Siendo este el escenario, es oportuno que el Derecho preste atención a los cambios repentinos y a las evoluciones de la técnica, ya que tienen efectos considerables en la sociedad. En modo alguno esto implica la equiparación entre artefactos y seres humanos, ni mucho menos. Las máquinas quedan al servicio del hombre, tanto en el caso de que la máquina sea dirigida por el ser humano, como cuando ella actúe de forma independiente. Y para que estén al servicio del hombre, el Derecho positivo debe reflejar este axioma.

De no ser así, es decir, de desvincular la actuación inhumana del control humano, se daría rienda suelta a la derresponsabilización moral de los individuos, que lleva aparejada una inaceptable debilitación de la tutela de los bienes jurídicos⁷³. A modo de ejemplo, cualquier sujeto interesado en cometer una conducta delictiva podría alcanzar su objetivo amparándose en la utilización de una máquina y en su capacidad de autoaprendizaje, habida cuenta de las dificultades probatorias y causales que propiciarían su impunidad.

Resumiendo, para contestar a la pregunta que impulsó estas reflexiones, es oportuno y necesario que el Derecho se plantee la posibilidad de

⁷² Cfr. *supra*, *sub* nota 1.

⁷³ Así advierte BASILE, *Intelligenza artificiale*, cit., p. 29, mencionando, *sub* nota 106, la «deshumanización de las ofensas» descrita por F. CONSULICH, *Il nastro di Mobius. Intelligenza artificiale e imputazione penale nelle nuove forme di abuso del mercato*, en «Banca Borsa Titoli di Credito», 2/2018, p. 204.

‘reprochar’ a los sistemas inanimados de inteligencia artificial, a fin de evitar una quiebra evidente de los bienes jurídicos ya protegidos por el ordenamiento. No resulta difícil imaginar la época en que los accidentes de tráfico se deban a coches conducidos por sí mismos, las pizarras digitales de los centros de educación se conviertan en robots autónomos capaces de lesionar a los discentes, los mismos robots adquieran impulsos sexuales aptos para llevarlos a cometer violaciones, o los algoritmos aprecien el valor del dinero hasta llegar a sustraerlo de unas cuentas bancarias y traspasarlo a otras. Así pues, el Derecho debe reaccionar antes de perder el control sobre la ordenación de una sociedad que paulatinamente está avanzando hacia nuevas fronteras; en otras palabras, antes de que la tenencia de un carné de conducir se convierta en algo tan insólito como, a día de hoy, no tenerlo.

Ahora bien, al haber apreciado la oportunidad (*an*) de atribuir subjetividad activa a los sistemas de inteligencia artificial, procede analizar la manera (*quomodo*) en que dicho reconocimiento se puede llevar a cabo en el marco del ordenamiento penal. Hasta la fecha, la responsabilidad penal gira en torno a la comisión u omisión de conductas realizadas exclusivamente por el ser humano, no sólo desde la perspectiva determinista, ya que también son delictivos algunos hechos realizados sin que subsista la intención del agente. En este sentido, de nada sirve objetar que también las personas jurídicas poseen subjetividad activa en Derecho penal. De hecho, como se dirá en breve, entre la *societas* y la *machina* hay diferencias significativas, empezando por la dirección siempre ‘humana’ de la persona jurídica, que al contrario de los productos de inteligencia artificial, no tiene autonomía cognitiva y decisional propia.

Lo que ahora se plantea, en cambio, es el traspaso de actuaciones penalmente relevantes del ser humano a una pluralidad de máquinas independientes. De algún modo se sale de la esfera de actuación humana para entrar en un escenario donde, por leves que sean, los comportamientos se desarrollan autónomamente por sistemas inanimados, lo que conlleva una alienación de la responsabilidad del individuo⁷⁴. Debe advertirse sin más tardanza que se presupone un contexto en el que las máquinas son capaces de actuar independientemente del control humano; de lo contrario, las consideraciones que figuran a continuación carecerían de sentido.

El análisis se desarrollará siguiendo dos directrices paralelas que sólo se escinden por razones de comodidad expositiva. Por un lado, se intentará examinar las implicaciones del reconocimiento de la subjetividad activa de los sistemas inanimados en la teoría del delito; por otro, se estudiará

⁷⁴ Cfr. C. BAGNOLI, *Teoria della responsabilità*, Il Mulino, Bologna, 2019, p. 77.

el mismo fenómeno desde un punto de vista penológico, pues no es posible imaginar un sistema de normas penales que despliegue sus efectos coercitivos sin hallarse respaldado por la posibilidad de conminar una sanción efectiva.

En el marco de la teoría del delito y, más detenidamente, del contenido de la culpabilidad, aunque se cuestione la idea de un concepto unitario⁷⁵ no cabe duda de la necesidad de integrar los varios elementos establecidos por el Derecho positivo para apreciar la existencia de un delito. Ahora bien, en la medida en que se teorice una autoría mecánica, esta implicaría la acreditación de una culpabilidad ‘deshumana’⁷⁶, o en sentido similar pero distinto ‘deshumanizada’. En ambos los casos, hay que afrontar la dificultad, por un lado, de adaptar conceptos que son propios de las personas físicas y, por otro, de reformularlos según las peculiaridades de los nuevos sujetos activos.

En el primer caso, para que se pueda hablar de una culpabilidad ‘deshumana’ se debería adaptar el régimen penal ordinario para centrarlo en las máquinas, cuya responsabilidad penal tendría que superar los peldaños exigidos por la dogmática (más bien, por el Derecho positivo). En este sentido, para que se pueda apreciar la existencia de un delito, la adaptación del Derecho penal conllevaría la incorporación de conceptos todavía ambiguos, elaborados por las distintas ciencias que operan en el sector de la inteligencia artificial. En otras palabras, el Derecho penal debería aceptar incondicionalmente los fundamentos científicos del *machine learning*, según los cuales las propias máquinas dispondrían de capacidad natural, o bien de capacidad de entender y querer, serían capaces de comprender el alcance de la antijuridicidad de una conducta y, más aún, podrían realizar comportamientos que se encajan en el complejo marco del dolo y de la imprudencia.

Ahora bien, aunque ningún ordenamiento se haya atrevido a acogerla, en la doctrina hubo manifestaciones aisladas de apoyo a dicha postura. Entrando en este posicionamiento, los problemas más evidentes están relacionados con el dolo y la imprudencia, además de la conciencia de la antijuridicidad y su contenido. Efectivamente, ¿es posible imaginar que un algoritmo posea la voluntad consciente de realizar un hecho delictivo, o más aún, sea capaz de desatender una regla cautelar que fundamente un juicio de reprochabilidad por imprudencia? Las opiniones que admiten dichas posibilidades⁷⁷ parten de la configurabilidad en abstracto de una tipicidad mecánica, dejando al margen como mínimo los problemas relativos a la

⁷⁵ Cfr. *supra*, sub nota 23.

⁷⁶ BASILE, *Intelligenza artificiale*, cit., p. 30.

⁷⁷ HALLEVY, *op. ult. cit.*, pp. 171 ss.

estructura de la antijuridicidad en los sistemas jurídicos continentales y las críticas que surgirían desde la perspectiva procesal.

Aunque se trate de un modelo de responsabilidad directa abstractamente configurable, se trataría de condenar por acciones y omisiones naturalísticas –*rectius*, actividades e inactividades– a las que el Derecho penal otorga relevancia de por sí, sin que siquiera existan instrumentos para averiguar el elemento subjetivo subyacente. Se ha hablado de *general intent*⁷⁸, pero no resulta claro cómo este vínculo psicológico entre la conducta y la máquina pueda resultar probado a efectos penales. Además, con respecto a la antijuridicidad, los modelos antropocéntricos de algunas causas de justificación –y lo mismo debe predicarse sobre algunas de las demás eximentes consideradas, por la mayoría, causas de exculpación⁷⁹– podrían suponer, de forma sistemática, la impunidad de la máquina.

Piénsese, a modo de ejemplo y sin entrar en casos más complejos, en el estado de necesidad alegado por (¿quién representa?) el sistema de inteligencia artificial que, para evitar un mal propio (en el que cabría simplemente su alteración o descomposición mecánica), lesione un bien jurídico de un ser humano; o incluso en el sistema inanimado que, diseñado por o al servicio de la autoridad, intente cumplir el deber (legítimo) de realizar la tarea encomendada y para eso se vea obligado a realizar una conducta delictiva. Todo ello sin perjuicio de los problemas de carácter procesal⁸⁰, empezando por la dificultad de encontrar un procurador dispuesto a representar la máquina en el juicio sin poder siquiera comunicar con ella, hasta llegar a enflaquecer la intolerancia del delito «ad ogni articolazione che faccia in qualche modo ricadere sull'imputato l'onere della prova (o il rischio della mancata allegazione) di elementi d'ordine positivo che ne caratterizzano la struttura»⁸¹.

Acoger este modelo de responsabilidad directa se hace, pues, inviable por suponer la creación de un tejido de problemas⁸² –entre ellos los aquí referidos– difícilmente solucionables. Mientras se ‘humanizan’ los sistemas de inteligencia artificial, permanecerían ‘deshumanos’ –y, entonces, sin contenido– algunos de los elementos que se han plasmado en la teoría del delito sobre la base de las exigencias y especificidades del ser humano.

Por otro lado, con un grado de viabilidad más elevado se puede plantear

⁷⁸ *Ivi*, pp. 177 ss.

⁷⁹ Cfr. el concepto de antijuridicidad especialmente cualificada (*supra*, *sub* nota 23).

⁸⁰ Al respecto, v. R. CASTILLO FELIPE, S. TOMÁS TOMÁS, *Proceso penal e inteligencia artificial: prevenciones en torno a su futura aplicación en la fase de juicio oral*, en esta obra.

⁸¹ BRICOLA, *op. ult. cit.*, p. 92.

⁸² Cfr., *amplius*, A. CAPPELLINI, *Machina delinquere non potest? Brevi appunti su intelligenza artificiale e responsabilità penale*, en «Criminalia», 2018, pp. 499 ss.

la elaboración de una culpabilidad ya no deshumana, sino ‘deshumanizada’, en el marco de una dogmática que sabe adaptarse a las renovadas necesidades sociales, que constan de una creciente demanda de tutela frente a posibles ataques de los sistemas de inteligencia artificial lesivos de bienes jurídicos constitucionalmente protegidos.

La elaboración de este modelo de responsabilidad penal de los sistemas de inteligencia artificial supondría un cambio de paradigma, es decir, la creación de una *fictio iuris* capaz de otorgarles personalidad jurídica y aptitud autorial. Se trataría de una responsabilidad autónoma frente a la que seguiría perteneciendo a las personas físicas. De hecho, como ya se advirtió anteriormente, aquí se toman en consideración los comportamientos de los sistemas inanimados que actúan de forma independiente, pues, de actuar conjuntamente con un sujeto humano, la implementación de este modelo no perjudicaría la responsabilidad individual.

La creación de dicha ficción jurídica se vería facilitada por el antecedente significativo de las *societates*, aunque existan diferencias considerables entre las dos figuras. Desde el punto de vista de la regulación aplicable, quizás convendría dotar a los sistemas de inteligencia artificial de personalidad jurídica a través de la inscripción en registros públicos, para que pueda plasmarse coherentemente su peculiar imputabilidad. Del mismo modo, sería oportuno restringir el área de lo penalmente relevante mediante la selección de un *numerus clausus* de hechos punibles, bien a través de una cláusula extensiva genérica, bien con la incorporación de tipos autónomos de aplicación exclusiva a los sujetos activos de nuevo cuño, además de rediseñar el sistema de las eximentes para adaptarlo a exigencias ya no humanas.

Sin embargo, algunas soluciones escogidas para las personas jurídicas no serían adecuadas para regular el fenómeno de la inteligencia artificial. En primer lugar, no procedería restringir la subjetividad activa excluyendo los sistemas inanimados pertenecientes al Estado, ni podría estigmatizarse un catálogo de personas físicas al hilo del primer apartado del actual art. 31 *bis* del Código penal español. En esto radica la principal diferencia entre personas jurídicas y máquinas: estas últimas carecen del sustrato humano que dirija su actuación. Entiéndase: en el caso de que las máquinas cometiesen un delito en virtud de la dirección humana, no cabe duda de que sea la persona física la que debe responder penalmente, con arreglo a la legislación existente; el problema surge cuando la máquina actúa independientemente de las determinaciones (aun dolosas o imprudentes) humanas.

Por ello, también hay que regular de forma distinta el criterio de imputación subjetiva para dichas actuaciones mecánicas. Pues, la tesis predominante en el ámbito de las personas jurídicas acerca de la necesidad

de acudir a la persona física que realmente llevó a cabo la conducta delictiva, encuentra respaldo en la existencia de ese sustrato humano que aquí, en el caso de los sistemas autónomos de inteligencia artificial, no se puede apreciar. Ante la carencia de pruebas científicamente satisfactorias que permitan averiguar un verdadero elemento volitivo de las máquinas, y frente a la imposibilidad de escoger un ‘agente modelo’ que presente *eaedem condiciones et professiones*, quizás el caso de la inteligencia artificial puede constituir el único supuesto de responsabilidad objetiva que la Constitución esté dispuesta a tolerar. De ser así, la responsabilidad penal de los sistemas inanimados prescindiría de la existencia del vínculo psicológico entre el ‘autor’ y el hecho, y asimismo de la cognoscibilidad de la Ley penal por parte del sujeto activo.

Por último, después de haber intentado examinar algunas de las implicaciones del reconocimiento de la subjetividad activa de los sistemas inanimados, queda por abordar el mismo fenómeno desde un punto de vista penológico. No hace falta detenerse en la exigencia de que, en relación con todo delito, las penas sean efectivas y proporcionadas al hecho cometido⁸³; de no ser así, la tipificación del delito quedaría sin efecto alguno.

Acogiendo la adaptación del Derecho penal que se ha propuesto para admitir que los sistemas de inteligencia artificial puedan cometer delitos, el debate sobre el fundamento y las funciones de las penas correspondiente requeriría un examen mucho más amplio. Aun así, se intentará desarrollar unas breves consideraciones tanto desde la perspectiva teleológica, con respecto a las tres teorías clásicas sobre la finalidad de la sanción, como desde la perspectiva práctica, pues si la sanción efectiva no tiene la aptitud para lograr el objetivo fijado a nivel abstracto, se hace innecesaria y el ejercicio del *ius puniendi* estatal representaría un mero desgaste. Debe advertirse previamente que, mientras «el fundamento del castigo penal que recibiría la persona jurídica sería la existencia de defectos en su organización, la tenencia de una estructura opaca, la inexistencia de controles sobre sus empleados»⁸⁴, en el caso de las máquinas no puede limitarse a eso.

La idea que debería apoyarse es que la elaboración del modelo de

⁸³ Ya advertía dichas exigencias C. BECCARIA, *Tratado de los delitos y de las penas*, Heliasta, Buenos Aires, 1993, pp. 68 ss.

⁸⁴ A. NIETO MARTÍN, *Problemas fundamentales del cumplimiento normativo en el derecho penal*, en *Compliance y teoría del Derecho penal*, dirigido por L. Kuhlen, J.P. Montiel, Í. Ortiz de Urbina Gimeno, Marcial Pons, Madrid, 2013, pp. 21 ss. Se expresa en contra de la responsabilidad penal de las personas jurídicas L. GRACIA MARTÍN, *Crítica de las modernas construcciones de una mal llamada responsabilidad penal de la persona jurídica*, en «Revista Electrónica de Ciencia Penal y Criminología», 2016, pp. 1 ss.

responsabilidad penal de los sistemas en cuestión, en el que se incluye la aplicación de las sanciones correspondientes, tiende exclusivamente a la protección última del ser humano. Por dicha razón, y dado que en los modernos sistemas constitucionales se ha ido sentando la idea de que la retribución subyacente a la pena sólo tiene sentido como medio preventivo, con las sanciones dirigidas a las máquinas se intentaría alcanzar, por un lado, una finalidad de prevención especial, que está encaminada a prevenir la reiteración delictiva por parte del ‘autor’, y por otro lado, un objetivo de prevención general, porque la vigencia de estas normas, además de tranquilizar a la colectividad, podría tender a evitar que se diseñen sistemas de inteligencia artificial parecidos. Con ello, como consecuencia del acogimiento del *versari in re illicita*, se debería renunciar a la función reeducativa prevista en el régimen general.

A nivel práctico, las sanciones que pueden conminarse a las máquinas no se enmarcan en el catálogo tradicional de las penas. En este sentido, es hasta antieconómico aplicar a una máquina la pena de prisión, así como imposible atacar un patrimonio que no posee. Con las dificultades que conllevan las hipótesis más complejas –*e.g.*, ¿cómo se plantea la destrucción o inutilización de un algoritmo?– se trata de estudiar un nuevo sistema penológico que esté de acuerdo con las finalidades que merecen ser perseguidas. Al efecto, téngase en cuenta que las dos sanciones quizás más fáciles de imaginar (el apagado definitivo y temporal de la máquina) traicionarían la teleología fijada arriba. Mientras el apagado temporal del sistema inanimado no lograría prevenir la reiteración delictiva por el mismo, su destrucción (o inutilización, o apagado definitivo) –que en el fondo refleja los caracteres de la pena capital– sería aceptable al haber renunciado al propósito reeducativo, pero a lo mejor obstaculizaría la posibilidad de mejorar el diseño de las demás máquinas.

3.3. La inteligencia artificial como instrumento de apoyo para el ejercicio de la justicia

Aunque no se puedan abordar en esta sede los aspectos procesales implicados, parece oportuno hacer mención a las consecuencias del uso de la inteligencia artificial como instrumento de apoyo para el ejercicio de la justicia.

Por un lado, hay que hacer hincapié en la posible afectación del modelo de libre valoración de la prueba por parte del juez, que se vería afectado en el caso de intervenir instrumentos de inteligencia artificial,

como los polígrafos, en la detección de la veracidad de las pruebas. Dicha implementación podría intervenir tanto de manera dura, si al juez no se le deja la posibilidad de moldear los resultados elaborados por la máquina, como de manera débil, si la mediación del juez persona física permite mantener en vigor el principio del *iudex peritus peritorum*. Es evidente que, en el marco de una justicia que ha de ser administrada en favor del reo, las máquinas que analizan los hechos, estudian las pruebas e incluso producen resoluciones judiciales, no pueden alcanzar la autonomía que supondría la opción de intervención más dura⁸⁵. Se hace necesario, ahora más que nunca, el control constante del órgano judicial para escapar de los supuestos en que la mecánica se apodera del control sobre las libertades fundamentales de los individuos. Así se reafirman las exigencias de culpabilidad que imponen la huida de Derechos penales desvinculados del hecho y apegados a los procesos emocionales internos del autor.

Por otro lado, aunque parezca ciencia ficción, no se puede descartar *a priori* que alguien llegue a formular una teoría del delito en la que el medio tecnológico de apoyo a la justicia se convierta en auténtico elemento negativo del hecho. Casi actuando como una causa de justificación, el reproche penal pasaría a ser el producto de factores que incluyen entre sí la estimación de la máquina acerca de la producción (o no) de un hecho delictivo. Las mismas consideraciones que se acaban de formular con respecto a la necesidad de un juicio ‘natural’ fundamentan, *a fortiori*, el rechazo de tal deterioro de la teoría del delito.

4. La incidencia de los sistemas de inteligencia artificial en la culpabilidad como criterio para la determinación de la pena en concreto

La segunda cara de la moneda ‘culpabilidad’ tiene que ver con el ámbito de la determinación de la pena en concreto⁸⁶. Desde esta perspectiva, como ya se ha adelantado, por un lado la inteligencia artificial pondría en peligro la operatividad de la proporcionalidad de la pena, al prever mecanismos que podrían favorecer la aplicación de sanciones fijas, y por otro lado, puede ocurrir que la máquina, al aplicar la pena en lugar del juez persona física, genere una condena injusta o arbitraria, o bien sea dictada como consecuencia de influencias sobornadoras externas.

Este último caso, antes de convertir la máquina en la autora de un delito

⁸⁵ Cfr. CASTILLO FELIPE, TOMÁS TOMÁS, *op. ult. cit.*, § 2.4.

⁸⁶ Cfr. las consideraciones críticas efectuadas *supra*, *sub* notas 63 y 64.

de cohecho o prevaricación, no representa sino un ulterior argumento de apoyo a los razonamientos sobre la inoportunidad de que los sistemas de inteligencia artificial sustituyan al juez persona física. En el marco de una intervención débil de las máquinas como instrumentos de apoyo al ejercicio de la justicia, no surgirían estos problemas. El único culpable de haber dictado una resolución injusta o ilícita seguiría siendo el juez, de acuerdo con las leyes que regulan su responsabilidad penal.

En cambio, un ataque más evidente se dirige al principio de proporcionalidad, cuyo fundamento constitucional le impide ser afectado por el uso de máquinas en el ejercicio de la justicia. Estas, quebrantando los matices de la discrecionalidad judicial clásica, no lograrían tomar en consideración todos los segmentos que constituyen un determinado marco edictal, llevando en última instancia a la aplicación de penas fijas, de por sí incompatibles con el ordenamiento jurídico-penal vigente. Para analizar este fenómeno, es suficiente pensar en el funcionamiento de los algoritmos que estarían detrás del cálculo de las penas operado por los sistemas de inteligencia artificial. Cada uno de los factores que intervienen en la averiguación de la verdad judicial conllevaría, en términos de porcentaje o unidad mínima (un día), el aumento o la disminución de pena empezando por el mínimo establecido para el caso específico.

Pese a que un listado exhaustivo de 'factores' (o circunstancias) a los que se debe atribuir un determinado efecto (aumento o rebaja) penológico es materialmente inconcebible, la conminación final de la pena sería el producto de operaciones matemáticas llevadas a cabo por el algoritmo. Para ejemplificar, habría grupos de circunstancias favorables, que llevarían aparejada una modificación (positiva) de la pena, así como grupos de factores desfavorables, con la opuesta repercusión (negativa) en la sanción a aplicar. Por muy detallado que sea el catálogo de las circunstancias 'matemáticas' teorizado e introducido en el algoritmo, siempre existirían segmentos de la horquilla que, aun formando parte de la pena en abstracto, pasarían a ser evanescentes en concreto.

Así pues, se generaría una casuística cerrada (presencia de un factor desfavorable: modificación en x puntos porcentuales superiores a la pena mínima; presencia de dos factores favorables y uno desfavorable: disminución en y puntos porcentuales; y así sucesivamente) que no es sino una manifestación del automatismo del proceso sancionador. En definitiva, acoger este mecanismo perverso significaría instrumentalizar la 'popularidad' de una peligrosa seguridad jurídica celando, al mismo tiempo, el sacrificio de ineludibles garantías constitucionales frente al poder punitivo del Estado.

5. Conclusiones

Las breves reflexiones desarrolladas en estas páginas permiten formular las siguientes conclusiones:

- i) La inteligencia artificial se ha desarrollado hasta llegar a influir no sólo en la realidad social, sino también en las instituciones básicas del Derecho penal. La implementación de máquinas capaces de autodeterminarse y, en cierta medida, de actuar independientemente de la voluntad humana, ha generado la exigencia de replantear la regulación jurídica frente a los posibles ataques sufridos y causados por estos sistemas inanimados.
- ii) Al analizar la influencia de la inteligencia artificial en el Derecho y, más en detalle, en el Derecho penal, es imprescindible tomar como referencia la definición de delito que hace hincapié en los mandatos constitucionales, para evitar el menoscabo de los derechos y las libertades fundamentales, conquistas del Estado de Derecho contemporáneo: «possiamo definire il reato come un fatto previsto in forma tassativa dalla legge, di realizzazione esclusiva dell'agente o in ogni caso al medesimo riconducibile tramite un atteggiamento colpevole (doloso o colposo), idoneo ad offendere un valore costituzionalmente significativo, minacciato con una pena proporzionata anche alla significatività del valore tutelato e strutturalmente caratterizzato dal teleologismo costituzionalmente attribuito alla sanzione penale e, infine, intollerante rispetto ad ogni articolazione probatoria che faccia in qualche modo ricadere sull'imputato l'onere della prova (o il rischio della mancata allegazione) di elementi d'ordine positivo che ne caratterizzano la struttura»⁸⁷.
- iii) La influencia de los sistemas de inteligencia artificial se manifiesta ante todo en la teoría del delito, según el sistema inanimado aparezca como potencial víctima o autor de un hecho delictivo, o bien como instrumento de apoyo para el ejercicio de la justicia.
- iv) Al intentar conceder subjetividad pasiva a los sistemas de inteligencia artificial, surgen las dudas teleológicas con respecto a la aplicación de la sanción correspondiente al autor del hecho. No parece conveniente anticipar la intervención penal hasta las conductas dirigidas a los sistemas inanimados, concretamente no lesivas de bienes jurídicos pertenecientes a los seres humanos. La finalidad educativa que se perseguiría con dicha intervención penal se enmarcaría en un Derecho penal de la voluntad o de la actitud interna, que sin embargo son incompatibles con el princi-

⁸⁷ BRICOLA, *op. et loc. ult. cit.*

pio del hecho, impuesto por los mandatos constitucionales.

- v) En cambio, parece oportuno conceder a los sistemas de inteligencia artificial subjetividad activa, si bien a través de una *fictio iuris* que supone la renuncia a algunas exigencias dogmáticas básicas. Habida cuenta de la creciente autonomía que el *machine learning* les va proporcionando, resulta necesario tutelar al ser humano frente a los ataques realizables por estos sistemas inanimados. La hipótesis de una responsabilidad directa de las máquinas con las herramientas del Derecho actualmente vigente supondría la implementación de una culpabilidad ‘deshumana’, que erigiría su estructura aceptando postulados científicos ajenos al Derecho penal y todavía envueltos por mucha ambigüedad. Por tanto, se propone la creación de un modelo alternativo de responsabilidad penal de los sistemas de inteligencia artificial, basado en una ficción jurídica ‘des-humanizada’ que permita renunciar incluso a las más elementales exigencias del delito ‘humano’, hasta llegar a posibilitar el único supuesto de responsabilidad objetiva que la Constitución esté dispuesta a tolerar.
- vi) La utilización de la inteligencia artificial como instrumento de apoyo para el ejercicio de la justicia conlleva el peligro de alterar la estructura del principio *iudex peritus peritorum*, cuyo alcance fundamenta la libre valoración de la prueba en el proceso penal. Ante la amenaza de dicha alteración, es oportuno que la inteligencia artificial siga siendo utilizada bajo el control, la supervisión y la responsabilidad del juez persona física.
- vii) Con respecto a los efectos de la inteligencia artificial en la culpabilidad entendida como criterio para la determinación de la pena en concreto, se rechaza íntegramente la creación de algoritmos que, de algún modo, puedan quebrantar el contenido del principio de proporcionalidad, constitucionalmente garantizado frente a la arbitrariedad de las decisiones legislativas y judiciales.

6. Bibliografía

- AGUADO CORREA T., *Principio de inexigibilidad de otra conducta en las categorías del delito*, en «Revista Digital de la Maestría en Ciencias Penales de la Universidad de Costa Rica», 3/2011.
- ALCÁCER GUIRAO R., *Los fines del Derecho Penal. Una aproximación desde la filosofía política*, en «Anuario de Derecho Penal y Ciencias Penales», 1/1998.

- ALEX Y R., *Teoría de la argumentación jurídica. La teoría del discurso racional como teoría de la fundamentación jurídica*, Centro de Estudios Políticos y Constitucionales, Madrid, 1989, traducción de M. Atienza e I. Espejo.
- ALONSO SALGADO C., *Acerca de la inteligencia artificial en el ámbito penal: especial referencia a la actividad de las fuerzas y cuerpos de seguridad*, en «Ius et Scientia», 1/2021.
- ARIAS P.C., *Artificial Intelligence & Machine Learning: a model for a new judicial system?*, en «Revista Internacional Jurídica y Empresarial», 2020.
- BAGNOLI C., *Teoria della responsabilità*, Il Mulino, Bolonia, 2019.
- BASILE F., *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, en «Diritto Penale e Uomo», 10/2019.
- BASILE F., *L'alternativa tra responsabilità oggettiva e colpa in attività illecita per l'imputazione della conseguenza ulteriore non voluta, alla luce della sentenza Ronci delle Sezioni Unite sull'art. 586 c.p.*, en «Rivista italiana di diritto e procedura penale», 3/2011.
- BECCARIA C., *Tratado de los delitos y de las penas*, Heliasta, Buenos Aires, 1993.
- BERTOLINO M., *Declinazioni attuali della pericolosità sociale: pene e misure di sicurezza a confronto*, en «Archivio penale», 2/2014.
- BRICOLA F., voz *Teoria generale del reato*, en «Novissimo digesto italiano», XIV, 1973.
- CAPPELLINI A., *Machina delinquere non potest? Brevi appunti su intelligenza artificiale e responsabilità penale*, en «Criminalia», 2018.
- CASTILLO FELIPE R., TOMÁS TOMÁS S., *Proceso penal e inteligencia artificial: prevenciones en torno a su futura aplicación en la fase de juicio oral*, en esta obra.
- CAVALIERE A., *Misure di sicurezza e doppio binario. Considerazioni introduttive*, en «Rivista italiana di diritto e procedura penale», 1/2022.
- COBO DEL ROSAL M., *Societas delinquere non potest*, en «Anales de derecho», 2012.
- COBO DEL ROSAL M., VIVES ANTÓN T.S., *Derecho penal. Parte general*, Tirant lo Blanch, Valencia, 1999.
- CONSULICH F., *Il nastro di Mobius. Intelligenza artificiale e imputazione penale nelle nuove forme di abuso del mercato*, en «Banca Borsa Titoli di Credito», 2/2018.
- CORCOY BIDASOLO M., *Exigibilidad en el ámbito del conocimiento y control de riesgos: teorización*, en «Revista catalana de seguretat pública», 13/2003.

- CORCOY BIDASOLO M., *Delitos de peligro y protección de bienes jurídico-penales supraindividuales*, Tirant lo Blanch, Valencia, 1999.
- DANAHER J., *Robotic Rape and Robotic Child Sexual Abuse: Should They be Criminalised?*, en «Criminal Law and Philosophy», 1/2017.
- DE LA CUESTA AGUADO P.M., *Inteligencia artificial y responsabilidad penal*, en «Revista penal México», 2020.
- FIANDACA G., MUSCO E., *Diritto penale. Parte generale*, Zanichelli, Bolonia, 2019.
- FORNARI L., *Misure di sicurezza e doppio binario: un declino inarrestabile?*, en «Rivista italiana di diritto e procedura penale», 1993.
- GALLO M., *Diritto penale italiano. Appunti di parte generale*, vol. I, Giappichelli, Torino, 2021.
- GALLO M., *Il concetto unitario di colpevolezza*, Giuffrè, Milán, 1951.
- GALLO M., *La teoria dell'azione «finalistica» nella più recente dottrina tedesca*, Giuffrè, Milán, 1950.
- GAMERO CASADO E., *El enfoque europeo de inteligencia artificial*, en «Revista de Derecho Administrativo», 2021.
- GARDNER G., *Intelligence Reframed. Multiple Intelligences for the 21st Century*, Basic Books, New York, 2000.
- GIALUZ M., *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, en «Diritto penale contemporaneo», 29 de mayo de 2019.
- GIRALDI A., *Algorithms and Big Data Towards a Crime-Preventing Groupware*, en «Roma Tre Law Review», 2/2021.
- GIRALDI A., *Política, discrecionalidad y Derecho en las implicaciones empíricas del principio de proporcionalidad de la pena*, en «Revista Internacional de Doctrina y Jurisprudencia», 22/2020.
- GRACIA MARTÍN L., *Crítica de las modernas construcciones de una mal llamada responsabilidad penal de la persona jurídica*, en «Revista Electrónica de Ciencia Penal y Criminología», 2016.
- HALLEVY G., *The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control*, en «Akron Intellectual Property Journal», 2/2010.
- HASSEMER W., *¿Por qué castigar? Razones por las que merece la pena la pena*, Tirant lo Blanch, Valencia, 2016, traducción de M. Cancio Meliá y F. Muñoz Conde.
- HEFENDEHL R., HIRSCH A.V., WOHLERS W. (eds.), *La teoría del bien jurídico. ¿Fundamento de legitimación del Derecho penal o juego de abalorios dogmático?*, Marcial Pons, Madrid, 2016, traducción de R.A. Guirao, M. Martín Lorenzo, Í. Ortiz de Urbina Gimeno.

- JAKOBS G., *La autocomprensión de la ciencia del Derecho penal ante los desafíos del presente*, en *La ciencia del Derecho penal ante el nuevo milenio*, dirigido por A. Eser, W. Hassemmer, B. Burkhardt, Tirant lo Blanch, Valencia, 2004.
- JAKOBS G., *El principio de culpabilidad*, en «Anuario de Derecho Penal y Ciencias Penales», 3/1992.
- JAKOBS G., CANCIO MELIÁ M., *Derecho penal del enemigo*, Civitas, Madrid, 2003.
- JESCHECK H.H., *Evolución del concepto jurídico penal de culpabilidad en Alemania y Austria*, en «Revista Electrónica de Ciencia Penal y Criminología», 2003.
- KAPLAN A., HAENLEIN M., *Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence*, en «Business Horizons», 1/2019.
- KAPLAN J., *Artificial Intelligence. What Everyone Needs to Know*, Oxford University Press, Oxford, 2016.
- LAMPE E.J., *Das personale Unrecht*, Duncker & Humblot, Berlín, 1967.
- LIMA D., *Could AI Agents Be Held Criminally Liable? Artificial Intelligence and the Challenges for Criminal Law*, en «South Carolina Law Review», 3/2018.
- LLEDÓ BENITO I., *El impacto de la robótica. La inteligencia artificial y la responsabilidad penal en los robots inteligentes*, en «Foro galego. Revista jurídica», 2020.
- MANZINI V., *Trattato di diritto penale italiano*, vol. III, Utet, Torino, 1981.
- MARAS M.H., SHAPIRO L.R., *Child Sex Dolls and Robots: More Than Just an Uncanny Valley*, en «Journal of Internet Law», 2017.
- MARINI G., voz *Colpevolezza*, en «Digesto delle Discipline Penalistiche», II, 1988.
- MARINUCCI G., DOLCINI E., GATTA G.L., *Manuale di Diritto Penale. Parte Generale*, Giuffrè, Milán, 2018.
- MASSARO A., *Pericolosità sociale e misure di sicurezza detentive nel processo di "definitivo superamento" degli ospedali psichiatrici giudiziari: la lettura della Corte costituzionale con la sentenza n. 186 del 2015*, en «Archivio penale», 2/2015.
- MCCARTHY J., MINSKY M.L., ROCHESTER N., SHANNON C.E., *A proposal for the Dartmouth summer research project on artificial intelligence*, 31 de agosto de 1955, disponible en <<http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>> (consultado el 18 de noviembre de 2022).
- MENCARELLI A., *Le misure di sicurezza detentive al vaglio della Corte europea dei diritti dell'uomo*, en *La tutela della salute nei luoghi di detenzione. Un'indagine di diritto penale intorno a carcere, REMS e CPR*, dirigido por A. Massaro, RomaTrE-Press, Roma, 2017.

- MIR PUIG S., *Derecho penal. Parte general*, Editorial Reppertor, Barcelona, 2011.
- MIR PUIG S., *Función de la pena y teoría del delito en el Estado social y democrático de Derecho*, Bosch, Barcelona, 1982.
- MIR PUIG S., *Preterintencionalidad y límites del art. 50 del Código Penal*, en «Revista Jurídica de Cataluña», 1979.
- MORILLAS CUEVA L., *Sistema de Derecho Penal. Parte general*, Dykinson, Madrid, 2018.
- MORILLAS FERNÁNDEZ D.L., *Implicaciones de la inteligencia artificial en el ámbito del Derecho Penal*, en esta obra.
- MUÑOZ CONDE F., GARCÍA ARÁN M., *Derecho Penal. Parte General*, Tirant lo Blanch, Valencia, 2019.
- MUÑOZ CONDE F., *Introducción*, en C. ROXIN, *Política criminal y sistema del derecho penal*, Hammurabi, Buenos Aires, 2002, traducción de F. Muñoz Conde.
- MUÑOZ VELA J.M., *Inteligencia Artificial y responsabilidad penal*, en «Derecho Digital e Innovación. Digital Law and Innovation Review», 11/2022.
- MUSCO E., voz *Misure di sicurezza*, en «Enciclopedia del diritto», I, 1997.
- NAVARRETE URIETA J.M., *El nacimiento de la idea de culpabilidad*, en «Aldaba», 18/1992.
- NIETO MARTÍN A., *Problemas fundamentales del cumplimiento normativo en el derecho penal*, en *Compliance y teoría del Derecho penal*, dirigido por L. Kuhlen, J.P. Montiel, Í. Ortiz de Urbina Gimeno, Marcial Pons, Madrid, 2013.
- NOTARO L., *Intelligenza artificiale e giustizia penale*, en *Intelligenza artificiale e giustizia penale*, dirigido por A. Massaro, Paruzzo Editore, Caltanissetta, 2020.
- NUVOLONE P., voz *Misure di prevenzione e misure di sicurezza*, en «Enciclopedia del diritto», XXVI, 1976.
- OTTO H., *Personales Unrecht, Schuld und Strafe*, en «Zeitschrift für die gesamte Strafrechtswissenschaft», 3/1975.
- PÉREZ ARIAS J., *Algoritmos y big data en la responsabilidad penal: el reto de la cibercriminalidad en el Derecho Penal*, en esta obra.
- PÉREZ ARIAS J., *Sistema de atribución de responsabilidad penal a las personas jurídicas*, Dykinson, Madrid, 2014.
- PÉREZ MANZANO M., *Culpabilidad y prevención. Las teorías de la prevención general positiva en la fundamentación de la imputación subjetiva y de la pena*, Universidad Autónoma de Madrid, Madrid, 1990.

- PERIS RIERA J.M., *Delitos de peligro y sociedad de riesgo: una constante discusión en la dogmática penal de la última década*, en *Estudios penales en homenaje al profesor Cobo del Rosal*, dirigido por J.C. Carbonell Mateu, B. del Rosal Blasco, L. Morillas Cueva, E. Orts Berenguer, M. Quintanar Díez, Dykinson, Madrid, 2005.
- QUINTERO OLIVARES G., *Parte General del Derecho Penal. Adaptada al programa de ingreso en las carreras judicial y fiscal*, Aranzadi, Pamplona, 2015.
- QUINTERO OLIVARES G., *Las vicisitudes del dolo y la subsistencia de la preterintencionalidad*, en *Constitución, derechos fundamentales y sistema penal. Semblanzas y estudios con motivo del setenta aniversario del profesor Tomás Salvador Vives Antón*, dirigido por J.C. Carbonell Mateu, J.L. González Cussac, E. Orts Berenguer, Tirant lo Blanch, Valencia, 2009.
- RIONDATO S., *Robotica e diritto penale (robots, ibridi, chimere e "animali tecnologici")*, en *Genetics, robotics, law, punishment*, dirigido por D. Provolo, S. Riondato, F. Yenisey, Padova University Press, Padua, 2014.
- ROXIN C., *Culpabilidad y prevención en Derecho penal*, Editorial Reus, Madrid, 1981, traducción de F. Muñoz Conde.
- ROXIN C., *Derecho Penal. Parte General. Tomo I. Fundamentos. La estructura de la teoría del delito*, Civitas, Madrid, 1997, traducción de D.M. Luzón Peña, M. Díaz y García Conlledo, J. de Vicente Remesal.
- SANTAMARIA D., voz *Colpevolezza*, en «Enciclopedia del diritto», VII, 1960.
- SCHANK R.C., *What Is AI, Anyway?*, en «AI Magazine», 4/1987.
- SOLUM L.B., *Legal Personhood for Artificial Intelligences*, en «North Carolina Law Review», 4/1992.
- SPINOZA B., *Tratado político*, Alianza Editorial, Madrid, 1986, traducción de Atilano Domínguez.
- SQUILLACI E., *'Ragioni' e 'costi' dell'attuale paradigma di prevenzione*, en «Archivio penale web», 2/2020.
- TAGLIARINI F., voz *Pericolosità*, en «Enciclopedia del diritto», XXXIII, 1983.
- TORIO LÓPEZ Á., *El concepto individual de culpabilidad*, en «Anuario de Derecho Penal y Ciencias Penales», 2/1985.
- TORRAS GENÍS C., DELICADO MORATALA L., *Blanco y Frío como El Polo Norte y El Polo Sur: Seres humanos y robots*, en «Sociología y tecnociencia: Revista digital de sociología del sistema tecnocientífico», 1/2022.
- VALLINI A., *Antiche e nuove tensioni tra colpevolezza e diritto penale artificiale*, Giappichelli, Torino, 2003.
- VALLS PRIETO J., *Inteligencia artificial, Derechos humanos y bienes jurídicos*, Aranzadi, Pamplona, 2021.

VELÁZQUEZ FERNÁNDEZ H., *¿Qué tan natural es la inteligencia artificial? Sobre los límites y alcances de la biomímesis computacional*, en «Naturaleza y Libertad. Revista de estudios interdisciplinarios», 12/2019.