

## CAPÍTULO III

### *Implicaciones de la inteligencia artificial en el ámbito del Derecho Penal*

David Lorenzo Morillas Fernández

SUMARIO: 1. Enfoque y estado de la cuestión – 2. Situación actual de la responsabilidad penal por un hecho delictivo cometido por inteligencia artificial – 2.1. ¿Puede considerarse a la inteligencia artificial autora única del delito? – 2.2. ¿Quién debe responder penalmente entonces? – 3. Pronunciamientos jurisprudenciales expresos – 4. Conclusiones – 5. Bibliografía.

#### *1. Enfoque y estado de la cuestión*

Aunque a muchos les pese, la forma más sencilla y, en mi opinión, recomendada de comenzar cualquier trabajo, no debe ser otra que contextualizar el ámbito en la que el autor pretende moverse y eso, en determinadas cuestiones, como la que me ocupa, no es sencillo dada la complejidad y aristas de la temática a tratar, máxime cuando el potencial lector no tiene por qué entender o saber inicialmente el concreto ámbito o las especificidades de la temática a abordar. Dar por supuestos determinados conceptos, contenidos e implicaciones resulta adecuado cuando el destinatario de la información es un público especializado en la materia si bien lo aquí buscado choca diametralmente con semejante afirmación, reuniendo el actual artículo la simple y mera finalidad de resultar accesible a juristas en general, para lo cual se van a ir confrontando y exponiendo ideas desde abajo para ir progresivamente avanzando en intensidad hasta alcanzar un nivel medio.

Así las cosas, la mejor forma de ejemplificar la inteligencia artificial, por mor de que a los más puristas en la materia le resulte una vaguedad o eufemismo, debe partir necesariamente de ejemplos concretos comprensibles para el conjunto de la ciudadanía. En este sentido, me permito traer a colación, dos claros ejemplos conocidos – creo – por todos: i) la serie de televisión de los años ochenta, «El coche fantástico», donde un hombre luchaba contra el

crimen junto con un coche que poseía habilidades y destrezas propias que le permitían realizar funciones de manera completamente independiente y siempre bajo un determinado prisma: proteger la vida humana, siendo capaz de desarrollar conductas ilícitas -básicamente daños- por sí mismo siempre que el fin fuera la protección de la vida<sup>1</sup>; o ii) la película «Terminator» – y sus diversas secuelas y precuelas – donde un sistema de inteligencia artificial, llamado Skynet, lidera un ejército de máquinas poseyendo la capacidad para controlar el arsenal militar de los Estados Unidos con independencia de la voluntad humana. En consecuencia, si se tomaran de ejemplo determinadas conductas visionadas en la anterior serie y saga, surgen preguntas relacionadas con la responsabilidad jurídica, ya sea penal o extra-penal, como ¿quién respondería cuándo Michael Knight realiza una persecución y conduce a más de 200 kms/h? Parece claro que el protagonista cometería un delito contra la seguridad vial del artículo 379.1 del Código Penal; pero ¿y si fuera KITT, el coche, quien hace uso de su autonomía para realizar una persecución sin conductor a más de 200 kms/h? ¿a quién atribuir la responsabilidad penal cuando, en alguna de las secuencias más famosas de la serie, KITT realiza una conducción autónoma y, haciendo uso de su turbo, salta y atraviesa una pared de ladrillo causando unos evidentes daños en la propiedad o, con su salto, destroza un enorme espejo que está siendo trasladado en un camión de grandes dimensiones? ¿Cambiaría la calificación de los hechos si Michael pidiera ayuda por su transmisor por ver peligrar su vida o salud y, como quiera que KITT está diseñado para preservar la vida, desarrolla tales comportamientos? ¿Podría aplicarse una causa de justificación a la conducta ilícita de una inteligencia artificial? Del mismo modo, yendo a parámetros más extremos, ¿de quién sería la responsabilidad penal si un humano crea un robot capaz de matar? Parece claro, al igual que en el primer ejemplo anterior, que de la persona, siendo la máquina simplemente el instrumento de ejecución del delito pero ¿y si fuera otra máquina -por ejemplo, Skynet, por seguir la representación- quien, de conformidad con su capacidad de autoaprendizaje, crea un robot (Terminator) que mata a personas (como en las distintas secuelas de la saga)? ¿Tendría relevancia el hecho de que la máquina no hace otra cosa que cumplir con el objetivo para el cual ha sido programada -acabar con la vida de John Connor-?

---

<sup>1</sup> Conviene llamar la atención en este punto en el hecho de que, al igual que sucede en la realidad, todo invento orientado a la mejora de la sociedad o de su calidad de vida, más pronto que tarde presenta también implicaciones ilícitas y así, en la citada serie, existía otro modelo -KARR- carente del sentido de protección de la vida y antítesis de KIT, por lo que el desarrollo de hechos ilícitos, representado por sentimientos humanos como la ira o la venganza, operaban en él.

Los ejemplos anteriores -y otros muchos- que en la década de los ochenta podían parecer auténtica ciencia-ficción no dejan de ser una realidad completa, con las lógicas limitaciones propios de la ficción en pleno año 2022, donde la sociedad se mueve al ritmo que le marca la tecnología en multitud de ámbitos. Piénsese, a tal efecto, en la propia conducción de vehículos, donde no sólo ya hay sistemas automáticos para aparcar, frenar automáticamente en caso de emergencia, detectar la salida de la calzada y proceder inmediatamente a la protección de los ocupantes tensando los cinturones de seguridad y activando un sistema de absorción de impactos en los asientos con el propósito de proteger la columna vertebral de los usuarios (...) sino igualmente vehículos conducidos autónomamente -vid., por ejemplo, el modelo Tesla, que saca el coche del garaje y lo deja en la calzada, o, yendo más allá, permite la conducción autónoma por vía pública<sup>2</sup>-, una ingeniería que cada día va a más y recibe el nombre de vehículo autónomo o robótico. Pero es más, ¿qué sucede en el ámbito de la medicina? Los avances de la inteligencia artificial en el diagnóstico y tratamiento de enfermedades son una constante diaria, lo que permite una mejora de la calidad de vida de la población. ¿Y su utilización con fines bélicos? No se trata ya de robots desactivadores de bombas que funcionan con control remoto sino de máquinas con capacidad de lucha autónoma -robots de combate con inteligencia artificial; vehículos no tripulados para transportar armas pesadas, donde la inteligencia artificial valora situaciones de combate y toma decisiones autónomas; el URAN-9, vehículo de combate destinado, entre otras funciones, a destruir otros vehículos blindados- (...).

Con semejantes planteamientos no es de extrañar que, como refirió Pérez Arias, en la era de la ciber-civilización, la sociedad muestre su preocupación por conocer el nivel de protección que el Derecho confiere a víctimas de delitos cometidos a través de un medio tecnológico. Y dentro de esta protección, es aún más lógico que la sociedad se cuestione (y con mayor motivo los intérpretes) si el Derecho Penal, ha sido debidamente actualizado y modernizado hasta el punto de otorgar protección frente a determinados delitos que discurren por medios, y fines, por completo desconocidos, hasta fechas muy recientes<sup>3</sup>.

---

<sup>2</sup> Aun cuando hoy día no está recomendado ya que la activación de esta función exige la supervisión por parte del conductor -exigencia interesante por parte de la compañía que, llegado el caso, pudiera eximirle de responsabilidad penal en caso de atropello, como se planteará más adelante-, no son pocos los vídeos que pueden encontrarse en redes sociales donde los potenciales conductores del vehículo graban la conducción autónoma del Tesla por carretera.

<sup>3</sup> J. PÉREZ ARIAS, *Cibercriminalidad: hacia la nueva realidad -virtual- del Derecho Penal*,

Como ha podido observarse, hoy día, la inteligencia artificial juega un papel nuclear en el devenir cotidiano de la ciudadanía con las consiguientes implicaciones jurídicas, entre otras muchas. Ahora bien, si se observa la casuística anterior, no todo responde en términos absolutos a inteligencia artificial ya que, aún dentro de ella, se establecen categorías. En este sentido, Salvadori ha distinguido entre inteligencia artificial y aprendizaje automático como parámetros delimitadores de los llamados agentes artificiales -utilizados para identificar los software- y sistemas de inteligencia artificial -capaces de desarrollar funciones para otro agente, ya sea humano o artificial-, los cuales están apoyando o sustituyendo a las personas en la realización de actividades, siendo en este ámbito donde cabría incluir los ejemplos anteriores vinculados al sector médico-quirúrgico, asistencia a enfermos o ancianos, coches autónomos o algoritmos de justicia predictiva; esto es, no se habla ya de meros instrumentos u objetos sino de verdaderos agentes artificiales capaces de interactuar con la realidad circulante, con otras máquinas inteligentes y seres humanos, siendo capaces de tomar en muy poco tiempo decisiones complejas de indudable relevancia social<sup>4</sup>.

Cuatrecasas Monforte ha clasificado a la inteligencia artificial, entre otros criterios, según el procedimiento en virtud del cual lleva a cabo el aprendizaje, distinguiendo entre aprendizaje supervisado, no supervisado y por refuerzo: i) el supervisado se caracterizaría por la presencia de un programador humano que introduce datos etiquetados y establece distintas variables de entrada en función de las cuales deberá asignarse una etiqueta de salida, mostrando el algoritmo las conclusiones a las que debe llegar; ii) no supervisado, se identificaría por la ausencia de un programador que introduzca datos y parámetros de identificación, siendo el propio algoritmo el que detecta la repetición de patrones y lleva a cabo un auto entrenamiento y aprendizaje, sacando sus propias conclusiones sin indicación externa alguna; y iii) por refuerzo, se basaría en la máxima prueba-error, existiendo una interacción permanente con el sistema, mediante el envío de mensajes notificaciones que le informan de si es correcto o incorrecto lo que va realizando, pudiendo extraer conclusiones tras múltiples valoraciones y determinar qué caminos son los que con más probabilidad llevan al éxito<sup>5</sup>.

La velocidad de los avances tecnológicas es tan creciente e intensa que no es de extrañar la afirmación recogida en la Resolución del Parlamento

---

en «Revista Electrónica de Doctrina y Jurisprudencia», n. 26, 2021, pp. 175 y 176.

<sup>4</sup> I. SALVADORI, *Agentes artificiales, opacidad tecnológica y distribución de la responsabilidad penal*, en «Cuadernos de Política Criminal», n. 133, 2021, pp. 138 y 139.

<sup>5</sup> C. CUATRECASAS MONFORTE, *La Inteligencia Artificial como herramienta de investigación criminal*, La Ley, Madrid, 2022, pp. 32-36.

Europeo, de 3 de mayo de 2022, sobre la inteligencia artificial en la era digital (2020/2266(INI)), la cual habla de que estamos ante una cuarta revolución industrial, caracterizándose ésta, en comparación con las anteriores, iniciadas por la introducción del vapor, la electricidad y luego los ordenadores, porque la cuarta oleada extrae su energía de la abundancia de datos combinada con potentes algoritmos y capacidad informática, expansiva a nivel mundial, con una rápida convergencia y el enorme impacto de los nuevos avances tecnológicos en los países, las economías, las sociedades, las relaciones internacionales y el medio ambiente, lo cual supone un cambio radical de enorme magnitud que repercute diversamente en las distintas partes de la sociedad, en función de sus objetivos, ubicación geográfica o contexto socioeconómico. Ahora bien, semejante transformación digital debe hacerse siempre dentro del pleno respeto de los derechos fundamentales y con el único propósito de que las tecnologías digitales sirvan a la humanidad.

Sea como fuere, conviene tener unas cautelas mínimas a la hora de abordar la problemática en el uso y la regulación de sistemas de inteligencia artificial. En este sentido, 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 (2020/2016(INI)), ha establecido unos considerandos muy útiles que delimitan los parámetros mínimos a tener en cuenta al respecto y que conviene resaltar:

- i) Las tecnologías digitales en general y la proliferación del tratamiento y el análisis de datos por medio de inteligencia artificial ofrecen enormes riesgos para los derechos fundamentales y las democracias basadas en el Estado de Derecho ya que la inteligencia artificial no debe considerarse un fin en sí misma sino un instrumento al servicio de las personas, con el objetivo primordial de aumentar el bienestar humano, las capacidades humanas y la seguridad.
- ii) No existen programas que puedan igualar a la flexibilidad humana en ámbitos más amplio o tareas que exijan la comprensión del contexto o un análisis crítico. Algunas aplicaciones de inteligencia artificial han alcanzado los niveles de rendimiento de expertos y profesionales humanos – por ejemplo, la tecnología jurídica o ‘legal tech’ –, pudiendo ofrecer resultados a una velocidad muchísimo mayor y una escala más amplia.

- iii) Algunos países hacen uso de aplicaciones o sistemas de inteligencia artificial con fines coercitivos y judiciales debido a la ausencia de regulación y las diferencias normativas. Eso hace que el creciente uso de la inteligencia artificial en el ámbito del Derecho Penal se base en particular en las promesas de que reduciría determinados tipos de delincuencia y daría lugar a decisiones más objetivas. Sin embargo, estas promesas no siempre se cumplen.
- iv) Para que los sistemas de inteligencia artificial resulten fiables, deben diseñarse con el propósito de proteger y beneficiar a todos los miembros de la sociedad, particularmente a las personas vulnerables y marginadas, no debiendo ser, en ningún caso, discriminatorias y tomar decisiones explicables y transparentes, respetando siempre la autonomía humana y los derechos fundamentales.
- v) El uso de aplicaciones de inteligencia artificial ofrece grandes oportunidades para mejorar los métodos de trabajo de las autoridades policiales y judiciales, aumentando la eficacia en la lucha contra determinados tipos de delitos, especialmente los de naturaleza financiera -blanqueo de capitales y financiación del terrorismo-, abusos sexuales, explotación sexual en línea y determinados tipos de ciberdelincuencia. No obstante, también pueden entrañar al mismo tiempo riesgos significativos para los derechos fundamentales de las personas.
- vi) Se hace imprescindible contar con un modelo claro para establecer la responsabilidad penal por los posibles efectos nocivos de los sistemas de inteligencia artificial.

Así las cosas, para concretar el conjunto de considerandos anteriores, resulta fundamental partir de una definición concreta de lo que se entiende por inteligencia artificial, algo que no va a resultar sencillo ya que existen multitud de aristas que dificultan una concreción terminológica unitaria al respecto y así, como refirió Miró Llinares, la premisa inicial resulta por sí compleja debido a las disquisiciones filosóficas no sólo respecto del concepto inteligencia sino porque además según el campo en el que nos movamos los objetivos definitorios serán igualmente modulables<sup>6</sup>. En términos muy parecidos se ha manifestado De la Cuesta Aguado al señalar que el concepto

<sup>6</sup> F. MIRÓ LLINARES, *Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots*, en «Revista de Derecho Penal y Criminología», n. 20, 2018, pp. 91-92.

de inteligencia, como característica propia del género humano, dista mucho de ser pacífico, tanto en cuanto a su contenido (¿qué es la inteligencia?) como a su extensión (¿solo son inteligentes los humanos?)<sup>7</sup>. Ahora bien, tal y como ha señalado Palma, la cuestión de la inteligencia artificial y la neurociencia se complica en el instante en el que una decisión inteligente puede ser calculada como un algoritmo e imitada por procesos matemáticos<sup>8</sup>.

Sea como fuere, son muchas y muy variadas las definiciones existentes al respecto, presentándose a continuación algunas, de manera meramente orientativa y próximas al ámbito del Derecho Penal:

- El enfoque inicial de la cuestión lo trae perfectamente a colación Valls Prieto cuando refiere que la inteligencia artificial es considerada como la capacidad de la máquina de pensar, de razonar y actuar con inteligencia; esto es, se trataría de un sistema que puede realizar tareas de manera idéntica a como la realizaría un humano en determinados contextos, si bien existen supuestos en los que semejante tarea designada obtiene un rendimiento mejor que en el caso que la llevaran a efecto personas, tal y como puede ser el ejemplo de la búsqueda de información, reconocimiento de tumores en radiografías o jugando al ajedrez. De acuerdo con lo anterior, se podrían diferenciar dos tipos de inteligencia artificial: una, basada en el conocimiento; y otra, en datos. En definitiva, las máquinas tienen la posibilidad de realizar comportamientos automáticos inteligentes en la medida en que pueden realizarlos de forma parecida a los humanos dentro de un contexto concreto en el que tienen que elegir entre diferentes posibilidades y que, en determinadas facetas de la vida, aportan soluciones útiles para las personas. Ahora bien, no conviene olvidar, siendo este extremo especialmente relevante para los aspectos que referiré posteriormente, que la inteligencia artificial es una herramienta creada por las personas, lo cual lleva consigo que, aunque pueda crear cosas -entre ellas más inteligencias artificiales-, siempre va a ser mediante unos patrones impuestos por los humanos, siendo éste el origen de la actividad que desarrolle y, por tanto, la máquina no va a realizar nada que no sea predeterminado por un humano, resultando esta cuestión nuclear al objeto de determinar la responsabilidad penal operable al respecto<sup>9</sup>.

---

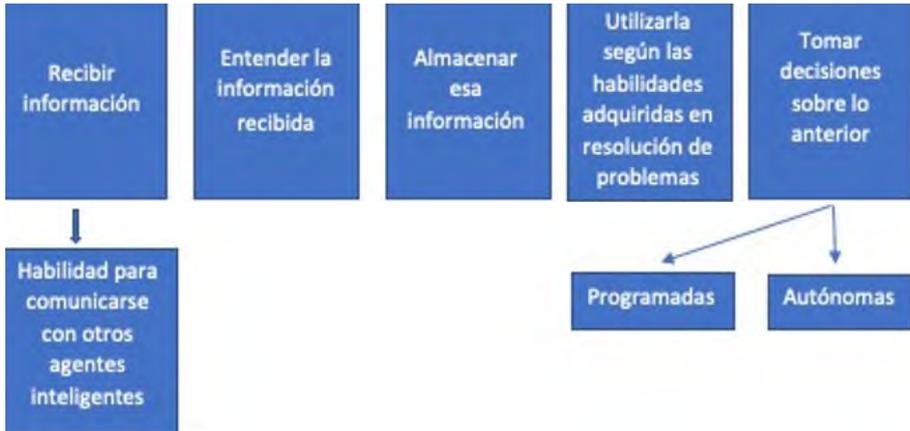
<sup>7</sup> P. DE LA CUESTA AGUADO, *Inteligencia artificial y Derecho Penal*, en «Revista Penal México», n. 16 y 17, 2019-2020, p. 52.

<sup>8</sup> M<sup>a</sup> F. PALMA, *El Derecho como universo de sentido para la interfaz cerebro e inteligencia artificial*, en «Derecho Penal y comportamiento humano», dirigido por E. Demetrio Crespo, Tirant lo Blanch, Valencia, 2022, p. 176.

<sup>9</sup> J. VALLS PRIETO, *Sobre la responsabilidad penal por la utilización de sistemas inteligentes*,

Pero claro, como ya se ha apuntado anteriormente, ¿qué cabe entender por inteligencia? Comparto totalmente lo expresado por De la Cuesta Aguado cuando afirma que la inteligencia es una característica propia del ser humano y no existe una significación concreta al respecto, no siendo tampoco éste el momento para profundizar en cuestiones vinculadas a ella, máxime por los múltiples enfoques científicos que podrían operar en la referida cuestión. Es por ello que, siguiendo los postulados manifestados por la citada autora, y con un mero carácter introductorio o ilustrativo a los efectos de este trabajo, voy a identificar la inteligencia de acuerdo con las capacidades que describió: i) capacidad para recibir información; ii) capacidad para entender la información recibida -entendiendo por tal aquella capaz de hacer utilizable por el agente en el contexto y en función de objetivos la información recibida; iii) capacidad para almacenar la información recibida; iv) capacidad para utilizar la información obtenida y procesada mediante las habilidades anteriores en la resolución de problemas; y v) capacidad para tomar decisiones<sup>10</sup>.

#### ESQUEMA DE LAS CAPACIDADES BASE PARA DEFINIR LA INTELIGENCIA<sup>11</sup>



en «Revista Electrónica de Ciencia Penal y Criminología», 24-27, 2022, pp. 5 y 9.

<sup>10</sup> P. DE LA CUESTA AGUADO, *Inteligencia artificial y Derecho Penal*, cit., p. 52.

<sup>11</sup> Elaboración propia sobre los postulados descritos por DE LA CUESTA AGUADO.

Así las cosas, como manifestaron en su día Corvalán y Ciraudo, con la inteligencia artificial se busca que las tecnologías permitan que los sistemas computacionales adquieran autodependencia, reconfiguración autoadaptativa, negociación inteligente, comportamiento de cooperación, supervivencia con intervención humana reducida, entre otros rasgos. La inteligencia artificial no busca crear robots o computadoras inteligentes ya que resulta imposible reproducir el cerebro humano en máquinas, dada su complejidad y la ausencia de datos fiables que permitan determinar a ciencia cierta su funcionamiento integral, lo que se busca de ella es que realice exitosamente ciertas actividades sin necesidad de que estos sistemas se asemejen imiten o copien a un cerebro humano<sup>12</sup>.

El error, entiende Cuatrecasas Monforte, radica en realizar una comparación directa entre ambas inteligencias, no siendo el mejor criterio para poder llegar a entender qué es la inteligencia artificial ya que, por un lado, las máquinas pueden realizar tareas que las personas no pueden en absoluto y, por otro, los humanos cuentan con una sensibilidad un sentido común y una capacidad de improvisación que las máquinas, al menos por el momento, no llegan a alcanzar. Así, la comparación entre la inteligencia humana y la artificial implicaría la equiparación de dos ideas con contenido no idéntico que podría inducir a error y/o confusión, por lo que la citada autora considera que la mejor opción es otorgar a la inteligencia artificial una definición autónoma<sup>13</sup>.

Por su parte Valls Prieto, quien ha realizado una muy acertada compilación de diversos documentos de la Unión Europea, refiere las conclusiones a las que llega el grupo de expertos de alto nivel en inteligencia artificial de la Comisión Europea y que, muy sucintamente, puede resumirse, a los efectos que ahora interesa, en lo siguiente: i) la reiterada dificultad para definir lo que es inteligencia, ya sea en humanos o en máquinas, por tratarse de un concepto vago, el cual no debe orientarse hacia la inteligencia en sí misma si no derivarlo al concepto de racionalidad, entendiéndose por ésta la habilidad de elegir la mejor opción para conseguir un determinado objetivo, dados determinados criterios y los recursos disponibles; ii) terminológicamente no es correcto hablar de inteligencia artificial per se sino que prefieren utilizar el término sistema de inteligencia artificial para definir a cualquier componente

---

<sup>12</sup> J.G. CORVALÁN, D. CIRAUDO, *Inteligencia artificial al servicio de la justicia penal, contravencional y de faltas. 'Prometea' en el ámbito de la ciudad de Buenos Aires*, en *Cibercrimen II*, dirigido por D. Dupuy, Editorial B de f, Buenos Aires, 2018, p. 269.

<sup>13</sup> CUATRECASAS MONFORTE, *La Inteligencia Artificial como herramienta de investigación criminal*, cit., p. 24.

basado en aquella, ya sea software o hardware<sup>14</sup>. Como puede observarse, el primer considerando rehúye de la utilización del término inteligencia en favor de racionalidad, lo cual traduce en una destreza para lograr un fin específico, resultando necesario unos criterios previamente establecidos al respecto, lo cual anularía la absoluta independencia de la máquina, si bien, como se verá posteriormente, pese a ser semejante postulado claro la situación concreta puede complicarse cuando el agente artificial es capaz de ‘aprender’ sobre semejantes parámetros y recurrir a otros criterios basados en el autoaprendizaje.

- El Grupo de Expertos de Alto Nivel en Inteligencia Artificial de la Unión Europea define la inteligencia artificial como aquellos sistemas de software (y posiblemente también de hardware) diseñados por humanos que, dado un objetivo complejo, actúan en la dimensión física o digital percibiendo su entorno mediante la adquisición de datos, interpretándolos de manera estructurada o no estructurada, razonando sobre el conocimiento, o procesando la información, derivada de ellos y decidiendo las mejores acciones a realizar para alcanzar el objetivo dado<sup>15</sup>.

Así pues, de la citada definición y sus implicaciones pueden comprobarse dos extremos fundamentales que conviene tener muy presentes a la hora de trasladar al ámbito jurídico: i) el origen del sistema de inteligencia artificial se encuentra embrionariamente ligada al ser humano; y ii) como el propio Grupo reconoce expresamente en el citado documento, la inteligencia artificial integra unos sistemas que muestran un comportamiento inteligente analizando su entorno y emprendiendo acciones -con cierto grado de autonomía- para alcanzar objetivos específicos<sup>16</sup>, lo cual, en mi opinión,

---

<sup>14</sup> VALLS PRIETO, *Sobre la responsabilidad penal por la utilización de sistemas inteligentes*, cit., p. 8.

<sup>15</sup> INDEPENDENT HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *A definition of AI: Main capabilities and disciplines*, European Commission, Bruselas, 2019, p. 6.

<sup>16</sup> *Ibidem*, p. 1. Pero es más, incluso en los procesos de aprendizaje automático, como puede ser el caso del aprendizaje profundo, la red neuronal tiene varias capas entre la entrada y la salida que permiten aprender la relación global de entrada-salida en pasos sucesivos, logrando que el enfoque general sea más preciso y con menos necesidad de orientación humana; o el aprendizaje por refuerzo, donde se deja que el sistema de inteligencia artificial tome libremente sus decisiones, a lo largo del tiempo, y en cada decisión se le proporciona una señal de recompensa que le indica si ha sido una buena o una mala decisión con el propósito de que el objetivo del sistema, a lo largo del tiempo, sea maximizar la recompensa positiva recibida (*Ibidem*, p. 4). Sin embargo, tanto en uno como en otro -como en los demás procesos de aprendizaje autónomo- la intervención humana sigue siendo, en mayor o menor medida, necesaria, no pudiendo prescindirse de ella.

vuelve a resultar nuclear ya que la declaración realizada respecto a la ausencia de autonomía absoluta en los sistemas de inteligencia artificial, apuntando nuevamente hacia un cierto grado de autonomía pero sin ser capaz de alcanzar, inicialmente, una independencia total, resulta un postulado esencial al objeto de delimitar el ámbito de la responsabilidad penal derivada de actuaciones ilícitas, máxime cuando un sistema de inteligencia artificial no puede crear nada de la nada; esto es, toda su actuación está basada en datos obtenidos previamente. Miró Llinares ha considerado que la autonomía de una inteligencia artificial no provendría de la capacidad real de adaptar las decisiones a un contexto distinto de aquel para el que ha sido programada; esto es, o bien el aprendizaje por el que se toma la decisión no hubiera devenido de una acción humana, sino de la propia máquina que ha aprendido por sí misma o de otra máquina educadora anterior, o bien su comportamiento no dependiera de ese aprendizaje anterior y de las decisiones atribuidas previamente en la inteligencia artificial de algo ajeno a ello y propio del mismo sistema, siendo, a fecha de hoy, el ser humano quien determina la situación concreta a la que se ve expuesta la inteligencia artificial y su reacción ante una lista cerrada de estímulos<sup>17</sup>. Valls Prieto ha sido muy claro al afirmar que, hoy por hoy, no existe un sistema inteligente capaz de reaccionar como un humano, aunque pueda realizar tareas humanas con mayor eficacia<sup>18</sup>.

Salvadori ha establecido cuatro tipos de agentes artificiales según el grado/nivel de automatización y autonomía del ente:

- i) De primer nivel, los cuales operan automáticamente estando sujetos al control humano, tal y como podría ser el caso de los coches inteligentes disponibles en el mercado automovilístico actual, en los cuales una persona puede supervisar su funcionamiento e intervenir en cualquier momento para corregir la conducción.
- ii) De segundo nivel, cuyo funcionamiento depende de algoritmos deterministas, pudiendo resolver problemas ya previstos por sus programadores y desarrolladores o reaccionar de manera autónoma ante situaciones preestablecidas, tal y como puede ser el caso, por ejemplo, del software que reconoce señales de tráfico y permite que los vehículos autónomos se detengan ante un semáforo en rojo.

<sup>17</sup> MIRÓ LLINARES, *El sistema penal ante la inteligencia artificial: actitudes, usos y retos*, en *Cibercrimen III*, dirigido por Dupuy D., Corvalán J. G., Editorial B de f, Buenos Aires, 2020, pp. 87 y 88.

<sup>18</sup> VALLS PRIETO, *Sobre la responsabilidad penal por la utilización de sistemas inteligentes*, cit., p. 10.

- iii) De tercer nivel, los cuales incorporan algoritmos de aprendizaje autónomo, capaces de aprender, corregir y mejorar su comportamiento de acuerdo con la experiencia que adquieren. Aquí la intervención humana se mantiene en las fases de aprendizaje, pudiendo el programador hacer un seguimiento del citado proceso de aprendizaje e intervenir en los datos que se recogen y procesan.
- iv) El cuarto nivel o sistema multiagente puede interactuar con otros agentes y objetos adaptando autónomamente su comportamiento en función del entorno en el que opera. El ejemplo más claro se encontraría en los vehículos completamente autónomos, los cuales pueden compartir informaciones que van desde el estado del tráfico, las obras, condiciones meteorológicas, espacios en un parking (...) no sólo con el centro de información de carreteras sino también con otros coches con los que se encuentran conectados, perdiendo, con ello, y de manera progresiva, el hombre el papel de conductor del vehículo y adquiriendo el rol de mero pasajero<sup>19</sup>.

- El Parlamento Europeo ha considerado que la denominación inteligencia artificial es un término paraguas que abarca una amplia gama de tecnologías, técnicas y enfoques antiguos y nuevos, siendo más correcto referirse a 'sistemas de inteligencia artificial', entendiendo cualquier sistema basado en máquinas que a menudo tienen poco más en común que estar guiadas por un conjunto determinado de objetivos definidos por el ser humano, con diversos grados de autonomía en sus acciones, y que realizan predicciones, recomendaciones o toman decisiones basadas en los datos disponibles<sup>20</sup>.

Una vez más, puede constatarse la intrínseca relación existente entre la persona física y el sistema, no existiendo una independencia absoluta al respecto por parte del ente artificial sino que la intervención del ser humano resulta fundamental en el diseño de la posterior intervención de la máquina, aspecto nuclear, como he venido refiriendo en cada definición, a tener muy presente a la hora de delimitar la responsabilidad penal por el hecho ilícito producido.

- Doctrinalmente existen un sinnúmero de definiciones, si bien, entre todas ellas me permito destacar las llevadas a cabo por: i) Corvalán, quien refiere

---

<sup>19</sup> SALVADORI, *Agentes artificiales, opacidad tecnológica y distribución de la responsabilidad penal*, cit., pp. 147-149.

<sup>20</sup> PARLAMENTO EUROPEO, *Report on artificial intelligence in a digital age*, 2022, p. 12. Recurso electrónico disponible en: <<https://www.europarl.europa.eu>>

que la inteligencia artificial se basa en obtener, por otros métodos artificiales, lo que alcanzamos con la inteligencia humana: el reconocimiento de patrones para alcanzar objetivos o resolver problemas<sup>21</sup>; ii) Kaplan y Haenlein, quienes identifican la inteligencia artificial como la capacidad de un sistema para interpretar correctamente datos externos, aprender de ellos y utilizarlos para alcanzar objetivos y tareas específicas mediante una adaptación flexible<sup>22</sup>; o iii) Nilsson, quien la identificó con aquella actividad dedicada a hacer que las máquinas sean inteligentes, y la inteligencia es esa cualidad que permite a una entidad funcionar de manera y con previsión en su entorno<sup>23</sup>, añadiendo Cuatrecasas Monforte la capacidad para hacer generalizaciones adecuadas de un modo oportuno, basándose en datos limitados, para después, en su caso, arrojar resultados individualizados lo cual implica la necesidad de observar analizar interpretar y tomar decisiones por sí mismas, de modo que aporten alguna información o solución nueva al usuario<sup>24</sup>. En consecuencia con lo anterior, se trata de identificar mayoritariamente hoy día a la inteligencia artificial con la característica de racionalidad del agente, elemento clave que, como ya se ha referido, debe concretarse no tanto con la inteligencia sino con la capacidad de desarrollar actuaciones racionales similares o próximas al ser humano y que además va a aportar diferencias sustanciales a la hora de exigir un marco de responsabilidad penal concreto, tal y como referiré posteriormente.

Las formas en las que la inteligencia artificial puede repercutir en la comisión de hechos delictivos resultan muy plurales y diversas, lo cual ha hecho que un amplio sector de la doctrina opte explícitamente por usar el término AI-Crime para referirse a los delitos cometidos a través de inteligencia artificial<sup>25</sup>. En este sentido, Salvadori ha compilado algunas prácticas que pueden hacer entender la magnitud e implicaciones concretas recurriendo a ejemplos comunes y perfectamente comprensibles

<sup>21</sup> J.G. CORVALÁN, *Presentación: inteligencia artificial, automatización y predicciones en el Derecho*, en *Cibercrimen III*, cit., p. 14.

<sup>22</sup> 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», n. 62-1, 2018, p. 17.

<sup>23</sup> N. NILSSON, *The Quest for Artificial Intelligence: A History of Ideas and Achievements*, Cambridge University Press, Cambridge, 2010, p.13.

<sup>24</sup> CUATRECASAS MONFORTE, *La Inteligencia Artificial como herramienta de investigación criminal*, cit., p. 25.

<sup>25</sup> Vid., entre otros, T. KING, N. AGGARWAL, M. TADDEO, L. FLORIDI, *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foresseeable Threats and Solutions*, en «Science and Engineering Ethics», n. 26, 2020, p. 2.

al ser habituales en el interactuar diario, siendo los más conocidos los denominados spear phishing<sup>26</sup>:

- i) Los modernos sistemas de inteligencia artificial pueden estudiar los hábitos de los usuarios online y enviar mensajes de correo electrónico fraudulentos mucho más sofisticados que los que han sido elaborados a través de tradicionales técnicas de social engineering. De este modo, aumentan las posibilidades de inducir a error a los internautas obteniendo de forma ilícita datos personales contenidos en tarjetas de crédito, credenciales para acceder a home banking, datos sanitarios, etc.
- ii) Los sistemas de inteligencia artificial pueden ser programados para acceder ilegalmente a sistemas informáticos, para llevar a cabo ataques distribuidos en la red (los llamados ataques distribuidos de denegación de servicios) o difundir peligrosos malware o ransomware.
- iii) Las denominadas social media bots contribuyen de forma significativa a la proliferación en las redes sociales de discursos de odio y campañas de desinformación, alimentadas de fake news, creadas con el objetivo de manipular la opinión pública, de crear consenso político o de desestabilizar los equilibrios geopolíticos<sup>27</sup>.

Respecto a la regulación de semejantes prácticas ilícitas, la Unión Europea considera que, como en todo, siempre hay niveles de interacción y, en este sentido, diferencia entre los usos de inteligencia artificial de alto y bajo riesgo, siendo la regulación existente al efecto directamente proporcional a ello; esto es, necesitando las prácticas calificada de ‘alto riesgo’ -debe verificarse su uso concreto y el contexto, la naturaleza, la probabilidad, la gravedad y la irreversibilidad potencial del daño que puede esperarse que se produzca en violación de los derechos fundamentales y de las normas de salud y seguridad establecidas en el Derecho de la Unión- de estrictas

---

<sup>26</sup> Típicos correos electrónicos, aunque también páginas webs o mensajes de texto, falsos, aunque su apariencia pueda parecer real -hay un estudio cibernético previo de la víctima-, en donde se solicita acceso o información en la que el sujeto necesariamente debe abrir sesión en alguna plataforma. En este sentido, da igual la naturaleza de la información asociada, lo importante es la obtención del acceso ya que los usurpadores parten de la premisa de que los usuarios poseen muy pocas claves y la obtenida puede abrir otros accesos; esto es, por utilizar un símil, se trataría de obtener una llave que puede abrir muchas puertas. Esto se complica, aún más, en el caso de inserción de algún tipo de software malicioso en un segundo plano en el equipo de la víctima pudiendo espiar de manera libre.

<sup>27</sup> SALVADORI, *Agentes artificiales, opacidad tecnológica y distribución de la responsabilidad penal*, cit., pp. 141-143.

salvaguardias legislativas adicionales, mientras que los casos de uso de 'bajo riesgo' pueden, en muchos casos, requerir requisitos de transparencia simplemente para los usuarios finales y los consumidores<sup>28</sup>.

## *2. Situación actual de la responsabilidad penal por un hecho delictivo cometido por inteligencia artificial*

La praxis de la cuestión formulada puede resumirse muy sucintamente en la resolución de las siguientes preguntas: ¿puede la inteligencia artificial considerarse autora de un delito y, consecuentemente, ser objeto de pena? Si la respuesta fuera negativa, ¿quién respondería penalmente por semejante ilícito? Es aquí donde, como muy bien ha referido De la Cuesta Aguado, el jurista debe precisar cuándo o con qué grado de inteligencia se supera el umbral de la responsabilidad; es decir, el instante en el que la inteligencia artificial pueda ser considerada responsable de sus decisiones<sup>29</sup>. Pérez Arias ha manifestado en esta obra que el principal problema con el que se encuentra el Derecho penal aparece en la difícil -si no imposible- conexión causal entre la conducta del autor y el algoritmo en el que finalmente se desencadena la conducta ideada. En consecuencia, el problema penal se puede volver muy peliagudo, porque todo puede partir de un algoritmo que aprende en la red y que no solo depende del autor intelectual (persona física) sino del desarrollo autónomo que generará una vez entre en la red<sup>30</sup>.

Valls Prieto parte de la premisa de que cuanto más preciso es un análisis realizado por esta tecnología más difícil es conseguir una explicación de por qué la máquina ha tomado esa decisión. Este oscurantismo, que se describe con el concepto de 'black box', hace imposible establecer un nexo causal claro entre la lesión del bien jurídico protegido y el proceso de toma de decisiones por parte del sistema inteligente. Para evitar esta situación se debe realizar una evaluación en dos momentos cruciales: i) antes de empezar a diseñar el artefacto; y ii) antes de su lanzamiento al público en general. Semejante evaluación quedará registrada de tal forma que el juez podrá determinar si se han cumplido con las medidas adecuadas para mitigar los

---

<sup>28</sup> PARLAMENTO EUROPEO, *Report on artificial intelligence in a digital age*, cit., pp. 34 y 35. Recurso electrónico disponible en: <<https://www.europarl.europa.eu>>

<sup>29</sup> DE LA CUESTA AGUADO, *Inteligencia artificial y Derecho Penal*, cit., p. 52.

<sup>30</sup> PÉREZ ARIAS, *Algoritmos y Big Data en la responsabilidad penal: el reto de la cibercriminalidad en el Derecho penal*, en esta obra, § 4 de su Capítulo.

riesgos conocidos y, en algunos casos, también los desconocidos<sup>31</sup>.

Sea como fuere, Demetrio Crespo considera que los desafíos para la teoría de la acción penal van hoy día más allá debido a la irrupción de la inteligencia artificial, que pone a prueba la capacidad de rendimiento del actual sistema de imputación jurídico penal<sup>32</sup>.

### 2.1. *¿Puede considerarse a la inteligencia artificial autora única del delito?*

La respuesta inicialmente debe ser negativa, toda vez que se trata de una máquina que no hace más que obedecer/interactuar sobre los parámetros asignados por su primigenio creador; esto es, lo único que haría sería obedecer las instrucciones asignadas al respecto. Ahora bien, tal y como advierte Salvadori, el funcionamiento de los agentes de software y robots que utilizan tecnologías relacionada con inteligencia artificial y el aprendizaje automático es diferente ya que los sistemas de inteligencia artificial operan sobre la base de algoritmos complejos a través de los cuales adquieren la capacidad de autoaprendizaje, de mejorar su desempeño en el tiempo y de adoptar soluciones dinámicas; esto es, la inteligencia artificial favorece el desarrollo de los agentes artificiales con un nivel de autonomía tal que son capaces de realizar actividades que ni sus programadores, ni desarrolladores ni productores podrían haber previsto o querido. Es más, respecto de los agentes artificiales dotados de una alta autonomía<sup>33</sup>, sus actuaciones no pueden ser determinadas ni previstas ex ante por los operadores que lo diseñaron, desarrollaron, probaron o produjeron al depender directamente de los input que adquieren con su experiencia y procesan de manera independiente<sup>34</sup>.

En este sentido, como ha señalado De la Cuesta Aguado, la clave podría radicar en que la inteligencia fuera capaz de articular decisiones autónomas; esto es, aquéllas tomadas exclusivamente por el propio agente, el cual sería

---

<sup>31</sup> VALLS PRIETO, *Sobre la responsabilidad penal por la utilización de sistemas inteligentes*, cit., p. 23.

<sup>32</sup> DEMETRIO CRESPO, *El Derecho penal ante el desafío neurotecnológico y algorítmico: reflexiones preliminares*, cit., pp. 24 y 25.

<sup>33</sup> Conviene llamar la atención en este punto ya que se dice expresamente ‘alta autonomía’, no existiendo, por tanto, una autonomía plena, lo cual puede resultar muy relevante a la hora de proceder a la atribución de responsabilidad penal por hechos delictivos; esto es, la ‘alta autonomía’ no conlleva una total independencia del origen sino que sigue manteniéndose una minoritaria dependencia.

<sup>34</sup> SALVADORI, *Agentes artificiales, opacidad tecnológica y distribución de la responsabilidad penal*, cit., pp. 143 y 144.

capaz de dirigir su comportamiento o el de otros con su decisión; o, en otras palabras, incluiría la capacidad de actuación sin intervenciones externas, y en particular, humanas, si bien, conviene advertir, que, de momento y dado el estado de la ciencia, por decisiones autónomas debe entenderse autonomía en cuanto a la conducta o comportamiento más adecuado a un fin u objetivo, excluyendo la autonomía de la decisión sobre el fin o el objetivo; esto es, se limita el análisis a problemas en los que los fines u objetivos han sido fijados previamente de forma externa a la gente artificial inteligente y, conforme a ello se implica a una persona (o a varias) en la toma de decisiones, por lo que verdaderamente se le hace responsable a ese sujeto de las consecuencias de los actos del agente artificial. Piénsese en un robot para el control de fronteras que tiene la orden de disparar cada vez que detecta, en una franja de tierra, a personas en movimiento, respondería a la premisa de autónomo, en el sentido en el que pueda actuar sin intervención externa, pero no sería inteligente. Ahora bien, la autonomía relacionada con la inteligencia sería aquella que permite decidir al robot qué debe hacer según las circunstancias, con la posibilidad, incluso, de adoptar comportamientos para los que no ha sido predeterminado en su software. Esta autonomía abriría el paso a la responsabilidad derivada de la libertad de decisión como concepto contrario al de necesidad de decisión. Hoy día existen agentes artificiales inteligentes capaces de adoptar decisiones de comportamiento para la consecución de objetivos en función de fines predeterminados y en atención a circunstancias, lo cual se traduce en decisiones autónomas que hacen que no dependan de autorizaciones o controles de personas<sup>35</sup>.

Es aquí, precisamente, donde Salvadori focaliza la cuestión en el sentido de considerar a los agentes artificiales capaces de aprender y de autodeterminarse donde puede surgir la problemática alrededor de si el Derecho Penal debe reconocer su autonomía efectiva y plena y equipararlos con sujetos (humanos) penalmente responsables<sup>36</sup>. Siguiendo esta línea argumental, lejana en el tiempo a fecha de hoy, Romeo Casabona llama la atención sobre el instante futuro en el que la máquina pudiera tener esa intención criminal de manera idéntica a un humano, lo cual implicaría indefectiblemente la necesidad de modificar el actual sistema de responsabilidad penal para adaptarlo a los entes autónomos<sup>37</sup>, una situación

---

<sup>35</sup> DE LA CUESTA AGUADO, *Inteligencia artificial y Derecho Penal*, cit., pp. 52 y 53.

<sup>36</sup> SALVADORI, *Agentes artificiales, opacidad tecnológica y distribución de la responsabilidad penal*, cit., p. 146.

<sup>37</sup> C.M. ROMEO CASABONA, *Criminal responsibility of robots and autonomous artificial intelligent systems?*, en «Comunicaciones en propiedad industrial y derecho de la competencia», n. 91, 2020, pp. 170-172.

respecto de la cual, como afirma Miró Llinares, queda aún lejos, toda vez que no existe una inteligencia artificial autónoma como tal, siendo la actual una suma de algoritmos de predicción utilizados para la realización de acciones o recomendaciones para actuar a partir de un conjunto de datos existentes identificación en ellos de patrones y probabilidades, siendo los seres humanos los que determinan completamente el actuar de la máquina, no necesitando ningún tipo de cambio actual en el sistema de atribución de responsabilidad penal, algo que indefectiblemente sucederá en un futuro cuando algún sistema de inteligencia artificial tenga rasgos de autonomía<sup>38</sup>.

Sea como fuere, Salvadori distingue la situación actual de la hipotética de futuro. Así, mientras ahora los agentes artificiales más avanzados no tienen todavía una conciencia de sí mismos que les permita comprender el valor social negativo de sus acciones y captar la función preceptiva del Derecho penal, quedan igualmente considerables dudas sobre la (presunta) posibilidad de actuar dolosamente, es decir, con 'voluntad consciente' de cometer un hecho constitutivo de delito, o de forma imprudente, como actitud de incumplimiento de la norma cautelar. De forma resumida: un agente artificial (hasta la fecha) no puede ser culpable, como tampoco puede percibir el efecto intimidatorio ni disuasorio de la pena y, menos aún, garantizar la función especial de prevención ya que la sanción no sería adecuada para reducir ni resocializar al agente artificial, quedando únicamente la vía de restablecer los algoritmos que han determinado o influido en su toma de decisión. Ahora bien, de futuro, resulta previsible que los agentes artificiales lleguen a ser completamente autónomos, lo que significaría la posesión de una capacidad propia para actuar y, en algunos aspectos, ser capaces de percibir la realidad que les rodea y de operar de forma 'consciente', lo que devendría en el reconocimiento de la personalidad jurídica a la 'entidad' no humana, algo que tampoco resultaría una novedad, dado que en muchos ordenamientos jurídicos, tanto del common law como del civil law, se reconoce pacíficamente a las entidades y personas jurídicas incluso a efectos penales. A ello debe contribuir, para superar las limitaciones actuales, el hecho de que la existencia de la libertad de la voluntad en la cabeza del hombre no tiene un fundamento ontológico, sino que es el resultado de una construcción social que podría hacerse valer también contra los agentes artificiales completamente autónomos, desde el momento en el que, como las personas, interactúan e intervienen en las relaciones sociales y tienen un poder de conformación de la realidad. En definitiva, concluye

---

<sup>38</sup> F. MIRÓ LLINARES, *El sistema penal ante la inteligencia artificial: actitudes, usos y retos*, cit., pp. 89 y 90.

el autor, la verdadera cuestión subyacente no es si los agentes artificiales están dotados de la libertad de actuar como seres humanos, sino si, en un futuro próximo, podrán ser considerados responsables de la realización de conductas socialmente desviadas<sup>39</sup>.

En este punto, Dupuy tiene clara la actuación directa desde una perspectiva de responsabilidad civil derivada de los daños ocasionados, no cabiendo realizar una traslación idéntica desde la perspectiva penal debido al referido problema de la culpabilidad, por lo que la autora sugiere dos vías de actuación: aplicando las reglas del Derecho vigente en relación con esta nueva realidad -excluiría una intervención penal- o creando un corpus iuris propio de la robótica, lo que requiere como presupuesto central el reconocimiento de cierta entidad (personalidad jurídica) a los robots como sujetos de Derecho, dando pie al origen del denominado y rechazado Derecho Penal de los robots<sup>40</sup>.

Llevar a término esta última opción requiere, como ya se ha referido, dotar de personalidad a la inteligencia artificial, algo inviable conforme a la normativa española actual ya que, como ha referido Santos González, no pueden ser considerados: i) persona física, ya que chocaría frontalmente con el artículo 30 del Código Civil al requerir unos requisitos biológicos que no pueden darse en un robot; ii) como persona jurídica, sería posible su configuración, requiriendo ser destinatario de derechos y obligaciones, si bien nuevamente vuelven a surgir imposibilidades debido, entre otras circunstancias, a que la persona jurídica se configura sobre la unión de un grupo de personas físicas, algo claramente alejado de los robots, al igual que el hecho de generar daños por sí mismo, algo respecto de lo cual la inteligencia artificial tendría capacidad mientras le persona jurídica no; iii) como animal, tampoco enlazaría con el citado reconocimiento al no existir una base genética o biológica que permita experimentar a la fecha sentimientos naturales; y iv) como cosa, la cual se identifica como algo inanimado, carente de vida, frente a los robots, respecto de los cuales se persigue justamente lo contrario, que realicen conductas inteligentes, tengan autonomía para moverse e interactuar, debiendo considerarse como algo animado<sup>41</sup>.

Sin embargo, poco a poco, va cobrando cada vez más peso la idea de crear una personalidad jurídica electrónica, lo cual, según recoge Busato, sería el

<sup>39</sup> SALVADORI, *Agentes artificiales, opacidad tecnológica y distribución de la responsabilidad penal*, cit., pp. 151-155.

<sup>40</sup> DUPUY, *Inteligencia artificial aplicada al Derecho Penal y Procesal Penal*, en *Ciberdelincuencia II*, cit., p. 297.

<sup>41</sup> M<sup>a</sup> J. SANTOS GONZÁLEZ, *Regulación legal de la robótica y la inteligencia artificial: retos de futuro*, en «Revista Jurídica de la Universidad de León», n. 4, 2017, pp. 39-43.

producto de dos aspectos distintos: uno, el legislativo, que ha pretendido establecer una dinámica de control jurídico de la evolución científica en el ámbito de la robótica; y, otro, el argumento de equivalencia funcional entre personalidad humana y electrónica, el cual debe tomar como punto de partida, la concepción de acción anclada en la mecánica newtoniana y la expresión de la subjetividad a partir del modelo cartesiano del cogito ergo sum. Este segundo aspecto puede observarse en el hecho de que la evolución de la inteligencia artificial ha llegado a tal punto que pueden observarse en ella características humanas como la comunicación, la internalización del conocimiento, la capacidad de aprendizaje, la toma de decisiones, incluso la creatividad y la espontaneidad, lo que significaría que las máquinas razonan de forma similar a un humano, con lo que merecerían idéntico tratamiento jurídico, concluyendo que los robots dotados de inteligencia artificial son capaces de realizar tanto el actus reus -la capacidad de acción se correspondería con la capacidad de producción de un movimiento físico que produce un resultado- como el mens rea -capacidad de pensar, la cual se vincularía al aprendizaje autónomo- exigidos para la responsabilidad penal; es decir, resultan ser capaces de atender a los requisitos objetivos y subjetivos de la imputación<sup>42</sup>.

Dupuy considera finalmente que los robots, por más autonomía que tengan -drones, vehículos sin conductor, etc.- deben mantener, a fecha de hoy, la configuración y el tratamiento jurídico de las cosas<sup>43</sup>. Por lo tanto, su utilización por un sujeto para cometer un delito significará que el dominio del hecho siempre le corresponderá al humano y, en base a ello, la imputación se realizará con base en la comisión directa y personal de la acción típica, abriendo un nuevo debate sobre qué persona y en qué condiciones se encontraría detrás de la responsabilidad penal<sup>44</sup>.

---

<sup>42</sup> No obstante, el citado autor concluye indicando que no se trata de considerar a la inteligencia artificial como persona incapaz de culpabilidad, sino, más bien, de reconocerlas como incapaces de acción (P. C. BUSATO, *De máquinas y seres vivos*, en *Derecho Penal y comportamiento humano*, dirigido por Demetrio Crespo E., Tirant Lo Blanch, Valencia, 2022, pp. 352-354 y 372).

<sup>43</sup> No obstante lo anterior, hay un corriente que viene demandando la creación de una nueva categoría de personalidad y así, como ha referido Santos González, puede surgir la opción de identificar a los robots como 'persona electrónica' o 'robot inteligente artificialmente' (*Ibidem*, pp. 43-46).

<sup>44</sup> DUPUY, *Inteligencia artificial aplicada al Derecho Penal y Procesal Penal*, cit., pp. 297 y 298.

## 2.2. ¿Quién debe responder penalmente entonces?

Delimitada la ausencia de responsabilidad directa sobre la inteligencia artificial, en los términos referidos en el apartado anterior, debe concretarse, a título doloso o imprudente, la persona que debe responder por las consecuencias jurídicas derivadas de la actuación de un sistema de inteligencia artificial, debiendo recurrir a criterios que permitan verificar cómo y por qué se ha producido el fallo correspondiente o bien la existencia de una voluntad criminal previamente establecida.

De la Cuesta Aguado ha ejemplificado las distintas hipótesis delictuales posibles respecto de los daños generados por un dron sobre tres modelos de responsabilidad penal:

- a) Responsabilidad personal por acciones o daños u otros resultados típicos, cuando la aeronave actúe bajo el control de una persona.
- b) Responsabilidad personal por acciones o daños u otros resultados típicos cuando la aeronave actúe dirigida por una persona pero en la decisión en concreto hayan intervenido factores ajenos al individuo que dirige o desarrolla el software.
- c) Responsabilidad personal por acciones o daños u otros resultados típicos cuando la aeronave actúe autónomamente, en cuyo caso habría que introducir dos sub-variantes: i) como consecuencia de errores, déficits o deficiencias de cualquier tipo en el software que establece fines, objetivos, límites y condiciones que incidieron en la toma de decisiones; y ii) subsidiariamente cuando no concurriera el supuesto anterior, por la decisión autónoma y derivada del propio aprendizaje de la aeronave<sup>45</sup>.

Valls Prieto ha establecido una metodología de actuación basada en la fase de interacción. En este sentido, considera la existencia de tres niveles de imputación: i) compuesto por desarrolladores y fabricantes del producto que lo diseñan y mandan al mercado; ii) profesionales que utilizan estos sistemas inteligentes para realizar una parte de su trabajo y, por tanto, interacciona con la máquina de forma diferente a cómo lo realizarían los desarrolladores; y iii) los usuarios finales, en muchos casos consumidores del sistema. Como puede comprobarse cada uno de ellos tiene una relación y control diferente respecto al sistema inteligente, lo que, según este autor, deriva en una responsabilidad distinta. No obstante, lo anterior debe ser complementado para determinar si se ha producido un uso inadecuado

---

<sup>45</sup> DE LA CUESTA AGUADO, *Inteligencia artificial y Derecho Penal*, cit., p. 54.

e irresponsable de la inteligencia artificial. Para ello, debe igualmente verificarse cómo se ha producido la recolección de datos, su procesamiento y la posterior interpretación de los resultados finales por parte del sistema de inteligencia artificial<sup>46</sup>.

Miró Llinares ha concluido que mientras no puede atribuirse autonomía a las entidades con inteligencia artificial el sistema de la teoría del delito sigue siendo totalmente válido para resolver los diferentes problemas causales y de atribución de responsabilidad, generalmente imprudente, si bien resulta esencial monitorizar la evolución de la inteligencia artificial desde una perspectiva de atribución de responsabilidad para evitar llegar a situaciones en las que el aprendizaje de las máquinas no permita decir que nadie ha tomado una decisión negligente pese a que existan daños<sup>47</sup>.

Ahora bien, yendo paso a paso, puede afirmarse que, en primer lugar, no parece revertir problema jurídico-penal alguno el hecho de que un sujeto utilice consciente y voluntariamente un sistema de inteligencia artificial para desarrollar un comportamiento criminal, respondiendo dolosamente en concepto de autor ya que en estos casos se utilizaría al sistema de inteligencia artificial como herramienta o medio para la comisión del delito -piénsese, por ejemplo, en el uso de un dron para acabar con la vida de alguien-.

Por el contrario, la verdadera problemática subyacerá cuando no existe semejante grado de conocimiento respecto del posible resultado ilícito o simplemente cuando se genera el delito sin ser ésta la voluntad inicial del individuo, abriendo en estos casos la puerta a la imputación a título imprudente.

Sobre este extremo, Valls Prieto hace especial hincapié en la necesidad de determinar cuáles son las normas de cuidado que se tienen que respetar y que son exigibles a los ciudadanos en el uso de sistemas inteligentes para poder delimitar quién ha cometido el delito. Así pues, dentro de los diferentes códigos éticos que existen los principios expuestos por las guías éticas de la Comisión Europea resultan, a juicio de este autor, la fuente más completa, lo cual se traduciría, siguiendo los parámetros de algunos autores<sup>48</sup>, en trasladar el sistema de compliance de las personas jurídicas a los sistemas inteligentes, si bien el modelo y la responsabilidad no podría ser idéntico por los problemas presentados, a saber: los programas de

---

<sup>46</sup> VALLS PRIETO, «*Sobre la responsabilidad penal por la utilización de sistemas inteligentes*» cit., pp. 25 y 26.

<sup>47</sup> MIRÓ LLINARES, *El sistema penal ante la inteligencia artificial: actitudes, usos y retos*, cit., pp. 112 y 113.

<sup>48</sup> Vid., a tal efecto, ROMEO CASABONA, *Criminal responsibility of robots and autonomous artificial intelligent systems?*, cit., p. 173.

compliance son realizados por personas jurídicas de forma obligatoria y en caso de cumplir les exime de responsabilidad penal. Por contra, el sistema inteligente no puede realizar por sí mismo este cumplimiento normativo, siendo un externo el que tomará la decisión de realizarlo o no. Además, mientras que dentro de la empresa se pueden tomar este tipo de decisiones, en el sistema inteligente no es posible, siendo una persona externa al sistema quién tomará las decisiones. Finalmente, el cumplimiento normativo implica la necesidad de un marco jurídico que regule el incumplimiento, mientras en materia de inteligencia artificial se parte de principios éticos (no normativizados) y, en consecuencia, los sistemas de prevención de riesgos son más amplios que en el caso de compliance<sup>49</sup>.

De la Cuesta Aguado considera fundamental distinguir entre dos situaciones, según la responsabilidad provenga por el producto o por los daños. En el primer caso, piénsese, por ejemplo, en un software que contenga errores, la responsabilidad se compondría de círculos individuales, compartida parcialmente o no, en función de los propios ámbitos de competencia de cada uno de los intervinientes en la fabricación y uso del producto. Esta materia no es objeto ajeno al Derecho Penal y así podrían establecerse distintas responsabilidades individuales según los siguientes supuestos: i) de quien comete el error en su ámbito de competencia que puede implicar errores de distintas personas; ii) por culpa in vigilando; y iii) derivada de la creación o uso de una fuente de peligro. De otro lado, y en relación con la responsabilidad penal por daños, lesiones o, en general, resultado delictivo, aquí el agente artificial podría ser considerado como dependiente de la responsabilidad de una persona, ya sea por errores humanos de consecuencias previsibles o imprevisibles<sup>50</sup>.

Salvadori ha compilado los supuestos en los que podría generarse la aplicación de tipos imprudentes según el rol desempeñado por la persona, siempre que exista una previsibilidad del riesgo<sup>51</sup>, destacando las hipótesis siguientes:

---

<sup>49</sup> VALLS PRIETO, *Sobre la responsabilidad penal por la utilización de sistemas inteligentes*, cit., pp. 29, 30 y 33.

<sup>50</sup> DE LA CUESTA AGUADO, *Inteligencia artificial y Derecho Penal*, cit., p. 58.

<sup>51</sup> La previsibilidad del riesgo es el elemento clave para imputar responsabilidad penal por imprudencia, debiendo advertir la posibilidad, como apuntó Hallevy, de que queden fuera de la citada cobertura los casos en los que el resultado no pudiera ser anticipado por el programador o, en última instancia, por el usuario final (G. HALLEVY, *The Criminal Liability of Artificial Intelligence Entities from Science Fiction to Legal Social Control*, en «Akron Intellectual Property Journal», 2016, p. 172).

- i) Los hechos lesivos que tienen lugar debido a un error de programación o funcionamiento del agente artificial podrán ser imputados a título imprudente a quienes lo han programado, desarrollado, producido o probado.
- ii) Quien utiliza un agente artificial podrá ser llamado a responder de las consecuencias lesivas que se deriven de su uso si tiene poder efectivo para prevenir el hecho. Piénsese, como paradigma, en el vehículo semiautomático que, cerca de un cruce, no se detiene en la señal de stop, golpeando fatalmente a un peatón. El fallecimiento del peatón será atribuido al conductor si se demuestra que con la debida diligencia podría (y debería) haber tomado medidas para tomar el control del automóvil y evitar el golpe.
- iii) Cuanto mayor sea el nivel de autonomía del agente artificial, más elevada será la dificultad para imputar responsabilidad penal a personas físicas. Piénsese en un agente de software que actúa como piloto automático en las complejas fases de despegue y aterrizaje de un avión, no proporcionando posibilidad alguna de control manual por parte de los pilotos en caso de emergencia. Si a raíz de un error en el funcionamiento de los sensores que monitorean las condiciones climáticas, la aeronave pierde repentinamente altitud y se estrella contra el suelo provocando la muerte de los pasajeros, no puede hacerse ningún reproche imprudente a los pilotos, si resulta que no podían prever y evitar el accidente. No obstante, habría que hacer un rastreo hasta llegar a los sujetos que diseñaron, programaron y produjeron el software para analizar las complejas fases del vuelo; al fabricante del software, a quien se le podría reprochar a través de la imprudencia, cualquier error cometido en las fases de producción; e incluso al desarrollador o programador por un denominado ‘fallo de programación’ o por no haber previsto la posibilidad de permitir, ante situaciones de emergencia, la intervención manual de los pilotos en situaciones críticas.
- iv) La imputación debe realizarse extensiva en el caso de que verificados problemas de seguridad posteriores -piénsese, por ejemplo, en actuaciones de hackers- no se lancen actualizaciones necesarias para remediar fallos de software o, caso de no ser posible, no se retire del mercado.
- v) La responsabilidad a título imprudente no tiene por qué cerrarse en una única persona ya que si consigue verificarse el error de varios operadores, algunos ordenamientos jurídicos contemplan figuras expresas ante ello, como puede ser el caso del Código Penal italiano y el concurso de condiciones imprudentes independientes o la cooperación imprudente<sup>52</sup>.

---

<sup>52</sup> SALVADORI, *Agentes artificiales, opacidad tecnológica y distribución de la responsabilidad penal*, cit., pp. 155-159.

Sobre la casuística anterior, considero muy relevante la hipótesis expuesta relativa al caso de la conducción de un vehículo de manera autónoma<sup>53</sup> en el que, fruto de un error del sistema, por el fallo de un sensor o no reconocimiento de una señal -piénsese, en un semáforo en rojo no detectado o simplemente en la salida del garaje del vehículo para posicionarse en la calzada- se genera el atropello de una o varias personas. Inicialmente, parece claro que el ocupante del vehículo automatizado exigirá responsabilidad a la empresa responsable del vehículo, siempre y cuando el mantenimiento y estado de los sensores correspondientes fuera el adecuado, ya que ha adquirido una máquina con una determinada fiabilidad probada científicamente con una serie de test y otros protocolos de seguridad. Sin embargo, las empresas fabricantes y comercializadoras no dejan margen a la improvisación y, para evitar riesgos de este tipo, incluyen cláusulas referentes a que la conducción automática exige la supervisión del conductor quien debe estar en todo momento atento a cualquier incidencia que pueda acontecer para evitar riesgos indeseados o, dicho en otras palabras, hay un deber de cuidado o actuación implícito en el ocupante del vehículo quien, pese a no manejar la máquina, debe estar preparado para tomar el control ante la menor incidencia, pudiendo derivar su inatención de una inobservancia del deber de cuidado que le era exigible con la consiguiente traslación de la responsabilidad penal a su persona.

### 3. *Pronunciamientos jurisprudenciales expresos*

Pese a que, como ha manifestado Alonso Salgado, aún hoy nuestros Juzgados y Tribunales distan mucho, desde una perspectiva estricta, de ser espacios tecnológicos, máxime cuando siguen instalados todavía en el tránsito transformador del siglo XX al XXI, irrumpe un nuevo desarrollo hijo de la nueva era, que tensiona nuestra tradición legal y nuestra cultura jurídica, que pone a prueba nuestra capacidad de adaptación y que evidencia resistencias, reparos y objeciones. Y ello, en muy buena medida, porque, lejos de lo que inicialmente pudiera parecer, ese nuevo desarrollo al que se

---

<sup>53</sup> Los delitos contra la seguridad vial presentan un amplio espectro de posibilidades e implicaciones tanto incriminadores como exculpatorias, tal y como puede observarse en el caso comentado o en el siguiente epígrafe con el supuesto de la persona con altas dosis de alcohol que aparece inconsciente con su vehículo parado en mitad de la calzada y su posible exención de responsabilidad penal en el hipotético caso de que toda la travesía hubiera sido gestionada la conducción por el sistema de inteligencia artificial.

hace referencia, constituye algo más que un giro meramente cuantitativo<sup>54</sup>.

Así las cosas y pese a que existan un mayor número de resoluciones judiciales vinculadas a los contenidos expuestos y relacionados directamente con la inteligencia artificial, la búsqueda acotada bajo el campo abierto 'inteligencia artificial' y la jurisdicción 'penal' ha dado como resultado un total de doce manifestaciones jurisprudenciales, las cuales se vinculan a circunstancias accesorias, mayoritariamente, del relato de los hechos pero que pueden compilarse en las siguientes categorías:

A) Descripción meramente accesorio, donde se incluye la denominación en el relato de los hechos o los fundamentos jurídicos por cuestiones relacionadas con el objeto del litigio pero sin relevancia jurídica alguna. En este sentido, pueden destacarse, por ejemplo:

a1) El Auto de la Audiencia Provincial de Barcelona 517/2021, de 13 de julio [JUR 2021\377853], refiere el programa piloto llevado a cabo en Cataluña relativo al pronóstico de regresión del recluso en base al cálculo del riesgo de reincidencia mediante el uso de herramientas de inteligencia artificial, como es el caso de la implementación de un algoritmo predictor, si bien no se contempla en la legislación vigente como método científico adecuado de predicción que ayude en la valoración y ponderación del pronóstico.

a2) Llama poderosamente la atención, sobre todo antes de 2013, el hecho de que negocios que incluyen en su denominación la palabra 'Inteligencia Artificial' aparezcan de manera más habitual que otras nomenclaturas en comportamientos delictivos<sup>55</sup>, siendo igualmente llamativo que, cuando una persona se encuentra formada en semejante materia, se haga expresa mención en la resolución judicial, aun cuando no tenga una vinculación relevante o directa con los hechos enjuiciados o el *modus operandi*<sup>56</sup>.

---

<sup>54</sup> 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*», vol. 7, n. 1, 2021, p. 26.

<sup>55</sup> Vid., a modo de ejemplo, el Auto de la Audiencia Provincial de Madrid 683/2012, de 29 de diciembre [JUR 2012\20403], los querellantes pertenecen a la empresa 'Sistemas de Inteligencia Artificial'; proyecto comercial en materia de inteligencia artificial; Auto de la Audiencia Nacional 22/2005, de 20 de abril [JUR 2008\341211], el menor acudió al negocio denominado 'Central de Inteligencia Artificial SA', dedicado a la venta de equipos de cómputo; Sentencia 1141/2004, de 23 de noviembre [JUR 2005\5995], los acusados prestaban sus servicios profesionales para la entidad 'AIS Aplicaciones de Inteligencia Artificial'.

<sup>56</sup> Vid., a modo de ejemplo, la Sentencia de la Audiencia Provincial de Madrid 208/2010, de 30 de junio [JUR 2010\32127], donde se destaca del condenado, por un delito de

- B) Uso en los escritos de las partes. Se emplea el término para intentar explicar algunos comportamientos ilícitos. Tal puede ser el caso, por ejemplo, del Auto de la Audiencia Provincial de Burgos 839/2021, de 30 de noviembre, [JUR 2022\60508], donde, en la querella interpuesta, se recoge que «la recopilación de datos junto con el uso de técnicas modernas como la inteligencia artificial ha generado un amplio debate y preocupación respecto a los límites que deben aplicarse, entre los que se encuentra el derecho a la protección de datos de carácter personal».
- C) Autorización de implementación de medidas de vigilancia basadas en inteligencia artificial. El Auto de la Audiencia Provincial de Barcelona 72/2021, de 15 de febrero [ARP 2021\706], invoca el Libro Blanco sobre inteligencia artificial de la Comisión Europea, de 19 de febrero de 2020, con el propósito de verificar lo que se entiende por ‘autenticación’ vinculado al reconocimiento facial en el caso en el que dos sujetos fueron condenados por delito de robo con violencia en la personas en grado de tentativa a las penas de prisión, inhabilitación (...) y prohibición de acceso al centro comercial Mercadona situado en la calle (...) por tiempo de dos años. Ante la citada resolución judicial, la cadena de supermercados pretende, ante la imposibilidad práctica de asegurar el cumplimiento de la pena accesoria, dado que los trabajadores no podían estar pendientes de la gente que entraba en el supermercado y mucho menos de si éstos habían sido efectivamente condenados, lograr la autorización para usar medios electrónicos que permitan detectar la entrada de los penados a través de un circuito cerrado de video grabación, asegurando la proporcionalidad de los medios utilizados al tiempo que la efectividad de la medida, añadiendo los intereses públicos y legítimos de la mercantil de asegurar el cumplimiento de las resoluciones judiciales en las que la misma es la víctima/perjudicada. El funcionamiento del sistema que se pretendía implementar quedaría circunscrito al reconocimiento facial en tiempo real en virtud del cual se detectaría en 0,3 segundos, única y exclusivamente, la entrada de personas con sentencias firmes y medida cautelar de orden de alejamiento en vigor contra Mercadona o contra alguno de sus trabajadoras o trabajadores, procediendo a la eliminación de todo registro de persona que no cumpla con los citados caracteres. Finalmente, la Audiencia rechaza la petición por no resultar proporcional, necesaria ni idónea, además de no respetar las garantías adecuadas en orden a la protección de los derechos y libertades de los interesados,

---

injurias leves, sus estudios en informática y su especialización en realidad virtual y sistemas de inteligencia artificial; Sentencia del Tribunal Supremo 331/2003, de 5 de marzo [RJ 2003\2814], donde el querellante participaba en un programa de inteligencia artificial.

no ya sólo de los que han sido penados y cuya prohibición de acceso les incumbe, sino del resto de personas que acceden al citado supermercado.

D) La inteligencia artificial en los delitos contra la seguridad vial. La Sentencia de la Audiencia Provincial de Navarra 199/2011, de 15 de diciembre [JUR 2013\320202], recoge el caso de un delito de conducción de vehículos a motor bajo la influencia directa del alcohol, donde unos agentes de la policía municipal se encuentran un vehículo con el motor encendido, las luces puestas pero parado, ocupando uno de los carriles de circulación, estando durmiendo en su interior una persona con el cinturón de seguridad puesto, con un fuerte olor a alcohol<sup>57</sup> y restos de vómito reciente. Ante las evidencias, el Tribunal verifica que el coche no tuviera un sistema de inteligencia artificial que le permitiera circular sólo por las distintas vías, sin una inteligencia humana que lo guié, lo cual eliminaría la responsabilidad penal del tipo aplicado, cerciorándose, por tanto, de que tuvo que haber una inteligencia humana que lo guiara, siendo sólo posible que fuera el acusado, que en aquel momento tenía no sólo aquella, sino también la voluntad e inteligencia sensiblemente mermadas hasta el punto de que con cinturón de seguridad puesto, motor en marcha y luces encendidas del vehículo, se quedó dormido en el vehículo parado en la citada vía.

La anterior Sentencia recoge la esencia de la problemática que puede derivarse de futuro respecto al uso de los coches autónomos y las connotaciones que puede introducir en los delitos contra la seguridad vial y, más concretamente, los vinculados al artículo 379.2 del Código Penal ya que habrá que verificar, por ejemplo, que hubo conducción humana del vehículo aun cuando el sujeto se encontrara en el asiento del conductor con una ingesta alcohólica o de drogas que incidiera en la conducción.

#### 4. Conclusiones

1. La inteligencia artificial constituye hoy día una realidad creciente e intensa, pudiendo afirmarse que estamos, como especifica la Resolución del Parlamento Europeo, de 3 de mayo de 2022, ante una cuarta revolución industrial ya que su incidencia ha generado un enorme impacto a nivel tecnológico mundial, incidiendo directamente en ámbitos pro-

---

<sup>57</sup> Posteriormente, el sujeto dio un resultado de 1,01 mg/l de alcohol.

pios y sensibles de cada país, como puede ser el caso de la economía, la sociedad, las relaciones internacionales o el medio ambiente. Semejante hecho no comporta una carta abierta en virtud de la cual todo tipo de actuaciones queden legitimadas en pro de los ámbitos anteriormente referidos, existiendo un deber de respeto a los derechos fundamentales, lo cual conlleva observar unas cautelas mínimas a la hora de regular su uso, no debiendo olvidar que se trata de un instrumento al servicio de las personas cuyo objetivo primordial ha de ser el bienestar humano y la seguridad.

2. No existe un concepto pacífico respecto a lo que debe entenderse por inteligencia artificial y así son muchas las disquisiciones existentes al respecto, partiendo incluso de la propia definición de inteligencia. Ahora bien, lo que no debe negarse, en pleno año 2022, es la existencia, junto con la inteligencia de los seres humanos, de otros sistemas de inteligencia, en este caso particular artificial, entendiéndose por semejante adjetivación el origen no intrínsecamente humano del citado pensamiento; esto es, la capacidad propia de pensamiento y su ejecución de una máquina, la cual, por un complejo procedimiento de algoritmos, puede realizar tareas similares a las personas, optimizando sus resultados tanto en un prisma temporal como cualitativo, constituyendo ese su principal objetivo fundacional, que no debe ser otro que realizar de manera satisfactoria determinadas actividades de forma precisa, acortando los plazos tradicionales y sin necesidad de una posterior intervención humana.
3. No obstante lo anterior, hay dos elementos que deben tenerse muy presentes desde una perspectiva jurídica: i) el origen del sistema de inteligencia artificial se encuentra embrionariamente ligado al ser humano y; ii) la inteligencia artificial no goza de absoluta autonomía, no siendo capaz de alcanzar una independencia total, al menos a fecha de hoy, lo cual resulta fundamental al objeto de delimitar el ámbito de la responsabilidad penal derivada de actuaciones ilícitas en las que se ve implicada un sistema de inteligencia artificial o, dicho en otras palabras, toda su actuación está basada en datos obtenidos previamente, detrás de los cuales, en algún momento anterior, ha intervenido un ser humano.
4. Como ha demostrado la historia, todo sistema evolutivo que mejora las condiciones de vida del ser humano tiene igualmente connotaciones negativas, las cuales suelen vincularse al ámbito delictivo y, en este caso, se ha acotado el término AI-Crime para referirse a los delitos cometidos a través de inteligencia artificial. Es por ello que los estados se ven obli-

gados a regular y crear instrumentos de control respecto de las prácticas delictivas derivadas del uso ilícito de los sistemas de inteligencia artificial, volviendo a reabrirse debates vinculados a cómo regular los concretos sistemas de responsabilidad penal derivados por la presencia de la inteligencia artificial en la comisión del delito y la clásica pregunta de si son suficientes los mecanismos existentes, bastando únicamente con adaptarlos a la nueva situación criminal o, por el contrario, se requiere de un nuevo sistema normativo que se adecúe a esta realidad.

5. Las preguntas que conviene realizar al respecto son claras: ¿puede considerarse a la inteligencia artificial autora única del delito? Y, en caso de respuesta negativa, ¿quién debería responder penalmente por el acto ilícito producido?

Con relación a la primera, la respuesta debe ser necesariamente negativa ya que una máquina no hace más que obedecer/interactuar sobre los parámetros asignados por su primigenio creador; esto es, por más que quiera, el sistema de inteligencia no va a gozar, al menos hasta la fecha, de una autonomía plena, sí de una alta autonomía pero nunca absoluta, reduciéndose el problema a aquellos casos en los que sus actuaciones no puedan ser determinadas ni previstas *ex ante* por los operadores que la diseñaron, desarrollaron, probaron o produjeron, siempre que dependa directamente de los inputs que adquiera con su experiencia y procese de manera independiente. No obstante, lo que parece claro es que los sistemas de inteligencia artificial carecen de personalidad, capacidad de culpabilidad e intimidación ya que no son capaces de percibir el efecto disuasorio de la pena, lo cual supone un nuevo eslabón a la hora de atribuirles específicamente responsabilidad penal, lo cual no resulta óbice para que algún sector doctrinal empiece a reclamar una regulación específica de los robots, lo que derivaría en un derecho de la robótica y, en el ámbito que ahora me ocupa, un Derecho Penal de los robots, donde alcanzaría una dimensión fundamental un nuevo rol denominado 'personalidad jurídica electrónica', si bien todo esto, repito, no sean más que consideraciones de futuro.

Negada, por tanto, la capacidad jurídica para cometer delitos a los sistemas de inteligencia artificial, cabe plantear la cuestión relativa a ¿quién responde entonces? El sistema de atribución de responsabilidad penal será el propio de las personas físicas, aun cuando un sector doctrinal abogue también por la inclusión de un sistema o régimen similar al de las personas jurídicas. Lo relevante aquí es la determinación de la responsabilidad, a título doloso o imprudente, según el actuar del sujeto o sujetos que, conforme a la relación causal concreta del hecho, sean considerados

responsables del ilícito producido, debiendo retrotraer la intervención de cada uno de los agentes, desde el responsable consumidor o titular del sistema de inteligencia artificial hasta los programadores/trabajadores de la empresa fabricante o comercializadora, quienes, conviene advertir, para evitar riesgos al respecto suelen incluir cláusulas de advertencia tendentes a eliminar una hipotética responsabilidad por la generación de un hecho ilícito en el desarrollo de la actividad concreta de la inteligencia artificial -véase, por ejemplo, en el caso de la conducción automática-.

6. La descripción inteligencia artificial en la jurisprudencia penal española no cuenta con una excesiva representatividad si nos atenemos a su descripción taxativa. En este sentido, conviene llamar la atención en que su aparición expresa puede encuadrarse en alguna de las siguientes categorías: i) como descripción meramente accesoria al supuesto juzgado; ii) breves y escuetas referencias en el escrito de alguna de las partes; iii) como autorización de implementación de medidas de vigilancia basadas en inteligencia artificial; o iv) lo que considero el aspecto más relevante, su implicación en los delitos contra la seguridad vial, donde puede modificar los patrones de actuación, persecución y enjuiciamiento de estos delitos, abriendo un nuevo espectro de posibilidades y actuaciones.

## 5. Bibliografía

- 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*», vol. 7, n. 1, 2021.
- BUSATO P.C., *De máquinas y seres vivos*, en *Derecho Penal y comportamiento humano*, dirigido por Demetrio Crespo E., Tirant Lo Blanch, Valencia, 2022.
- CORVALÁN J.G., CIRAUDO D., *Inteligencia artificial al servicio de la justicia penal, contravencional y de faltas. 'Prometea' en el ámbito de la ciudad de Buenos Aires*, en *Cibercrimen II*, dirigido por Dupuy D., Editorial B de f, Buenos Aires, 2018.
- CORVALÁN J.G., *Presentación: inteligencia artificial, automatización y predicciones en el Derecho*, en *Cibercrimen III*, dirigido por Dupuy D., Corvalán J.G., Editorial B de f, Buenos Aires, 2020.
- CUATRECASAS MONFORTE C., *La Inteligencia Artificial como herramienta de investigación criminal*, La Ley, Madrid, 2022.

- DE LA CUESTA AGUADO P., *Inteligencia artificial y Derecho Penal*, en «Revista Penal México», n. 16 y 17, 2019-2020.
- DEMETRIO CRESPO E., *El Derecho penal ante el desafío neurotecnológico y algorítmico: reflexiones preliminares*, en *Derecho Penal y comportamiento humano*, dirigido por Demetrio Crespo E., Tirant Lo Blanch, Valencia, 2022.
- DUPUY D., *Inteligencia artificial aplicada al Derecho Penal y Procesal Penal*, en *Ciberdelitos II*, dirigido por Dupuy D., Editorial B de f, Buenos Aires, 2018.
- HALLEVY G., *The Criminal Liability of Artificial Intelligence Entities from Science Fiction to Legal Social Control*, en «Akron Intellectual Property Journal», 2016.
- INDEPENDENT HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *A definition of AI: Main capabilities and disciplines*, European Commission, Bruselas, 2019.
- 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», n. 62-1, 2018.
- KING T., AGGARWAL N., TADDEO M., FLORIDI L., *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*, en «Science and Engineering Ethics», n. 26, 2020.
- MIRÓ LLINARES F., *El sistema penal ante la inteligencia artificial: actitudes, usos y retos*, en *Ciberdelitos III*, dirigido por Dupuy D., Corvalán J. G., Editorial B de f, Buenos Aires, 2020.
- MIRÓ LLINARES F., *Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots*, en «Revista de Derecho Penal y Criminología», n. 20, 2018.
- NILSSON N., *The Quest for Artificial Intelligence: A History of Ideas and Achievements*, Cambridge University Press, Cambridge, 2010.
- PALMA M<sup>a</sup> F., *El Derecho como universo de sentido para la interfaz cerebro e inteligencia artificial*, en «Derecho Penal y comportamiento humano», dirigido por Demetrio Crespo E., Valencia, 2022.
- PARLAMENTO EUROPEO, *Report on artificial intelligence in a digital age*, 2022. Recurso electrónico disponible en: <<https://www.europarl.europa.eu>>
- PÉREZ ARIAS J., *Ciberdelictividad: hacia la nueva realidad -virtual- del Derecho Penal*, en «Revista Electrónica de Doctrina y Jurisprudencia», n. 26, 2021.
- PÉREZ ARIAS J., *Algoritmos y Big Data en la responsabilidad penal: el reto de la ciberdelictividad en el Derecho penal*, en esta obra.

- ROMEO CASABONA C.M., *Criminal responsibility of robots and autonomous artificial intelligent systems?*, en «Comunicaciones en propiedad industrial y derecho de la competencia», n. 91, 2020.
- SALVADORI I., *Agentes artificiales, opacidad tecnológica y distribución de la responsabilidad penal*, en «Cuadernos de Política Criminal», n. 133, 2021.
- SANTOS GONZÁLEZ M<sup>a</sup> J., *Regulación legal de la robótica y la inteligencia artificial: retos de futuro*, en «Revista Jurídica de la Universidad de León», n. 4, 2017.
- VALLS PRIETO J., *Sobre la responsabilidad penal por la utilización de sistemas inteligentes*, en «Revista Electrónica de Ciencia Penal y Criminología», n. 24-27, 2022.