Universidad del País Vasco Euskal Herriko Unibertsitatea

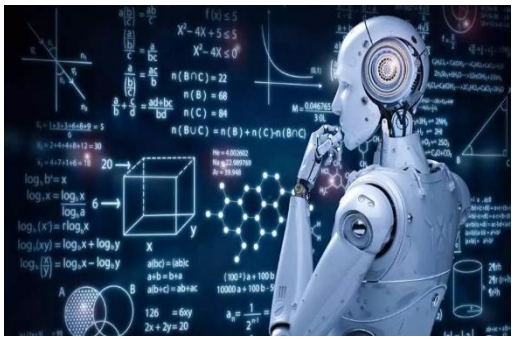
BILBOKO HEZKUNTZA FAKULTATEA
FACULTAD DE EDUCACIÓN DE BILBAO

Trabajo Fin de Grado

GRADO EN DERECHO

Curso 2020-2021

Inteligencia Artificial y Atribución de la Responsabilidad Penal



Autora / Autor: Ignacio Laguna Pérez

Directora / Director: Carlos María Romeo Casabona.

En Leioa, a 16 de junio de 2021

ÍNDICE

Inteligencia Artificial y Atribución de la Responsabilidad Penal

Ignacio Laguna Pérez
UPV/EHU

La Inteligencia Artificial está llamada a protagonizar uno de los mayores avances en la evolución de la especie humana. Hoy en día encontramos que sistemas IA se han implementado en determinados sectores de nuestra sociedad, como la industria, medicina… Su desarrollo es exponencial lo que implica que en escasos años veremos estas tecnologías en mayor medida, y más avanzadas. En este trabajo planteo los riesgos, implícitos y explícitos que estas tecnologías suponen y la manera de afrontarlos y gestionarlos a través del Derecho Penal. Las complicaciones que surgen para la atribución de la responsabilidad penal por delitos que cometen sistemas IA es una cuestión que ha de ser abordada, así como la implementación de un control humano significativo a estos sistemas.

1. Introducción

La Inteligencia Artificial y la robótica están llamadas a protagonizar una de las mayores transformaciones de la historia de la humanidad. Su desarrollo y evolución hacia sistemas más automatizados avanza a velocidad vertiginosa. Estos sistemas nos aportan grandes beneficios en nuestra sociedad, ambientales, sanitarios, industriales… aunque conllevan también un correlativo riesgo implícito, y a medida que avanza de forma exponencial la implementación de estos sistemas, cada vez más complejas, estamos más expuestos a dicho riesgo. Hemos escuchado a prestigiosos nombres mencionar esta cuestión: Bill Gates afirmó "que la inteligencia artificial es tan prometedora y peligrosa al mismo tiempo como la energía nuclear". (Merino, 2019). Igualmente, Elon Musk mostró su enorme preocupación ante el proyecto "*DeepMind*" de Google. Otros han llegado a ir más lejos, como Stephen Hawking que llegó a afirmar que "el desarrollo de la inteligencia artificial podría significar el fin de la raza humana" (Cellan-Jones, 2014).

Ante esta situación, el Derecho penal debe establecer mecanismos de garantías para proteger a todas las personas susceptibles de sufrir daños por estos sistemas. Del mismo

modo, establecer un sistema de prevención sólido para minimizar los riesgos. El presente trabajo tiene como objeto mostrar la situación en la que nos encontramos con la familia de tecnologías que forman los sistemas de Inteligencia Artificial, así como el correlativo riesgo que suponen de vulneración de bienes jurídicos protegidos penalmente por o a través de estos sistemas. Mostraré igualmente los mecanismos jurídico-penales que ostentamos actualmente para afrontar esta situación. Finalmente, establecer nuevos mecanismos jurídicos para fortalecer el control humano significativo sobre estos sistemas, cuya virtualidad reside en establecer garantías de carácter preventivo para aquellos sistemas que suponen un mayor riesgo, y que será la base de la individualización de la responsabilidad penal.

1.1 Desarrollo y adaptación del Derecho Penal al avance tecnológico.

La presencia que han tenido y siguen teniendo las tecnologías en el Derecho ha venido de la mano de la práctica jurídica. No obstante, los sistemas de inteligencia artificial, en lo sucesivo IA, están cada vez más presentes en nuestro día a día y plantean serias cuestiones a las que el Derecho ha de contestar; concretamente el Derecho Penal. A pesar de ser una cuestión novedosa, puesto que la robótica y la inteligencia artificial se encuentran ahora mismo experimentando un crecimiento exponencial, existen diversas hipótesis para configurar un sistema de atribución de responsabilidad penal por las lesiones a bienes jurídicos protegidos penalmente producidas por o a través de robótica o sistemas IA. En este trabajo veremos las distintas tesis que han planteado diversos autores.

La revolución tecnológica y digital, y en concreto la inteligencia artificial, aunque se encuentra en fase inicial, está llamada a protagonizar una transformación social como la que supuso en su día internet, puesto que se abre la posibilidad de que sistemas autónomos puedan adoptar decisiones al margen de la voluntad humana. Esto se traduce directamente en mayores márgenes de riesgo. Esta situación supone que existan nuevas formas de vulneración de bienes jurídicos a proteger por el Derecho Penal, como posibles delitos perpetrados por robots autónomos. Esto conlleva un especial desafío para el legislador en cuanto a cuestiones a resolver en torno a la atribución de la responsabilidad penal, ya que el Derecho suele ir detrás del avance tecnológico.

El avance tecnológico ofrece un aspecto sumamente positivo y progresivo en el desarrollo del ser humano, tanto en sus actividades como en el ámbito laboral, racionalizando y transformando la actividad laboral y sus medios. Ahora bien, hay que tener presente que toda transformación inexorable como la que supone la implementación de sistemas IA en nuestra sociedad, lleva implícita un riesgo. En caso de lesión a un bien jurídico protegido penalmente por un sistema de inteligencia artificial o valiéndose del mismo, habrá de atenderse a su uso, programación, desarrollo y objetivos a la ahora de atribuir la responsabilidad penal. Supone una especial complejidad jurídica si nos encontramos con un sistema de inteligencia artificial de desarrollo dinámico, autónomo y autodidacta. Con lo cual, además de las formas de atribución de la responsabilidad a personas físicas y jurídicas antes expuestas, hay que prestar especial atención a si un robot autónomo u otro sistema IA, puede ser penalmente responsable, y por ende, imponer una pena al mismo. En caso contrario, cómo individualizamos la responsabilidad penal.

1.2 Marco teórico de los sistemas IA y su interés jurídico penal.

> El progreso, desarrollo e innovación tecnológico ha experimentado un crecimiento exponencial durante el Siglo XX y XXI. La sexta revolución tecnológica, es donde ubicamos el desarrollo de los sistemas IA, junto a la biotecnología, ingeniería genética, ingeniería robótica, la ciencia cognitiva en asociación con las neurociencias… generan unas condiciones y presupuestos idóneos para que la humanidad experimente un evento al que llamaremos singularidad tecnológica, que implica entrar en un estadio de desarrollo de nuestras capacidades con límite desconocido para el curso de la evolución de la especie humana. El problema ético y jurídico que se plantea es el control humano sobre este desarrollo. (Domínguez & García-Vallejo, 2009, pág. 9)

Por ello, para tal evento, es imprescindible que exista una cooperación internacional e interdisciplinaria que permita integrar las máximas áreas científicas posibles, en las que la ciencia jurídica ocupa un papel fundamental. Para establecer un ámbito regulador para los sistemas inteligentes, en primer lugar, hemos de comenzar por su definición. Ésta ha sido abordada desde distintas áreas del conocimiento, siendo en ocasiones discordante unas con

otras. Sin embargo, podemos destacar dos perspectivas a la hora de concretar qué es la inteligencia artificial.

> La primera de ellas consiste en la tesis de Hayes, en la cual la IA es el estudio de la inteligencia como proceso, por lo que el objetivo principal de esta disciplina es la conducta inteligente y, en particular, la conducta humana. Se centra en el estudio de los procesos cognitivos, intentando obtener un desarrollo teórico sistematizado de las diversas actividades del intelecto que nos permitan un conocimiento más profundo y preciso de aquel. Desde la segunda, el fin de IA es la creación de sistemas automáticos, y así lo han definido tanto John McCarthy como Marvin Minsky, ambos considerados padres de esta disciplina. Para el primero, la IA «es la ciencia e ingenio de hacer máquinas inteligentes, especialmente programas de cómputo inteligentes» Minsky, por su parte, entiende la IA como «la ciencia de hacer que las máquinas hagan cosas que requerirían inteligencia si las hicieran las personas»; es decir, programar máquinas de forma que realicen tareas que, si fuesen llevadas a cabo por un ser humano, exigirían inteligencia por parte de la persona que las ejecuta. Actividades como leer un libro, conducir un coche o la comprensión de un lenguaje, se dice que requieren un cierto nivel de inteligencia. (Morales, 2021, pág. 6)

Estas definiciones nos otorgan luz en el área en el que se encuentran los sistemas de inteligencia artificial y robótica. Con todo ello, en el orden jurídico penal es necesario concretar de forma íntegra y específica de qué consiste la Inteligencia Artificial para que no dé lugar a lagunas normativas o injusticias. Es relevante igualmente de cara a establecer mecanismos legislativos suficientes para la gestión del riesgo que implica el desarrollo de determinados sistemas IA como los de alto riesgo. Por este motivo, la definición de Inteligencia Artificial que plantea la Comisión Europea en su propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de Inteligencia Artificial y se modifican determinados actos legislativos de la Unión del 21 de abril de 2021, es la más desarrollada a mi juicio hasta la fecha.

> La Comisión Europea expone que un sistema de IA consiste en “un programa informático desarrollado con una o varias de las técnicas y enfoques informáticos y que puede, para un conjunto determinado de objetivos definidos por el ser humano, generar resultados tales como contenidos, predicciones, recomendaciones o decisiones que influyen en los entornos con los que interactúan.

Junto a la definición, los sistemas de Inteligencia Artificial han sido categorizados en dos vertientes. En primer lugar, la Inteligencia Artificial "débil", y la Inteligencia Artificial "fuerte".

> Barona Vilar llega a hablar hasta de cuatro tipos de IA: sistemas que imitan cómo piensan los humanos, capaces de tomar decisiones, resolver problemas y con capacidad de aprendizaje; sistemas que actúan como humanos e imitan su comportamiento; sistemas que utilizan el pensamiento lógico racional humano, capaces de inferir una solución a un caso a partir de una información sobre un contexto dado; y sistemas que emulan la forma racional del comportamiento humano, como los sistemas inteligentes o expertos, de los que hablaré posteriormente. Todos ellos, no obstante, comparten una misma limitación: son un tipo de inteligencia específica, pero no de tipo general. Es importante el matiz de que no se trata de una inteligencia general (propia de los seres humanos), sino específica. El diseño y realización de inteligencias artificiales que únicamente muestran comportamientos inteligentes en un ámbito muy específico es lo que algunos autores han denominado «IA débil», en contraposición con la «IA Fuerte». Esta última implicaría que un ordenador convenientemente diseñado no simula una mente, sino que es una mente y, por consiguiente, debería ser capaz de tener una inteligencia igual o incluso superior a la humana. (Morales, 2021, págs. 6-7)

> De este modo, hablamos de IA débil para referirnos a programas que realizan tareas específicas sin necesidad de tener estados mentales. Buscar soluciones a fórmulas lógicas con muchas variables, la previsión meteorológica o los asistentes de voz (como "Siri" o "Cortana") son algunos ejemplos, todos ellos relacionados con la toma de decisiones. También se asocia con la IA débil el hecho de formular y probar hipótesis acerca de aspectos relacionados con la mente (por ejemplo, la capacidad para razonar deductivamente o de aprender inductivamente) mediante la construcción de programas que llevan a cabo dichas funciones. Estos programas, utilizados ampliamente en las últimas décadas, han proporcionado resultados satisfactorios, hasta tal punto que las voces más optimistas pronostican la posibilidad de que fuesen capaces de superar a la inteligencia humana e incluso reemplazarla en una actividad muy concreta (…). La diferencia radica, entonces, en el modo de procesar la información. En nuestro caso, la inteligencia humana, el cerebro procesa la información activando de forma coordinada las redes neuronales. Por

> su parte, en un sistema artificial, las instrucciones básicas son las propias de una computadora: operaciones aritmeticológicas, de lectura/escritura de registros y de control de flujo secuencial. (Morales, 2021, pág. 8)

Traigo a colación la partida de ajedrez entre el campeón mundial Garry Kaspárov y el "superordenador" de *IBM Deep Blue* en Nueva York, en mayo de 1997, en la que se impuso la IA. (Illescas, 2016). Puesto que, más allá del revuelo que ocasionó aquél suceso, el sistema *IBM Deep Blue* se trataba de una IA "débil", ya que la máquina pese a ser la mejor en su especialidad consistente en jugar al ajedrez, únicamente estaba programada para esta tarea.

> Todos los avances logrados hasta ahora en el campo de la IA son manifestaciones de IA débil y específica; y, entre estos sistemas de IA débil, destacan los denominados sistemas expertos o sistemas de conocimiento («*Knowledge Based System*»). Un sistema experto es aquel capaz de procesar y memorizar información, aprender y razonar en situaciones determinadas e inciertas, comunicarse con los hombres u otros sistemas expertos, tomar decisiones apropiadas y explicar por qué se han tomado tales decisiones. Un consultor que puede suministrar ayuda, o incluso sustituir, a los expertos humanos con un grado razonable de fiabilidad. Un sistema experto incluso puede obtener experiencia a partir de los datos disponibles, pero no dejará de ser un conocimiento de tipo específico, incapaz de actuar de manera inteligente más allá de las reglas previamente suministradas y basadas en situaciones concretas. Su hándicap será, siempre, no poder alcanzar un conocimiento de tipo general, que sí posee la inteligencia humana, y que es una inteligencia basada en módulos especializados para diferentes habilidades cognitivas. Módulos que no son otra cosa sino las diferentes adaptaciones que las personas hacemos a los problemas que surgen a lo largo de nuestra vida. Un conocimiento que proviene de millones de experimentos en los que participamos sin ser conscientes de ello, que denominamos «sentido común» y que vienen de la mano de nuestra capacidad sensorial. Podría decirse que sentido común y capacidad sensorial se retroalimentan, pues ese conocimiento es el que permite una completa comprensión del lenguaje y una interpretación profunda de lo que capta un sistema de percepción y, a su vez, el sistema de percepción es necesario para adquirir tal conocimiento. (Hidalgo, 1996, pág. 15)

En la actualidad, los avances en el campo de la IA "fuerte" son tan limitados que

siguen estando en una fase embrionaria. No obstante, esto no impide que se siga pensando en la posibilidad de un aprendizaje propio de las máquinas en un futuro, y en lo que, de un modo u otro, constituiría una IA autónoma. De ahí que ya se hayan comenzado a plantear los retos que tales sistemas de IA supondrían para el Derecho, así como las posibles soluciones que podrían darse. Será esta una de las cuestiones a analizar para abordar la relación entre la IA y el Derecho Penal.

> En el ámbito en el que nos encontramos, los sistemas IA, son de especial interés jurídico la llamada Inteligencia Artificial Fuerte. Por un lado, los sistemas *Machine learning"* basados en un aprendizaje automático e inductivo a partir de ejemplos, lo que conocemos como experiencia. Por otro lado, los sistemas más avanzados, las redes neuronales artificiales y su subcampo de aprendizaje profundo o "*Deep Learning"*, con la capacidad especialmente los dos últimos por los siguientes motivos. Las Redes Neuronales Artificiales, en adelante RNA, son modelos computacionales que procesan información imitando el funcionamiento de las neuronas biológicas. En este sentido, las redes artificiales están compuestas por nodos de entrada que reciben la información desde el exterior de la red (*input*), nodos de salida que transmiten y envían información hacia el exterior de la red (*output)*, y nodos ocultos que transmiten la información entre los nodos de la red. Estos últimos son llamados "capas de aprendizaje" y son los que predominan en las RNA, de forma que, a mayor cantidad de capas, mayor es la profundidad de la red y mayor es la capacidad de aprendizaje. En este contexto, los nodos de entrada reciben una serie de datos desde el exterior, estos datos son enviados al interior de la red hacia los nodos ocultos. Los nodos ocultos van procesando, modificando y transfiriendo la información de una capa a otra. Este proceso es lo que se conoce como "aprendizaje", pues cada capa de nodos ocultos va aprendiendo de las capas más externas. Dicha secuencia de aprendizaje es lo que da origen al *Deep Learning*. Cuando las redes neuronales son entrenadas, cada red crea, modifica o elimina conexiones entre los nodos con el fin de dar respuestas más acertadas ante el problema que busca resolver. (Lazcano, 2020)

La actual aparición de estos sistemas y su correspondiente industrialización e implementación en nuestra sociedad a corto-medio plazo es lo que conoceremos como la gran revolución de la Inteligencia Artificial. Los sistemas inteligentes autónomos, también conocidos como sistemas "superinteligentes" -*super-intelligent systems*-, suponen un gran

> paso en el desarrollo de estos sistemas, dada su capacidad de reprogramación a partir de la inferencia de información tomada de su entorno que son capaces de procesar y asimilar -aprender- por sí mismos (*deep learning* y *machine learning*). A partir de este aprendizaje pueden tomar decisiones, convirtiéndose así en auténticas entidades autónomas. Los sistemas "super-inteligentes" son capaces de aprender, proponer objetivos y construir planes para alcanzarlos más allá de las funciones específicas para las que sus diseñadores los programaron. (Romeo, 2020, pág. 169)

Este será el gran reto de la humanidad para los próximos años: cómo implementar estas tecnologías evitando al mismo tiempo que puedan causar perjuicios a los seres humanos y a sus bienes. Aspecto en el que el Derecho cobra un papel fundamental.

3. El nuevo reto para el Ordenamiento Jurídico penal que supone el desarrollo de sistemas IA y su creciente implementación en nuestra sociedad

> La inteligencia artificial, aunque de un modo incipiente, forma parte de la vida cotidiana, del mismo modo que la robótica. Vemos su creciente impacto en múltiples sectores de la actividad humana; por ejemplo, el aumento de la capacidad de producción industrial, el diagnóstico y tratamiento de enfermedades, en la previsión de daños contra la ciberseguridad, para combatir la crisis climática, y otras actividades humanas en expansión. (Romeo, 2020, pág. 169)

La influencia de la IA en los diferentes ámbitos de la sociedad humana, del mismo modo que su progresivo desarrollo en su autonomía conlleva un relativo riesgo y peligrosidad. La integración de sistemas inteligentes autónomos en nuestra sociedad puede suponer la comisión de delitos a través de estos sistemas. Del mismo modo, cabría cuestionarse la aparición de nuevos delitos, así como nuevos bienes jurídicos a proteger por el Derecho penal; de un modo semejante a lo que supuso la revolución digital y el cibercrimen.

Sin embargo, hay que tener especial cuidado en no incurrir en el "Derecho ficción". Es decir, hemos de analizar si hoy en día o a corto-medio plazo, los comportamientos de los sistemas

de inteligencia artificial, son susceptibles de llevar a cabo comportamientos que comprendan un ilícito penal. En virtud del principio de proporcionalidad, el Derecho Penal es un derecho subsidiario que ha de operar como última ratio, únicamente cuando el orden jurídico no pueda ser preservado y restaurado eficazmente mediante otras soluciones menos drásticas que la sanción penal.

> Con todo ello, existen argumentos a favor de la intervención del Derecho Penal como el de Miró Linares. La autora afirma que la IA disponible hoy en día ya plantea suficientes retos en relación con el Derecho Penal como para obviarlos. Se está refiriendo al uso que en la actualidad se da a la IA en la actuación policial para la prevención e investigación del delito, y en la determinación judicial y el tratamiento penitenciario. (Morales, 2021, pág. 10)

El principio de intervención del Derecho Penal incide en el desarrollo de la propuesta de Reglamento de la Comisión Europea, por el cual se plantean nuevas normas y acciones para la excelencia y la confianza en la Inteligencia Artificial en el territorio europeo. Los sistemas inteligentes autónomos suponen un riesgo susceptible de gestión a través del orden administrativo a través de sus respectivas sanciones. Del mismo modo, la propuesta antes nombrada, califica determinados sistemas inteligentes autónomos como inadmisibles o de riesgo muy alto. Los sistemas mencionados son susceptibles de lesionar bienes jurídicos protegidos, con lo cual, es pertinente la intervención del Derecho Penal en su "última ratio".

Con todo ello, hemos de tener en cuenta que el avance de la IA en sus diversas formas y aplicaciones es un logro de la creatividad humana a la que no estaremos dispuestos a renunciar, como ha ocurrido con otros avances tecnológicos anteriores. Por lo que del mismo modo que el Derecho Penal intervendrá bajo los principios de necesidad y proporcionalidad, el Derecho Administrativo hará lo propio a través de la gestión de riesgos. Este planteamiento se justifica puesto que una actuación de estos órdenes ilimitada y desproporcionada implicaría obstáculos en el avance e innovación tecnológica.

En todo caso, supone un reto para el legislador dar respuesta a nuevos conflictos e interrogantes que supone la implementación de sistemas IA. Para ello, es lícito apostar por una ampliación del ámbito penal de aplicación de tal manera que cubra la potencial vulneración de bienes jurídicos susceptibles de ser lesionados por sistemas IA. Todo ello, con

las precauciones propias que supone ampliar el campo legislativo del derecho penal sin incurrir en injusticias, lagunas normativas o directamente trabe e impida la transformación, desarrollo, innovación e investigación tecnológica. La propuesta de la Comisión Europea por la que se establecen normas armonizadas en materia de Inteligencia Artificial, hace hincapié en esta cuestión, cuando afirma *"esta propuesta presenta un enfoque regulador horizontal equilibrado y proporcionado que se limita a los requisitos mínimos necesarios para hacer frente a los riesgos y problemas los riesgos y problemas relacionados con la IA, sin restringir ni obstaculizar indebidamente el desarrollo tecnológico o que aumente de forma desproporcionada el coste de la comercialización de soluciones de IA en el mercado."*. (European Comission, 2021, pág. 3).

El legislador ha de estar a la altura de un reto como es la Sexta Revolución tecnológica, puesto que, como he explicado anteriormente, exige una cooperación interdisciplinar para su correcto desarrollo, sin que la ciencia jurídico-penal se vea rezagada, estableciendo un marco jurídico sólido y flexible. Sólida puesto que ha de definir de forma muy concreta el riesgo permitido, y establecer garantías suficientes para el cumplimiento de una serie de estándares por parte de los sistemas IA. Por otro lado, será flexible porque no debe crear restricciones innecesarias, el marco normativo tiene que responder al principio de proporcionalidad, lo que también implica que "*la intervención legal se adapte a aquellas situaciones concretas en las que exista un motivo justificado de preocupación*". (European Comission, 2021). Ha de comenzarse a tener en cuenta la robótica autónoma o sistemas IA con aprendizaje profundo como posibles medios para la comisión de delitos, así como atender al uso, programación y distribución de los mismos.

Sobre estas cuestiones relativas a la responsabilidad por las decisiones tomadas por sistemas IA se ha pronunciado el Parlamento europeo, la Comisión europea y el Consejo de Europa, es una cuestión de gran interés y extrema importancia a medida que la mejora de estas tecnologías crece en la industria y comienza a tener un impacto más directo en nuestra vida cotidiana. El Parlamento europeo también anunció que varias cuestiones relacionadas con la creación, el desarrollo y el funcionamiento de robots y sistemas de inteligencia artificial serán objeto de regulación legal.

Así, la Comisión Europea presentó el 21 de abril de 2021 una propuesta de Reglamento por el cual se plantean nuevas normas y acciones para la excelencia y la confianza en la Inteligencia

Artificial en el territorio europeo, de la que hablaré más adelante. Esto supone un primer marco jurídico sobre la Inteligencia Artificial. A su vez, presentó el Plan Coordinado 2021 sobre Inteligencia Artificial (IA), basado en la sólida colaboración entre la Comisión y los Estados miembros establecida durante el Plan Coordinado de 2018.

3. El riesgo de lesión de un bien jurídico penalmente protegido a través de sistemas inteligentes autónomos.

No cabe duda de que los robots y sistemas IA puedan lesionar bienes jurídicos protegidos por el Derecho Penal. Al no descartar que se produzca esta situación, hemos de plantearnos la atribución de responsabilidad penal a estos sistemas, especialmente a medida que desarrollan una creciente autonomía en la toma de decisiones. Para poder imputar la responsabilidad penal de un delito cometido a través de una IA en nuestro Ordenamiento jurídico penal, hemos de partir de dos cuestiones. En primer lugar, el tipo de IA con la que nos encontremos, y en segundo lugar, la forma de comisión del delito, a lo que habrá de atenderse a su uso, programación, diseño, fabricación y objetivos a la ahora de atribuir la responsabilidad penal.

La IA "débil" no debería presentar muchos problemas a la hora de atribuir la responsabilidad porque tienen un grado de autonomía limitado y están diseñadas para el ejercicio de una actividad concreta.

Determinados autores consideran la delincuencia a través de sistemas IA débiles como una evolución del cibercrimen. Con mayor frecuencia, las organizaciones delictivas comienzan a incorporar estos sistemas con el fin de facilitar sus actividades y maximizar los beneficios en el menor tiempo posible, beneficiándose también de la falta de conocimiento y control que existe aún en relación con estas nuevas herramientas delictivas. No obstante, en la medida en que esta IA débil consiste esencialmente en algoritmos de predicción utilizados para la ejecución de acciones o recomendaciones para actuar a partir de un conjunto de datos existente y en la que, por tanto, todo el contexto es otorgado por los seres humanos (que le brindan la información necesaria y que determinan su actuación),

> no requiere de ningún tipo de cambio esencial en el sistema de atribución de responsabilidad pensado para las personas físicas. En palabras de Quintero Olivares, esta posibilidad no supone ninguna dificultad, pues el que dispone de capacidad para programar una «actuación» de un robot (ya sea un humanoide o una máquina sin forma humana) con la finalidad de atacar algún objetivo, y pudiendo mantener el control hasta lograrlo, habrá cometido un delito que le será imputable sin lugar a duda. Seguiremos ante la delincuencia tradicional, pero que utiliza, ahora, estos sistemas como un instrumento más de comisión del delito.

Posteriormente analizaremos este tipo de supuestos en el epígrafe denominado "la IA como mero instrumento de las acciones humanas".

> Mayor repercusión presenta si se trata de sistemas que empiezan a tener cierta autonomía, esto es, sistemas dotados de IA fuerte. Son estos los que han llevado a la doctrina a plantearse importantes preguntas que, en último término, suponen un cambio en el modelo tradicional de atribución de responsabilidad penal, así como la búsqueda de nuevas respuestas jurídicas a los problemas que pueden llegar a surgir. (Morales, 2021, pág. 11)

A continuación, veremos una serie de supuestos en los que se puedan producir lesiones a bienes jurídicos protegidos por el Derecho Penal, para desentrañar con mayor exactitud el riesgo que suponen estos sistemas.

> Uno de los supuestos más habituales son los accidentes en plantas de producción relacionados con robots. Suceso que lo veremos con mayor regularidad en la medida en que se vaya implementando sistemas IA en los distintos sectores de la sociedad. Uno de los casos más recientes es el de la línea de montaje de Volkswagen en Wolfsburg, Alemania, en la cual un robot mató a un trabajador después de que lo agarrara por el pecho y lo aplastara contra una placa metálica. En este punto tras la apertura de la investigación por parte de la fiscalía, la cuestión a dilucidar era si se trataba de un fallo técnico del robot, o de un error humano. (Guelland, 2015)

Como vemos en este primer supuesto, no es sencillo definir el curso causal de la decisión tomada por la máquina. Ello obliga a que el Derecho Penal se cuestione si estos sistemas pueden llevar a cabo una conducta típica o si esto es solo posible por parte de los humanos.

Una hipótesis de riesgo podría ser la de un usuario de un sistema IA que sirviéndose del mismo cometiera algunos de los delitos tipificados en el CP. En tal caso, podría ser condenado como reo de este, con especial atención a su medio de comisión. Por ejemplo, si algún usuario de un dron dotado de Inteligencia Artificial con sistema "*Deep Learning"* se sirviera de él para matar a otro sujeto, debería ser condenado como reo de asesinato (138.1 del CP), o para causarle lesiones, con el agravante de uso de arma para ello (148.1 del CP). De esta forma podemos considerar al dron autónomo como un arma para la comisión del delito.

En todos estos casos nombrados anteriormente, cabría calificar a los robots como meros instrumentos del usuario. Esto quiere decir que el usuario, aun no realizando la acción típica de manera personal y directa, es responsable penal por realizar la acción típica por medio del sistema IA, valiéndose de la misma. Del mismo modo, es compleja la situación de la empresa distribuidora de dichos drones autónomos, que ha participado en el proceso de industrialización y distribución del sistema IA.

Sin poner en duda los beneficios explícitos que el diseño de estas tecnologías de carácter puntero va a suponer en nuestra sociedad, no podemos obviar sus riesgos implícitos. Especialmente en el aspecto de la ausencia de control o mando por parte de estos drones, ya que en caso de un fallo en el error de identificación de un enemigo, por ejemplo, implica dificultades en la atribución de la responsabilidad penal.

Surgen también interrogantes en caso de que el delito sea fruto de una imprudencia grave durante el proceso de fabricación y programación de la IA, con la mayor complejidad que supone que dicho sistema sea autónomo y autodidacta dotada de "RNA" o sistemas de *deep learning*. Esta situación implica reconocer cierta autonomía a las decisiones tomadas por los sistemas de IA, que ya no dependen de los humanos, o que no dependen significativamente de ellos. Ante esta situación de pérdida de dependencia del usuario o programador, total o parcial, sobre el sistema autónomo inteligente, se plantean cuestiones cruciales en términos de interés jurídico penal.

Así, ¿Deberían ser considerados penalmente responsables los seres humanos que mantienen de alguna manera deberes de supervisión y control sobre la IA y los robots?

¿Podría llegar a imputarse a los programadores del proyecto IA como reos de un delito de carácter imprudente, por dejación e incumplimiento de los deberes antes nombrados, que dé lugar a una lesión de un bien jurídico por mal funcionamiento de sus dispositivos? En tal caso, han de establecerse deberes de supervisión y control reforzados a los programadores de los sistemas autónomos inteligentes a través del Derecho Penal, y atender a las normas de la imprudencia punible del Ordenamiento jurídico penal (artículo 12 CP).

Igualmente, ¿Los sistemas inteligentes, especialmente los llamados autónomos, cumplirán siempre los requisitos asignados por los estudiosos del derecho y la jurisprudencia al concepto de delito y, por tanto, ¿se les podrá imputar penalmente? O, por el contrario, ¿hemos de acudir únicamente al Derecho Civil, a través de la responsabilidad civil extracontractual (1902 del CC), para indemnizar los daños causados por un sistema de IA, así como el Derecho Administrativo para gestionar los riesgos a través de sanciones? Es decir, ¿es suficiente conformarse con la reparación del daño y con la indemnización de los daños materiales y morales sufridos por la víctima, así como de una sanción administrativa impuesta al fabricante?

> Otro supuesto de posible vulneración de bienes jurídicos protegidos por sistemas de inteligencia artificial es el de cámaras dotadas de un sistema IA capaz de identificar el sexo de la persona, edad, color de piel, características de la ropa y hasta rasgos únicos, dotadas de un sistema de reconocimiento facial y machine learning. Vemos el caso de China que se ha llegado a dotar de 20 millones de estas cámaras como red de espionaje masivo (https://www.eldiario.es/tecnologia/sky-net-gran-hermano-china_1_3173654.html) (https://www.skynet.net/china). (Sarabia, 2017)
>
> También marcas como Facebook, tienen sus propios sistemas IA capaces de acceder a determinada información personal, a la cámara o al micrófono del dispositivo móvil, entre otras cosas, para ofrecer publicidad personalizada (en todo caso, accedemos a ello al aceptar los términos y condiciones del uso. (Schneier, 2018)

Esta tecnología, en caso de que cayera en las manos equivocadas son medios e instrumentos susceptibles de vulnerar bienes jurídicos como la intimidad o libertad informática (artículo 197 del CP) o en el caso más grave del artículo 584 del Código Penal, consistente en la

revelación de secretos. La recopilación de datos personales a través de la navegación en internet, la aceptación *Cookies*, la prestación del consentimiento de los términos y condiciones de servicios de determinadas aplicaciones… puede ser ampliada a través de sistemas de Inteligencia Artificial mediante herramientas de recopilación de datos. Esto puede implicar un aumento significativo en la capacidad de adquisición de información personal. Por ello, sería interesante atender a las deliberaciones del Consejo Europeo y del grupo de expertos de alto nivel sobre Inteligencia Artificial (AI HLEG) sobre protección de datos e intimidad en el contexto de los sistemas IA que expondré a continuación.

Entraña una especial complejidad jurídica el hipotético riesgo de un proyecto IA como el de un androide, dotado de una Red Neuronal Artificial con sistema deep learning, en cuyo proceso de autoaprendizaje, lleve a cabo una reprogramación que le lleve a la comisión de un determinado delito. Puede llegar a ser reprochable penalmente los actos y omisiones llevados a cabo en su proceso de programación, industrialización y fabricación por el siguiente motivo. Llevar a cabo un proyecto como éste, entraña riesgos, especialmente los sistemas IA categorizados por la Comisión Europea como potencialmente peligrosos. Por lo que el programador que lleve a cabo un determinado proyecto IA ha de ser conocedor de dichos riesgos, y deberá llevar a cabo un control significativo sobre el mismo. Ergo implica deberes reforzados de supervisión, vigilancia y control de su actividad atendiendo a las características concretas del sistema operativo inteligente. En caso de la omisión de dichos deberes, el artículo 31.bis del Código Penal atribuye la responsabilidad penal a las personas jurídicas, en nuestro caso concreto la empresa, usuaria o programadora, del proyecto IA. Si bien creo que debería preverse un tipo reforzado de responsabilidad de incumplimiento de los deberes de supervisión, control garantía.

Del mismo modo una persona física, como un programador, que con sus actos propios incumpliera los deberes antes descritos en la elaboración de su proyecto, podría ser responsable penal en grado de imprudencia grave.

> Dicho esto, cabe cuestionarse si el programador, o el órgano de administración de la empresa en cuestión, pueden llegar a ser exonerados de la responsabilidad penal, adoptando de manera eficaz e idónea todas las medidas de control, vigilancia y supervisión inherentes a la creación y uso de un proyecto IA; y en su caso a quién sería imputable dicha responsabilidad. (Romeo, 2020, pág. 175)

Lo anteriormente dicho, lo analizaré al abordar el modelo de *"Compliance"* inherente al control humano significativo.

Los antecedentes con los que contamos nos muestran como abordó el Ordenamiento jurídico penal la revolución que supuso en su día internet, a través del cibercrimen. A nivel legislativo, configuró el cibercrimen a través de "parches", dejando determinadas lagunas normativas que han supuesto algunos problemas en el proceso penal. Tenemos como referencia esta metodología legislativa para afrontar este nuevo reto que supone la implementación de la IA en nuestra sociedad. La comisión de delitos a través de sistemas IA se puede abordar, bien, a través de una reforma más profunda de la legislación penal. Me inclino por la última opción de realizar una reforma de profundidad con mecanismos reforzados de garantía, o que la IA pueda ser sujeto responsable del delito.

4. Mecanismos actuales de atribución de la Responsabilidad penal por delitos cometidos a través de Robots o IA en nuestro Ordenamiento Jurídico penal.

En el Ordenamiento jurídico penal español existen distintos mecanismos para atribuir la responsabilidad penal por posibles delitos cometidos por sistemas autónomos IA o a través de ellos, como veíamos en los ejemplos del anterior epígrafe. Asimismo, para imputar la responsabilidad penal por la comisión de delitos por sistemas de inteligencia artificial, "atenderemos a su programación, diseño, fabricación y objetivos, así como el tipo de IA con el que nos encontremos." (Romeo, 2020, pág. 168)

Nuestro Derecho Penal tiene material legislativo para imputar penalmente al usuario que cometa un delito doloso perpetrado a través de un sistema IA, bien sea persona física o jurídica. En el caso del usuario, consistiría en que, bien adquiriendo en el mercado o fabricando él mismo el sistema IA, se valiese de él para la comisión de un delito. Sin embargo, la cuestión es si un robot puede llegar a ser responsable criminalmente en concepto de autor directo. A continuación, me dispongo a analizar la responsabilidad criminal de robots y/o sus programadores o usuarios tal y como lo concibe nuestro Derecho Penal (artículos 27, 28 y 29 del CP)

4.1. La responsabilidad penal directa de los sistemas de inteligencia artificial, en concepto de autor directo. El robot o el sistema de inteligencia artificial como agente criminal.

Existen perspectivas desde todos los ámbitos. Una de ellas, sostiene que la inteligencia artificial "fuerte" o "profunda", puede ser también directamente responsable de los delitos que lleve a cabo por sus actos propios. En este modelo de responsabilidad penal no imputamos el delito a la persona que se encuentra detrás del sistema, sino al propio sistema en sí. Esta teoría plantea atribuir a determinados sistemas de inteligencia artificial la condición de sujeto activo de un delito o como agente criminal. Los autores que defienden esta tesis afirman que sistema únicamente se podría dar en dos supuestos. El primero consistiría en que las personas encargadas del sistema de Inteligencia Artificial *Fuerte*, programadores, ingenieros, usuarios…, sufren una pérdida de control sobre el sistema de forma que son total y absolutamente incapaces de prever el resultado lesivo. Es necesario que esa pérdida de control sobre el sistema no sea por culpa de las personas antes mencionadas, para que no sean criminalmente responsables por imprudencia estas personas intervinientes en el proceso. El segundo consistiría en aquellos delitos en los que en los que el sistema de IA que causa la lesión ha sido programado, a su vez, por otro dispositivo con las mismas características (es decir, detrás de él no habría ninguna persona humana, sino otra máquina). Con todo ya se observa el riesgo de disociar en la relación de causalidad los resultados delictivos de las conductas humanas, dejando impunes ciertas conductas que entrañen incluso un cierto dolo eventual.

Para dar respuesta a esta cuestión debemos abordar la estructura del delito conforme a nuestro Ordenamiento y analizar si encaja en el esquema que tenemos actualmente, o si es necesario realizar cambios significativos en la estructura del delito y, por consiguiente, en las fuentes de imputación jurídico-penales, para atribuirles responsabilidad. Para ello veremos el esquema que tenemos actualmente.

Son delitos las acciones y omisiones dolosas o imprudentes penadas por la ley (artículo 10 del CP) llevada a cabo una persona física o una persona jurídica. Del mismo modo, hemos de acudir a los elementos del delito en nuestro Derecho Penal como acción, típica, antijurídica,

culpable y punible. Suscita interrogantes entorno al aspecto cognitivo y volitivo propios de la culpabilidad, y necesarios para atribuir la responsabilidad.

Gabriel Hallevy, con todas las salvedades que implica que lo analice bajo la perspectiva del derecho anglosajón, desentraña los delitos cometidos por sistemas de Inteligencia Artificial en tres elementos; un primer elemento externo, un elemento positivo y un elemento negativo. A continuación, me dispongo a mencionar sus ideas y analizar cómo podemos trasladarlas a nuestro Derecho continental.

En lo que al elemento externo respecta, es lo que nosotros entendemos en los elementos del delito como acción. Me dispongo a analizar en primer lugar la estructura general del mismo, y posteriormente sus componentes.

> La estructura general de la acción la comporta el elemento fáctico, los hechos que involucran a un sistema IA, y que son constitutivos de un delito que da lugar a la responsabilidad penal, hablamos de una conducta delictiva del sistema IA. De esta conducta delictiva derivará una responsabilidad penal en la que analizaremos la forma en la que se ha perpetrado el delito. (Hallevy, 2015, pág. 47). Este elemento es fácilmente atribuible a los Sistemas IA, puesto que únicamente precisaría que el sistema controle por sí mismo su mecanismo general de movimiento, o el movimiento de alguna de sus partes, para que cualquier acto pueda considerarse realizado por el sistema de IA. (Hallevy, 2015, pág. 60)

> Cuando una máquina (por ejemplo, un robot equipado con tecnología de inteligencia artificial) mueve sus brazos u otros dispositivos suyos, se considera que actúa. Por lo tanto, si se diera el caso de que al accionar su brazo golpeara a una persona que esté cerca (por ejemplo, un trabajador si el robot se encuentra en una fábrica) causándole lesiones, dicho movimiento y su resultado serían la manifestación del elemento externo del delito de lesiones. Esto es correcto cuando el movimiento es el resultado de los cálculos internos de la máquina, pero no sólo entonces. Incluso si la máquina es manejada en su totalidad por un operador humano a través de un control remoto, cualquier movimiento de la máquina se considera un acto. Hemos de decir que esto no implica necesariamente que las máquinas sean criminalmente responsables por dichos actos, puesto que para imponer

> responsabilidad penal hemos de atender también al elemento subjetivo-interno, el mental, lo que en nuestro derecho conocemos como el elemento volitivo. (Hallevy, 2015, págs. 60-63)

La conducta como elemento inherente del elemento externo puede manifestarse a través de una acción como hemos visto, pero también a través de una omisión. Estas formas de conducta han de ser analizadas en función de los atributos, capacidades y características con las que fue diseñada la máquina. La tecnología dotada de inteligencia artificial es capaz de realizar "actos" que satisfacen el requisito de conducta; no sólo la IA fuerte, sino también para tecnologías mucho más bajas.

> Del mismo modo que una acción como levantar un brazo, desplazarse de carril… una máquina dotada de IA, puede ser medio susceptible de llevar a cabo delitos cometidos por omisión, en el que se castigue su inacción en un deber legítimo de actuar. No hay duda de que cualquier máquina es capaz de no hacer nada, por lo que cualquier máquina es físicamente capaz de llevar a cabo una omisión. El "*actus reus"* vendrá constituido ahora, por la ausencia de actuación, pues el fin último que tenía el sistema era actuar de una determinada manera. De esta forma vemos satisfecho el requisito de conducta. Si además tuviera voluntad propia, habría que plantearse si los sistemas de inteligencia artificial debieran constituirse como sujetos responsables criminalmente; es decir si en dichos sistemas concurre el elemento volitivo. (Morales, 2021, pág. 20)

> Visto el elemento externo, veremos el elemento positivo en el marco de las inteligencias artificiales, cuyo fundamento se basa en el elemento "mental" del delito. En el Ordenamiento jurídico-penal español, ubicamos dicho elemento en la culpabilidad, elemento del delito que establece que para poder atribuir a alguien de un hecho delictivo tiene que ser imputable (*nullum crimen sine culpa*). A continuación, veremos como la imputabilidad a robots autónomos o sistemas IA "fuertes" entraña una especial complejidad en el marco de los delitos dolosos. El motivo de esto es que, a día de hoy, es cuanto menos cuestionable que sistemas IA estén dotados de una auténtica conciencia artificial, que les permita tener deseo de realizar el acto delictivo con conocimiento de su antijuricidad. El reflejo de la intencionalidad en el elemento mental abarca aspectos volitivos en cuanto al propósito delictivo, y cognitivos que apoyan el mismo. El aspecto cognitivo de la intencionalidad requiere conciencia, entendiendo ésta en el término

jurídico penal como la percepción por los sentidos de los datos fácticos y su comprensión. (Hallevy, 2015, págs. 67-68)

Si hemos de referirnos a la percepción por los sentidos de los datos fácticos, entendemos que el Derecho Penal se refiere a la percepción sensorial humana. Sin embargo, el desarrollo de la ingeniería robótica e ingeniería de software permite hoy en día que la tecnología de los sistemas IA equipados con los dispositivos pertinentes, perciban incluso más datos fácticos que a través de los 5 sentidos humanos. Veremos que un sistema IA podrá procesar imágenes, sonido, presiones, temperatura, humedad… mediante cámaras, sonido y los respectivos sensores. Dichos datos son absorbidos con gran precisión como hemos dicho, y los transfieren a los procesadores correspondientes. Por consiguiente, los sistemas IA cumplen con creces con esta primera etapa de la conciencia que corresponde el aspecto cognitivo. La segunda etapa consiste en tener una percepción completa del entorno analizando dichos datos. Las IA no poseen un cerebro biológico para ello, pero se trata de analizar si su procesador incorporado es óptimo para dicha función, como las Redes Neuronales Artificiales que nombraba anteriormente. De esta manera, la Inteligencia Artificial cumpliría la segunda etapa de la conciencia en términos de Derecho Penal, pudiendo considerar que las máquinas dotadas de IA ostentan la misma, siendo lo realmente relevante para la imputación de la responsabilidad penal; pese a que entendamos que no tienen conciencia en sentido amplio. (Hallevy, 2015, págs. 86-93)

En lo que a la volición respecta, hemos de averiguar si podemos atribuir los distintos niveles de voluntad exigibles en la culpabilidad para atribuir la responsabilidad a los sistemas IA. Al igual que con la conciencia, hemos de entender la voluntad en términos jurídico-penales. En efecto, la intención implica la voluntad de realizar una acción calificada como delito por el Derecho Penal, además de la conciencia de realizar esa acción. Pese a que es cierto que un sistema de IA puede estar programado para tener un fin o propósito y ejecutar acciones para alcanzarlo, cuando hablamos de específica intención a la hora de cometer un delito nos estamos refiriendo a la existencia en el sujeto activo de sentimientos o estados mentales que le mueven a actuar de una determinada manera. Sentimientos como el amor, el odio, la envidia, rencor, etc., y que, hoy en día, no existen en ningún sistema de IA. Por otro lado, el actuar culpable (*'mens rea'*) presupone la capacidad del acusado de actuar de forma diferente a como lo hizo y de ser susceptible de recibir un reproche legal por haber actuado ilícitamente, ya que, por el contrario, se ha

> demostrado que el acusado podría haber actuado conforme a la ley. En definitiva, de forma distinta a la acción u omisión, dolosa o imprudente penada por la ley. (Hallevy, 2015, págs. 93-103)

Es de tener en cuenta que ciertas IA, aunque todavía en desarrollo, están dotadas de Redes Neuronales Artificiales, completadas con un sistema de "*deep learning*". Estas máquinas tratan de simular el cerebro humano, como explicaba en el epígrafe 2.b, tienen la capacidad de evaluar distintos escenarios y actuar de una manera u otra en consecuencia. De igual manera el sistema *"machine learning"* permite un aprendizaje inductivo del ordenador a partir de ejemplos, basado en la experiencia. Esto influirá también en el proceso de toma de decisiones por parte de la IA. En todo caso, el comportamiento de las tecnologías dotadas de IA orientado a un determinado objetivo será aquel al que ha sido programado. Esto cuadra con la regla de previsibilidad de la que parte la voluntad penal.

Un sistema de inteligencia artificial "fuerte", tiene la capacidad de evaluar las distintas probabilidades y opciones de conducta, y actuar en base a ello, tras procesar toda la información posible de su entorno. De tal forma, si su conducta y el resultado de esta constituye un delito, entendemos que la IA tenía intención de cometer el mismo. Con mayor motivo, si el ordenador tiene la capacidad de evaluar la probabilidad con más precisión que el ser humano, podríamos concluir que la IA era consciente de su actividad delictiva. Así pues, concluyo que el aspecto volitivo puede cumplirlo un ordenador siempre que esté dotado de un sistema de RNA (Redes Neuronales Artificiales) con *deep learning*, lo que también conocemos como inteligencia artificial fuerte.

De esta manera, hemos dicho que la robótica autónoma y ordenadores dotados de IA, pueden llevar a cabo acciones u omisiones que revistan un carácter delictivo de manera culpable y consciente en términos penales, cumpliéndose los elementos externos y positivos, tal y como los reviste Gabriel Hallevy. En términos de nuestro Ordenamiento Jurídico-penal implica el cumplimiento de una acción u omisión culpable.

En cualquier caso, podemos pensar que los conceptos de conciencia y voluntad que son atribuidos a la IA son superficiales como para achacar al robot autónomo el daño a un bien jurídico. Hemos de tener en cuenta que tanto personas físicas como jurídicas están sujetas al derecho penal, pero no hay ningún interés en otorgar personalidad a las IA, con el riesgo

inherente de desvincular las acciones de la IA de sus operadores o usuarios. Sin embargo, sin necesidad de un cambio conceptual de las premisas que establece el Derecho Penal para la imputación de la responsabilidad penal, nos encontramos con que dichos conceptos son aplicables a las IA, como he expuesto anteriormente.

Por consiguiente, cabría preguntarse si la puerta para la imposición de la responsabilidad penal a la tecnología de inteligencia artificial como delincuentes directos estaría, de algún modo abierta, en sus distintas formas de autoría, siempre que se cumplan todos los elementos del delito.

> La cuestión consiste en determinar si estos requisitos fundamentales, concebidos para humanos y aplicados durante siglos a ellos, son plenamente transferibles a los robots y a los sistemas autónomos inteligentes. No existe una única respuesta a esta cuestión y se plantean distintas posibilidades para afrontar esta situación. En primer lugar, la propuesta de Gabriel Hallevy es la de transferir los requisitos fundamentales para la atribución de la responsabilidad penal, a los robots y la IA “fuerte”, de forma que podrían ser considerados como autores de un delito por sus actos propios, es decir, considerarlos como agentes criminales. (Romeo, 2020, pág. 170)

> Para tal consideración, sería necesario revisar las características legales actuales de la teoría del delito y adaptarlas para dar cabida a la responsabilidad penal de los sistemas de inteligencia artificial. Sin embargo, no es menos cierto que esta modificación, para satisfacer determinados elementos, podría "contaminar" la concepción y los requisitos de la actual teoría del delito aplicable a los seres humanos, con el probable efecto de relajar algunas de esas exigencias o incluso de prescindir de ellas, en detrimento de las garantías que conllevan. Cuestiones relativas a la culpabilidad, por ejemplo, en cuanto a la intencionalidad, conciencia y voluntad, se verían cuanto menos simplificadas. (Romeo, 2020, pág. 171)

Otra cuestión debatible para proyectar esta teoría en el Ordenamiento jurídico es la de atribuir personalidad a estos sistemas autónomos inteligentes, una suerte de “personas artificiales”, del mismo modo que se hizo con las personas jurídicas, para que se encuentren sujetos a responsabilidad penal. El ordenamiento jurídico confiere como sujetos activos y pasivos de derecho penal a la persona física y jurídica desde la reforma del año 2015. El ordenamiento

jurídico penal ha tenido en cuenta para configurarlos, los rasgos propios de aquéllas, en torno a su cuerpo y "mente", y así mismo, a la colectividad y métodos de convivencia en los que se desenvuelven los bienes jurídicos a proteger.

> Autores que han analizado esta cuestión previamente como Gabriel Hallevy y África Morales son partidarios de atribuir, en mayor o menor medida, un cierto grado de personalidad a las IA autónomas y, por consiguiente, su posible responsabilidad penal. Existen incluso proyectos "*RoboLaw"* cuyo objetivo consiste en la creación de un "Derecho Robótico". (Morales, 2021, pág. 10)

Hay quienes pueden considerar válida esta teoría, al ser más equiparable un robot a un ser humano que una sociedad mercantil. En ambos casos nos encontramos ante ficciones jurídicas que realiza el Derecho Penal para dar respuesta a determinadas conductas delictivas. Hablábamos antes de la cognición y volición, e incluso una oportunidad para renovar y/o modernizar el Derecho Penal mediante una transformación profunda. Sin embargo, es cuestionable a mi parecer, que sea esta la dirección en la que ha de transcurrir el Derecho Penal. El orden penal se basa en principios fundamentales como la seguridad jurídica, el principio de necesidad, el principio de intervención... y si bien podemos llegar a la conclusión de que es procedente realizar ficciones para las personas jurídicas, puede parecer desproporcionado hacer lo propio con los sistemas de inteligencia artificial. No debemos olvidar que el principio de seguridad jurídica en el Derecho Penal implica realizar las menores ficciones posibles. Y en virtud del principio de necesidad, es cuanto menos opinable que en la actualidad nos encontremos con sistemas inteligentes que tengan una actividad potencialmente delictiva y de forma totalmente autónoma a corto plazo. Con lo cual, no se aprecia una necesidad imperiosa de dotar a la IA de personalidad a fin de atribuirles responsabilidad penal, dado que se encuentra aún en fase inicial, y primigenia, teniendo un carácter más bien instrumental. No hemos de olvidar que hay quienes consideran que aún estamos lejos de lo que constituiría una verdadera IA autónoma. Por ello, también habría que valorar el coste que podría suponer para la estabilidad del actual sistema penal dominante introducir modificaciones en la estructura jurídica del delito para dar cabida a la responsabilidad penal de los sistemas inteligentes.

,

> No obstante, no debemos olvidar que el Derecho, y el Derecho Penal en especial, es una creación humana, un instrumento que se utiliza, entre otras cosas, para asegurar una vida

> social pacífica. En otras palabras, esta poderosa y sofisticada herramienta jurídica puesta al servicio de la sociedad, puede subordinarse a las necesidades humanas de cada momento histórico y, por tanto, adaptarse y modificarse según las necesidades. (Romeo, 2020, pág. 171)

Como última cuestión a debatir también deberíamos preguntarnos si no sería mejor buscar otros medios de imputación penal dirigidos directa o indirectamente contra seres humanos, siempre que cumplan conjuntamente todos los requisitos legales -objetivos y subjetivos-, aunque los sistemas IA relacionados con ella también puedan verse afectados por las consecuencias legales del delito y las respuestas al mismo. La primera cuestión que se plantea es si las normas jurídicas actuales del derecho interno o de derecho comparado podrían aplicarse a los hechos realizados por robots o por sistemas inteligentes autónomos a través de sistemas de red que pueden provocar la muerte o lesiones de seres humanos o a daños relevantes en las cosas, incluyendo en este último caso otras formas de vida no humana y los ecosistemas.

A continuación, me dispongo a mencionar una de estas posibilidades que ofrece nuestro Ordenamiento jurídico penal como método de imputación de la responsabilidad a los seres humanos por actos u omisiones llevados a cabo por sistemas de inteligencia artificial que revistan carácter delictivo

4.2. La responsabilidad penal del programador o usuario. El robot o sistema de IA como mero instrumento de las acciones humanas.

En el presente epígrafe me dispongo a analizar la posibilidad de considerar como responsable penal al usuario o programador de un sistema de Inteligencia artificial que se sirve del mismo para llevar a cabo una conducta delictiva.

Cabe mencionar en la teoría de la autoría mediata de Hallevy, a mi parecer imprecisa para llevar a cabo la atribución de personalidad al sistema IA en cuestión; puesto que esta forma de autoría está pensada en nuestro Derecho para los seres humanos: tanto el autor mediato como el instrumento del que se ha servido son personas humanas. Hallevy considera que los usuarios y programadores en estos supuestos serían responsables en concepto de autores

mediatos. Considerando las acciones físicas cometidas por la tecnología de inteligencia artificial como si hubieran sido del programador, del usuario o de cualquier otra persona que utilice instrumentalmente la tecnología de inteligencia artificial. En este escenario no se atribuye a la tecnología de inteligencia artificial ningún atributo mental, necesario para la imposición de la responsabilidad penal, sino que se le imputa directamente al autor mediato esta intención delictiva.

Sin embargo, desde mi perspectiva no procede enmarcar en estos casos a los programadores y usuarios que se sirven de la IA como autores mediatos. En nuestro Derecho Penal concebimos el autor mediato como el que se sirve de otra persona para cometer un delito evitando ejecutarlo él directamente, y para lo cual utiliza al autor como mero instrumento, beneficiando el error o el dolo de éste, o sometiéndolo a violencia física o psicológica o a coacciones o amenazas, o utilizando organizaciones jerárquicas de poder. En este caso, la subordinación del sistema IA al plan de programación hecha por el ingeniero correspondiente, no es subsumible al supuesto arriba enunciado. Puesto que no cabe entenderse como una organización jerárquica interpersonal, ni concurren elementos subjetivos como el dolo o la violencia.

Una vez vista la responsabilidad penal directa, veremos la indirecta. La tecnología dotada de IA puede ser utilizada por otro como mero instrumento para la comisión de un delito. En el plano de la IA, consistiría en una persona que, de forma consciente, ejecuta un plan delictivo, realizando la conducta tipificada por el Derecho Penal, mediante el uso instrumental de un ordenador con inteligencia artificial.

En este caso hemos de considerar como autor del delito al que lo orquesta, y al robot como herramienta o instrumento sofisticado para la comisión de este. Este autor podrá ser el programador de software que diseñe una IA para cometer delitos a través de ella. Puede ocurrir también que el responsable penal como autor sea el usuario final de la tecnología IA.

> Esta teoría, aunque desde mi perspectiva es acertada, plantea determinados interrogantes. Uno de ellos consiste en la concreción del lugar de comisión del delito.” Puesto que nos encontramos en un supuesto en el que la IA es un instrumento sometido al control remoto de un usuario que se puede encontrar en un lugar completamente distinto al lugar en el que

> la IA o el robot realizan la acción. Se sumará a este problema el hecho de que, en función del lugar del mundo en el que nos encontremos, los valores, las normas sociales de conducta y lo legalmente permitido o prohibido variará. (Morales, 2021, pág. 12)

Todo ello sin perjuicio de que también se puedan imponer medidas contra el sistema inteligente para bloquear un riesgo objetivo de reincidencia. Nuestro Derecho penal prevé las llamadas penas accesorias. En los delitos dolosos (artículo 127.1 del Código Penal) como en los imprudentes con pena privativa de libertad superior a un año (artículo 127.2) está prevista la figura del decomiso susceptible de ser aplicada en aquellos delitos en los que la Inteligencia Artificial sirva de medio o instrumento. Una vez llevado a cabo el decomiso, el Juez en función de las circunstancias del caso y su gravedad, podrá llegar a ordenar la suspensión del proyecto, desconexión de la IA, o incluso su destrucción. Así mismo, encontramos penas accesorias de las personas jurídicas (artículo 33.7 del CP, letras *C* a *G*) que podrían servir de inspiración en una eventual reforma del Código Penal para los sistemas IA.

Debemos señalar que en los supuestos en los que la IA toma la decisión por sí misma basándose en su propia experiencia o conocimientos acumulados, o basada en cálculos avanzados de probabilidades, con una total ausencia de control operabilidad sobre el sistema no permite imputar responsabilidad penal por delitos dolosos a todos los sujetos intervinientes en el proceso de desarrollo del sistema. Eso sí, siempre que hayan actuado con toda la diligencia posible y pese a ello, no hayan podido evitar la conducta delictiva de la IA. En todo caso cabría preguntarse si es posible la individualización de la responsabilidad penal en el programador o en la empresa fabricante por imprudencia al incumplir un deber de cuidado, al entender que en la fabricación de dichos sistemas, ostentan unos deberes reforzados cercanos a la responsabilidad objetiva. Otra opción sería la vía de la Comisión Europea de prohibir una serie de sistemas IA por considerarlos contrarios a los principios de la Unión, y a su vez establecer un delito de riesgo por la mera creación de estos sistemas, como desarrollaré más adelante.

A continuación, veremos el supuesto antes mencionado en el que el programador o la empresa es susceptible de responder penalmente por imprudencia.

4.3. Responsabilidad penal por imprudencia o dolo eventual del programador o de la empresa.

> La imprudencia tiene como una de sus premisas la realización de una conducta que no responde al cuidado objetivamente exigible en el sector social en el que se ha realizado la hipotética acción castigada. Esta postura se fundamenta en la infracción de un deber de garante que nace en el programador al crear la fuente de riesgo que sería el propio sistema. Esto supondría partir de la inobservancia de una norma de cuidado, lo que, a su vez, implica tener en cuenta la conciencia mínima del riesgo y el principio de precaución. (Romeo, 2020, pág. 172)

> Aunque sea de mero carácter especulativo, existe la posibilidad, cada vez mayor, que un sistema inteligente pudiera cometer un daño ilícito a un bien jurídico penalmente protegido debido a un error o fallo en el momento de valorar la forma de realizar dicha acción. Un primer supuesto es el de un sistema de inteligencia artificial que, por falta de información suficiente en su procesador, le impide tomar la decisión más correcta, no pudiendo por ello prever las consecuencias perjudiciales o los riesgos implícitos en su actuación. Finalmente, la conducta del sistema de inteligencia artificial reviste un carácter delictivo. La falta de previsión en la toma de decisión del sistema autónomo podría deberse a un diseño defectuoso durante el proceso de programación, que no debería permitirle derivar una decisión o una propuesta de actuación sin disponer previamente de todos o suficientes elementos de valoración de la situación. (Hallevy, 2015, págs. 124-130)

La determinación del error humano que da lugar a la acción delictiva del sistema inteligente es esencial para la individualización de la responsabilidad penal. Por ello, el proceso de industrialización del sistema autónomo inteligente, más concretamente durante su programación, en la que participan los correspondientes ingenieros de software, estos ostentarán deberes de cuidado reforzados. Por consiguiente, hemos de entender que el cuidado objetivamente exigible a un programador de software que participa en el diseño y programación del sistema autónomo inteligente consiste en que dicho sistema sea absolutamente incapaz de actuar de forma contraria a las normas penales.

En virtud del principio de tipicidad, hemos de identificar si la conducta de la que resulta el daño o lesión producido constituye lo que conocemos como imprudencia punible. Para ello,

habrá que comprobar si se cumplen los elementos de ese delito, en particular la infracción de la diligencia debida. En este punto, también es necesario recordar la importancia del llamado "riesgo permitido" en la modulación del cuidado debido. Hemos de partir del límite máximo del riesgo razonable. Encontramos como referencia para establecer parámetros de dicho riesgo permitido, la propuesta de reglamento de la Comisión Europea que establece cinco niveles de riesgo de los sistemas de inteligencia artificial. La mencionada propuesta clasifica determinados sistemas autónomos inteligentes como de "riesgo inadmisible", criterio que nos serviría para el establecimiento del límite máximo del riesgo razonable. Esto es relevante puesto que, en caso de que el ingeniero en cuestión asuma el riesgo conscientemente, y dé la posibilidad de reprogramación al sistema inteligente, que determine la pérdida de control relevante y significativo sobre este, nos encontraríamos ante una responsabilidad dolosa en caso de haber tipificado dicha conducta como un delito de riesgo.

No debemos olvidar que el Derecho Penal debe cumplir su función preventiva, lo que es especialmente aconsejable, y a mi parecer también es factible en relación con estas tecnologías. Por lo que, en virtud del desarrollo exponencial de los sistemas IA, a medio plazo se implementará un marco regulativo desde distintos órdenes, no sólo el penal. Es indudable que el Derecho administrativo jugará un papel importante, mediante la intervención de las autoridades administrativas mediante normas que coordinen el funcionamiento permitido de estos sistemas de IA en función de sus características específicas. Mediante este marco regulativo se inscribe el espacio de riesgo permitido con respecto a las decisiones autónomas de los sistemas de inteligencia artificial.

> Esto significa que el marco en el que se inscribe el espacio de riesgo permitido en relación con las decisiones e intervenciones basadas en la IA debe construirse con una estructura cercana al modelo de "*compliance*", que constituye la base de la responsabilidad penal de las personas jurídicas. En el ámbito del delito imprudente, el modelo de "*compliance*", el cual desarrollaré más adelante, consistiría en primer lugar, en la existencia de estructuras organizativas y de gestión de riesgos que impidan la existencia o el desarrollo de una red que permita la comisión de delitos (objetivo esencial en relación con las personas jurídicas). En segundo lugar, dichas estructuras y procedimientos deben impedir la toma de decisiones autónomas que impliquen la desviación del plan inicial de programación. Es decir, que su capacidad de aprendizaje autónomo, no le lleve bajo ningún concepto a la toma de decisiones que revistan carácter delictivo. (Romeo, 2020, pág. 176)

5. Sistemas autónomos inteligentes en el ámbito de la ingeniería militar.

No podemos obviar que, entre los sistemas de inteligencia artificial, los relativos al ámbito militar son de los más desarrollados hasta la fecha. Buena parte del gasto en defensa de los Estados está destinado a financiar determinados sistemas autónomos para sus respectivos ejércitos, bien diseñándolos, o bien adquiriéndolos a terceros. Vemos que solo Estados Unidos supera los 1000 millones de dólares de inversión al año. (Fernández, 2021) Esto supone que hemos de establecer un marco regulativo, que establezca aún más garantías si cabe en el ámbito militar, puesto que no serviría de nada si establecemos garantías y procesos de gestión de riesgos en el ámbito civil, mientras que en el campo militar donde ostentan sistemas IA más avanzados no se establece ninguna garantía.

La tecnología militar de última generación es una de las industrias en auge hoy en día, concretamente, la industria de sistemas autónomos inteligentes. A continuación, nombraré algunos de los sistemas que están comenzando a formar parte de las fuerzas armadas con mayor frecuencia.

El caso de la mercantil "*Anduril*", que ha comercializado un nuevo dron autónomo denominado "Ghost 4 SUAS", que además de un uso militar, permite su adquisición para particulares (https://www.anduril.com/ghost).

Vemos que los enjambres de drones inteligentes son una realidad en la actualidad, habiendo varios proyectos en curso, incluso en nuestro país. La empresa española y tecnologías de vanguardia Escribano *Mechanical & Engineering* S.L. ha desarrollado el proyecto LISS (*Long Range Intelligent Security System*). Este proyecto consiste en la creación de un sistema de enjambre de drones inteligente capaz de realizar misiones complejas en escenarios de gran dificultad. Esto supone en palabras de Jesús Martín Sánchez responsable del área de UAV (*Unmanned Aerial Vehicle*), "poner toda la inteligencia y la toma de decisiones a bordo de la nave, permitiendo al enjambre operar de forma totalmente autónoma sin que nadie 'al otro lado' de la línea de comunicación esté al mando." (González, 2021)

Entre los sistemas autónomos inteligentes, la industria de los vehículos aéreos no tripulados es una de las predominantes. No alcanza únicamente a drones, sino también a los cazas de combate. Observamos el proyecto estadounidense llamado *Skyborg* por el cual "la inteligencia artificial autónoma que será el cerebro de los aviones de combate norteamericanos en las próximas décadas. Este caza de combate totalmente autónomo que podrá volar de forma independiente y tomar decisiones propias para atacar enemigos y defender a su líder humano. El objetivo final del programa *Skyborg* es reemplazar a pilotos de combate humanos cuando esté completamente operativo." (Díaz, 2021). Según afirma la Fuerza Aérea de los Estados Unidos, en adelante USAF, esta IA puede aprender a volar cualquier aeronave, como quedó acreditado en las pruebas con el vehículo aéreo Kratos UTAP-22 "Mako", un dron táctico que no tiene nada que ver con los que Boeing, Kratos y General Atomics están desarrollando para Skyborg. Otra cualidad de estos sistemas autónomos es su capacidad de integrarse en una formación y colaborar entre distintos vehículos aéreos.

La USAF, junto a *Skyborg*, tiene otros proyectos en marcha como el de *Golden Horde* o *NTS-3.* Este último "es quizás el más pragmático y desarrolla nuevos receptores GPS que incorporen múltiples señales para las unidades militares. Parece un simple avance en una tecnología cotidiana, pero resulta trascendental en caso de conflicto y para operaciones militares. Baste pensar la cantidad de ingenios y armas que hoy en día basan su precisión (su eficacia) en establecer su situación geográfica, rutas y objetivos mediante posicionamiento satelital." (Fernández, 2021). Por su lado la *Golden Horde "*se busca un sistema de armas, en general bombas o misiles guiados, capaces de funcionar en modo colaborativo unas con otras. La idea es que un grupo o enjambre de armas compartan entre sí determinados datos, como podría ser la ubicación de sus objetivos y de sus defensas."

6. Cuestiones relativas al Derecho comparado sobre la imputación de la responsabilidad penal valiéndose de una Inteligencia Artificial.

Es conveniente analizar el marco normativo que han establecido otros países en el contexto de la Inteligencia Artificial, puesto que la regulación en esta materia todavía es escasa. El

desarrollo exponencial de los sistemas IA y su progresiva implementación en múltiples sectores de nuestra sociedad, conlleva una necesidad imperiosa de establecer un marco legal sólido y armonizado entorno a la Inteligencia Artificial. A continuación, expondré las iniciativas regulatorias de los países en los que la Inteligencia Artificial está alcanzando mayores niveles de desarrollo.

En primer lugar, analizaré el marco regulador actual de Estados Unidos entorno a los sistemas de Inteligencia Artificial. El 1 de enero de 2021 el congreso de los Estados Unidos aprobó la Ley de la Iniciativa Nacional de la Inteligencia Artificial (DIVISION E, SEC. 5001). Esta proporciona "un programa coordinado en todo el gobierno federal para acelerar la investigación y la aplicación de la inteligencia artificial para la prosperidad económica y la seguridad nacional de la nación." A su vez, han creado una Comisión de Seguridad Nacional sobre Inteligencia Artificial (NSCAI) cuyo objetivo consiste en "hacer recomendaciones al presidente y al Congreso para "avanzar en el desarrollo de inteligencia artificial, aprendizaje automático y tecnologías asociadas para abordar de manera integral las necesidades de defensa y seguridad nacional de los Estados Unidos". Este órgano realizó un informe final el 19 de marzo de 2021, en el cual señala en su primer capítulo de los 16 realizados, una serie de amenazas emergentes en la era de la IA (NSCAI, 2021, págs. 45-60). En este capítulo, categoriza los riesgos en cuatro apartados. El primero consiste en los riesgos actuales derivados de sistemas avanzados de IA, en los cuales encontramos el riesgo de que un programa maligno -*malware*- provoque que la IA haga una copia de sí misma, lo que la Comisión denomina "proceso de autorreplicación". Otro posible riesgo actual mencionado en el informe consiste en la posibilidad de que una IA diseñe y dirija patógenos. Finalmente, la posibilidad de mejorar las campañas de desinformación -*Fake news*-. Por otro lado, el segundo capítulo consiste en las "nuevas amenazas generadas desde los sistemas IA, en las que encontramos principalmente: las falsificaciones profundas a través de procedimientos computacionales, y enjambres de IA", de los cuales hablé en el epígrafe anterior. El tercer apartado trata el riesgo de que una IA se ataque así misma, a través de la manipulación por parte de un agente externo. Finalmente habla de los riesgos futuros a tener en cuenta posibles gracias a los sistemas IA, en los que encontramos principalmente la proliferación de armas autónomas letales a grupos terroristas. Estos hipotéticos ataques terroristas a los Estados Unidos "serían con precisión milimétrica, a gran escala, y a mayor velocidad", gracias a estos sistemas. Llega a comparar la era IA con la era de los misiles.

Por su parte, las iniciativas y propuestas regulatorias en Reino Unido se centran en establecer garantías para el despliegue de la IA en el sector público. Esto lo hace el Comité de Normas de la Vida Pública de Reino Unido, el cual presentó en febrero de 2020 un organismo de garantía regulatoria con la principal función de identificar situaciones de urgente necesidad de regulación llamado *Centre For Data Ethics and Innovation*, en adelante CDEI. (Gobierno de Reino Unido, s.f.). Este órgano elaboró en junio de 2020 un informe denominado "Barómetro IA", en el cual analiza los riesgos en determinados sectores de la sociedad, entre ellos la justicia penal en su tercer capítulo. (CDEI, 2020, págs. 24-46). Este capítulo se centra en el uso de la Inteligencia Artificial en la Administración penal, destacando la tecnología de reconocimiento facial, "herramientas algorítmicas de toma de decisiones o de apoyo a la herramientas de apoyo (ADMT)", útiles por ejemplo para predecir el riesgo de reincidencia; también encontramos "los análisis predictivos de la delincuencia informan de las decisiones de planificación proporcionando "mapas de calor" de la actividad delictiva que ayudan a las fuerzas a decidir su despliegue de recursos y sus respuestas". Finalmente encontramos la "policía científica digital, donde las herramientas de análisis de datos pueden mejorar la capacidad y la velocidad con la que los investigadores pueden buscar entre las pruebas digitales de dispositivos, correos electrónicos y cuentas de redes sociales para determinar la relevancia de un caso, o lo que puede ser necesario revelar en un proceso judicial."

Del mismo modo, la Cámara de los Lores elaboró un informe denominado "*AI in the UK: No Room for Complacency*" en su séptimo informe del período de sesiones 2019-2021 y publicado el 18 de diciembre de 2020. Este texto consta de tres capítulos y dos apéndices. El informe señala principalmente: la trascendencia de la convivencia con la IA en nuestra vida de ahora en adelante. Así como las deficiencia y escasa materia legislativa para la gobernanza del Reino Unido sobre la IA.

> En febrero 2021, el gobierno del Reino Unido en respuesta al anterior informe (*Government response to the House of Lords Select Committee on Artificial Intelligence*) reconoció que el enfoque del gobierno debe centrarse en establecer los arreglos necesarios y precisos entre las instituciones: entre el gobierno y el sector público, entre los reguladores, así como con la academia y la industria. Este enfoque garantizará que el impulso ganado en los últimos años no se pierda, permitirá utilizar el liderazgo de UK en IA para resolver desafíos globales y entiende que es crucial desarrollar la comprensión y la confianza del público en la IA. (The Technolawgist, 2021)

7. Implementación de un control humano significativo y técnico sobre el desarrollo y programación de sistemas IA para la prevención del daño.

El desarrollo y avance tecnológico y digital durante el Siglo XXI es exponencial, lo que supone que la implementación de sistemas inteligentes autónomos en nuestra sociedad sea prácticamente de límite desconocido. El desarrollo y la expansión de la inteligencia artificial y los robots ha llevado a su uso en muchos sectores de nuestra sociedad. Bien es cierto que, determinados sistemas como las Redes Neuronales Artificiales y "*deep learning*", se encuentran en una fase inicial y experimental, en cambio, es indudable que se implantarán en nuestra sociedad de manera progresiva a corto-medio plazo, lo que conocemos como la gran revolución de la IA. El desarrollo de este ámbito científico forma parte de la singularidad tecnológica, que mencionaba en la introducción. Esto incide en el curso de la evolución humana, ante este nuevo escenario de capacidades con límite desconocido.

Este evento, junto con los múltiples aspectos positivos que supone, conlleva también un correlativo riesgo y peligrosidad, ya que determinados sistemas de inteligencia artificial pueden ser medios susceptibles de lesionar un bien jurídico protegido por el Derecho Penal. El papel que ha de adoptar el Derecho ante esta tesitura consiste principalmente en otorgar los mecanismos necesarios que permitan al ser humano "mantener el dominio sobre las tecnologías autónomas emergentes como los robots y otros sistemas de IA." (Romeo, 2020, pág. 182). Para que el desarrollo de este control humano significativo, en adelante CHS, sobre los sistemas autónomos inteligentes será necesario establecer un ámbito regulativo que no dé lugar a lagunas ni contradicciones, y del que me dispongo a hablar posteriormente.

En primer lugar, hemos de concretar en qué consiste ese control humano significativo. Ha teorizado sobre este control la UNIDIR *(United Nations Institute for Disarmament Research)*. Este Instituto autónomo de las Naciones Unidas ha realizado diversos informes desarrollando la idea del CHS para los Sistemas de Armas Autónomas Letales *(LAWS)* (United Nations Institute for Disarmament Research , 2016). Del mismo modo que UNIDIR, también ha desarrollado esta idea "Artículo 36: *Killer Robots*", una organización no gubernamental

británica que tiene como objetivo el control político y legal para evitar los daños del uso de armas autónomas totalmente letales. Policy_Paper (article36.org). Por lo tanto, de acuerdo con estas definiciones, el CHS es todo proceso que impida el "libre albedrío" de sistemas de armas letales autónomas. En ese control y dominio humano, hacen hincapié explícitamente en la calidad del control ("significativo") y atribuye implícitamente la responsabilidad a los agentes humanos por las decisiones tomadas por los sistemas autónomos. Sin embargo, el concepto de CHS podría ir más allá de armas letales, pudiéndose proyectar sobre otras tecnologías cada vez más autónomas, particularmente aquellas dotadas de una Inteligencia Artificial "Fuerte". Sin embargo, también es cierto que en el marco del derecho penal sigue siendo una expresión algo imprecisa, en particular por el término "significativo que es, sin embargo, la palabra clave de este concepto. La seguridad jurídica, como derivación del principio de legalidad, que es una de las bases fundamentales del derecho penal contemporáneo, exige que la descripción de las categorías jurídico-penales sea suficientemente concreta. (Romeo, 2020, pág. 183)

Habiendo visto en qué consiste, ahora hemos de establecer el ámbito en el que se desplegará el control humano significativo -*Meaningful Human Control (MHC)*-. Comienzo por definir el ámbito objetivo, es decir, sobre qué máquinas ha de desplegarse el CHS. Determinados casos como los de uso instrumental de sistemas autónomos inteligentes como robots, o drones para la comisión de determinados delitos, el CHS no determina la responsabilidad a efectos penales. Puesto que para atribuiremos la responsabilidad a la persona que se sirvió del robot, es decir a su operador. Sin embargo, hemos de analizar el posible desarrollo de sistemas de inteligencia artificial más avanzados a corto-medio plazo, que ostenten una autonomía prácticamente total.

Anteriormente mencionaba los mecanismos de atribución de la responsabilidad penal por delitos perpetrados por sistemas de inteligencia artificial "fuerte". Mencionaba un supuesto en el que un robot dotado de un sistema de "*deep learning*", es programado para que no cometa ningún delito en ninguna circunstancia. Sin embargo, la capacidad de aprendizaje del sistema supera las expectativas del programador. Por consiguiente, estas nuevas "capas de aprendizaje" que forman parte de su sistema RNA, carecen de directrices de programación, por lo que el sistema evade el control del humano y pueda llevar a cabo uno o varios ilícitos penales.

> Se está extendiendo la postura, que parte de los debates en torno a los Sistemas de Armas Autónomas Letales (LAWS) (United Nations Institute for Disarmament Research , 2016) y a los sistemas de vehículos autónomos, de que el Control Humano Relevante (CHS) es esencial para la responsabilidad en el ámbito internacional. Esto significa que los seres humanos -y no los ordenadores y sus algoritmos- deben mantener el control final, pudiendo así ser moral y quizás legalmente responsables por ello. El CHS debería convertirse en una norma de lo que debe ser la actitud del ser humano hacia cualquier tecnología que sea capaz de aumentar su autonomía respecto a los humanos que la crearon o utilizaron. (Romeo, 2020, pág. 183)

En todos estos supuestos que mencionaba, el control humano significativo (CHS) es el factor esencial para la atribución de la responsabilidad penal. Por lo tanto, el ámbito objetivo del CHS que se desplegará sobre todos aquellos sistemas con capacidades de autoaprendizaje, evaluación de su entorno y dotados de una autonomía suficiente para la toma de decisiones automatizadas. Una posibilidad de definir el ámbito objetivo del CHS es la de la propuesta de la Comisión Europea, puesto que categoriza los sistemas en función de riesgo; podríamos decir que el CHS ha de desplegarse sobre todos los sistemas de alto riesgo, e incluso de los de riesgo limitado.

A continuación, hemos de concretar el ámbito subjetivo del Control Humano Significativo, es decir, sobre qué personas se proyectará. Serán los programadores que han intervenido en el diseño y fabricación, las personas jurídicas que han distribuido el sistema IA, así como sus usuarios finales. Finalmente hay que recalcar que el contexto en el que se desenvolverá el CHS será el modelo de "*Compliance*" que sustentan la responsabilidad de las personas jurídicas como desarrollaré en los epígrafes posteriores.

7.1 Iniciativas de las instituciones de la Unión Europea y categorización del riesgo de los sistemas de Inteligencia Artificial.

La cooperación internacional es el eje por el que ha de pivotar la regulación de los sistemas IA, poniendo sus beneficios al servicio de las personas, y gestionando los riesgos de la forma más eficaz posible.

Las instituciones de la UE se han pronunciado en diversas ocasiones sobre esta cuestión. En 2017, el Consejo Europeo pidió con "sentido de urgencia para abordar las tendencias emergentes" incluyendo "cuestiones como la inteligencia artificial ..., garantizando al mismo tiempo una alta nivel de protección de datos, derechos digitales y estándares éticos (El Consejo Europeo, 2017, págs. 5-8). El Consejo Europeo también ha pedido que se determinen claramente los sistemas IA que deben considerarse de alto riesgo. (Consejo Europeo, 2020, pág. 6). Vemos que en este pronunciamiento el Consejo Europeo insta a los estados miembros a abordar cambios significativos en la legislación para adaptar un campo normativo a los desafíos que suponen las nuevas tecnologías, e incluso especifica que hay determinados sistemas IA que deben considerarse de alto riesgo. Esto implica definir de forma concreta cuál es el riesgo permitido en los delitos perpetrados a través de sistemas IA, estadio en el que el Derecho Penal ha de intervenir.

Del mismo modo, el Consejo de la Unión Europea, en sus Conclusiones de 2019 sobre el Plan coordinado de desarrollo y uso de inteligencia artificial, destacó la “importancia de garantizar el pleno respeto de los derechos de los ciudadanos europeos y pidió una revisión de la legislación pertinente existente para que se adapte a las nuevas oportunidades y desafíos planteados por la IA.” (Consejo de la Unión Europea, 2019). Unas conclusiones más recientes del 21 de octubre de 2021 del mismo órgano pedían además que “se abordara la opacidad, la complejidad, la parcialidad, un cierto grado de imprevisibilidad y el comportamiento parcialmente autónomo de ciertos sistemas de IA, para garantizar su compatibilidad con los derechos fundamentales y facilitar la aplicación de las normas jurídicas.” (Consejo de la Unión Europea, 2020, pág. 5). En el mismo texto nos indica que a su vez es necesario realizar esfuerzos para que el diseño, desarrollo, despliegue y uso indebido de los sistemas IA puedan entrañar riesgos para derechos fundamentales, la democracia y el Estado de Derecho. Nos dice que “para hacer frente a los posibles riesgos de forma eficaz, deben cumplirse requisitos específicos para el diseño, desarrollo, despliegue y uso de los sistemas de IA.” (Consejo de la Unión Europea, 2020, pág. 6). Es aquí donde reside la clave de la cooperación internacional para la elaboración de una IA fiable y segura. Han de elaborarse y desarrollarse normas técnicas por parte de los Estados miembros de la

UE, de forma coordinada y completa. Además de los estados miembros, la normativa comunitaria también ha de desarrollarse, adoptando el enfoque de la Comisión Europea en su Libro Blanco "Sobre la Inteligencia Artificial - Un enfoque europeo para excelencia y confianza". En este texto la Comisión insta a revisar la normativa adecuadamente, tanto los riesgos y las oportunidades, así como los requisitos de las aplicaciones de la IA, si puede aplicarse eficazmente y si son necesarios ajustes o nueva legislación, también en lo que respecta a la protección de nuestros principios y valores comunes.

El Parlamento Europeo ha realizado un trabajo considerable a lo largo de los últimos años en el campo de las nuevas tecnologías, especialmente de la Inteligencia Artificial. Son de especial atención las resoluciones emitidas este año sobre Inteligencia Artificial en materia penal, en la que encontramos la Resolución 2020/2016(INI) *"La inteligencia artificial en el derecho penal y su uso por las autoridades policiales y judiciales en materia penal"*

La Comisión Europea elaboró el 21 de abril de 2021 una propuesta de Reglamento del Parlamento europeo y del Consejo, por el cual se plantean nuevas normas y acciones para la excelencia y la confianza en la Inteligencia Artificial en el territorio europeo.

En su exposición de motivos recalca que "la IA es una familia de tecnologías en rápida evolución que puede aportar una amplia gama de beneficios económicos y sociales, al mejorar la predicción, la optimización de las operaciones y la asignación de recursos, y la personalización de la prestación de servicios. Sin embargo, los mismos elementos y técnicas que impulsan los beneficios socioeconómicos de la IA también pueden provocar nuevos riesgos o consecuencias negativas para los individuos o la sociedad." (European Comission, 2021). A lo largo de la propuesta queda reflejado el interés de la Unión Europea en preservar el "liderazgo tecnológico", lo que implica lograr una soberanía digital para poder actuar con autodeterminación en la esfera digital y fomentar la resiliencia de la Unión Europea.. Esto es inconcebible si el desarrollo de los sistemas de inteligencia artificial no se basa también en los valores, derechos fundamentales y principios de la Unión Europea. Por su lado, el Parlamento europeo ya ha expresado su deseo de que se adopten medidas legislativas por parte de los estados miembros que establezcan garantías para el correcto funcionamiento de sistemas IA. (Parlamento Europeo, 2020).

La propuesta de la Comisión Europea se elabora con unos objetivos específicos, de los cuales nos interesan especialmente el de "garantizar que los sistemas de IA comercializados y utilizados en la Unión sean seguros y respeten la legislación vigente en materia de derechos fundamentales y los valores de la Unión; así como mejorar la gobernanza y la aplicación efectiva de la legislación vigente en materia de derechos fundamentales y los requisitos de seguridad aplicables a los sistemas de IA" (European Comission, 2021, pág. 3). Esto es clave, puesto que partimos de la base de que todo proyecto de IA que sea elaborado deberá respetar la legislación vigente, con mayor motivo, la legislación penal, por lo que, si un sistema IA vulnerara un bien jurídico protegido por el Ordenamiento jurídico penal, tendrá relevancia para la individualización de la responsabilidad penal. En función del cumplimiento de las garantías necesarias para la prevención del daño por parte de los programadores y diseñadores.

La cooperación internacional para el desarrollo de una IA fiable, lícita y sólida se ve reflejada en la propuesta de la Comisión Europea, insta a los Estados miembros a establecer un sistema de gobernanza sobre los sistemas IA a través de sus mecanismos jurídicos. Del mismo modo, propone "crear un Consejo Europeo de Inteligencia Artificial, junto con otras medidas adicionales (…)" (European Comission, 2021, pág. 3).

La propuesta de la Comisión Europea exige "una total coherencia con la legislación de la Unión vigente aplicable a los sectores en los que ya se utilizan sistemas de IA de alto riesgo o en los que es probable que se utilicen en un futuro próximo." (European Comission, 2021, pág. 4). Esto motiva la propuesta de establecer un marco normativo para la regulación del uso de la inteligencia artificial, estableciendo distintos niveles de riesgo realizando la siguiente clasificación de sistemas IA.

Encontramos en primer lugar la mayor categoría de riesgo a la que la Comisión Europea considera como prácticas prohibidas en el título II de la propuesta. "La IA que contradice los Derechos Fundamentales de la Carta de la UE será prohibida, lo que incluye los sistemas de IA que se consideren una clara amenaza para la seguridad, los medios de subsistencia y los derechos de las personas." Esto abarca los sistemas o las aplicaciones de IA que manipulan el comportamiento humano para eludir la voluntad de los usuarios (por ejemplo, manipulación subliminal), "la explotación de niños o personas con discapacidad mental, resultando en daños físicos/psicológicos", "el reconocimiento facial o biométrico en espacios públicos", lo

que denomina vigilancia indiscriminada. En este punto nos planteamos la permisión de cámaras dotadas de softwares con sistemas IA, con capacidad de reconocimiento facial y otros rasgos físicos. Se necesitará autorización especial para el uso de "sistemas de identificación biométrica remota", como el reconocimiento facial, en espacios públicos para las excepciones de medidas antiterroristas, o la policía predictiva. Del mismo modo prohíbe los sistemas que crean puntuaciones de crédito social por parte de las administraciones públicas (aunque no de la empresa privada). Es decir, estará prohibido juzgar la fiabilidad de una persona en función de su comportamiento social o de los rasgos de personalidad previstos a través de sistemas de Inteligencia Artificial. (European Comission, 2021, págs. 43-45).

En virtud de esta categorización del riesgo, entiendo que la legislación penal debe ir encaminada a imputar la responsabilidad penal a todos los sujetos intervinientes en el proceso de diseño, industrialización, comercialización y uso de estos sistemas. Es interesante, como señalaba en el epígrafe 5.c, la posibilidad de establecer un tipo penal en el que el Estado reaccione imponiendo el castigo penal, no ante la causación de un resultado material de daño o lesión, sino ante el peligro que supone la mera creación de un sistema de estas características. Lo que conocemos como delitos de riesgo establecidos en el Código Penal en sus artículos 348 a 350.

Por otro lado, encontramos en el Título III de la propuesta de la Comisión, los sistemas IA de alto riesgo (*high-risk AI systems*). Los cuales la Comisión Europea propone integrarlos “en la legislación de seguridad sectorial existente para garantizar la coherencia, evitar duplicaciones y minimizar las cargas adicionales.” (European Comission, 2021, pág. 4). Estos sistemas, estando permitidos, se encuentran sujetos al cumplimiento de los requisitos para una IA fiable, ética y sólida, así como a la evaluación de la conformidad ex ante, establecidos en los artículos 8 a 15. En esta cuestión el modelo “*Compliance*”, del que hablaré en el epígrafe posterior, juega un papel fundamental.

Conforme a las reglas de clasificación para sistemas IA de alto riesgo del artículo 6 y 7 de la propuesta de la Comisión Europea, abarcan las tecnologías empleadas en: las infraestructuras críticas (que pueden poner en peligro la vida y la salud de los ciudadanos como, por ejemplo, los coches autónomos); la formación educativa o profesional, que puede determinar el acceso a la educación y la carrera profesional de una persona (por ejemplo, la puntuación en exámenes); los componentes de seguridad de los productos (por ejemplo, la aplicación de la

IA en cirugía asistida por robots); el reclutamiento de empleados (por ejemplo, los programas informáticos de clasificación de CV para procedimientos de contratación de recursos humanos); los servicios públicos y privados esenciales (por ejemplo, los sistemas de calificación crediticia que priven a los ciudadanos de la oportunidad de obtener un préstamo); la aplicación de las leyes, que pueden interferir con los derechos fundamentales de las personas (por ejemplo, la evaluación de la fiabilidad de las pruebas); la gestión de la migración, asilo y control de las fronteras (por ejemplo, la comprobación de la autenticidad de los documentos de viaje); y la administración de justicia y procesos democráticos (por ejemplo, la aplicación de la ley a un conjunto concreto de hechos). (European Comission, 2021, págs. 45-46)

En el título IV de la propuesta menciona los sistemas de IA de riesgo limitado, estableciendo unas obligaciones específicas de información y transparencia en su artículo 52. (European Comission, 2021, pág. 69) Será necesario notificar a las personas cuando interactúan con un sistema de IA, a menos que sea "obvio por las circunstancias y el contexto de uso". El caso de los robots conversacionales que interactúen con los usuarios, estos últimos deberán ser conscientes de que se encuentran con una máquina, para poder tomar una decisión informada de continuar o no. Finalmente, los de riesgo mínimo o nulo, en los que encontramos la inmensa mayoría de los sistemas IA hasta el momento. Claramente estos dos últimos tienen escasa relevancia para el Derecho Penal y debería obviarlos en virtud del principio de proporcionalidad y necesidad.

Finalmente propone la creación de un "Consejo Europeo de Inteligencia Artificial", formado por representantes de todos los países, para ayudar a la Comisión a decidir qué sistemas de IA se consideran de "alto riesgo" y recomendar cambios en las prohibiciones. Esto es fundamental puesto que el progreso y avance tecnológico puede dar lugar a algunas modificaciones en la metodología de clasificación de sistemas IA. El Consejo está trabajando en un convenio que trate la responsabilidad en la IA.

7.2 Modelo de "*Compliance*" penal eficaz que integre medidas para prevenir riesgos penales derivados de delitos cometidos por IA. Principios para el desarrollo de una Inteligencia Artificial lícita, ética y fiable.

El modelo de "*Compliance*", incorporado inicialmente al Código Penal en el año 2010 obligaba a las personas jurídicas a preparar un modelo de prevención de riesgos penales. Más tarde en la reforma del Código Penal del año 2015 por la que se amplía el artículo 31.bis, establece un contenido completo con las obligaciones específicas a cumplir.

Este sistema basado en el de las personas jurídicas, puede ser trasladable para la atribución de la responsabilidad penal por delitos perpetrados por sistemas autónomos inteligentes, puesto que permite definir el ámbito en el que enmarcamos el riesgo permitido.

Antes mencionaba la gran revolución de la inteligencia artificial y lo que ello implica. Uno de los elementos de este fenómeno consiste en la implementación de los sistemas de inteligencia artificial en múltiples sectores de la sociedad. Ergo, la industrialización y respectiva comercialización de estos sistemas, conlleva que estos robots puedan adquirirse a título particular. Ante este fenómeno, mi propuesta es la configuración de un modelo de "*Compliance"* penal para un Control Humano Significativo eficaz sobre los sistemas autónomos inteligentes.

En la utilización de instrumentos sofisticados como son las IA, no se va a confiar toda la responsabilidad al productor, sino que intervendrán más agentes. Ha de haber agencias independientes encargadas de validar su funcionamiento, por ejemplo. De acuerdo con el enfoque del modelo de "*compliance*", debe ir precedido de un examen general y detallado del sistema por parte de una agencia externa acreditada, formada por expertos técnicos autorizados e independientes. Estos evaluarán la idoneidad técnica del robot o del sistema de IA para el servicio o las actividades que iban a prestar, el grado de error previsible, su inocuidad básica para bienes jurídicos, etc. La evaluación positiva y la acreditación oficial darían lugar a la homologación y autorización del uso público o privado del sistema de IA. Esta acreditación debería indicar las revisiones y validaciones que deberían realizar los usuarios, o que deberían establecer sus empresas u organizaciones, sobre las conclusiones y propuestas del sistema; e indicando las funciones específicas para las que ha sido acreditado.

El modelo "*Compliance*" se proyectará sobre todo el proceso de "vida" útil del sistema IA. Esto comprende a todos los sujetos enumerados en la propuesta de la Comisión Europea en su artículo 3 consistente en la definición. (European Comission, 2021, págs. 39-43).

> Distinguimos a los denominados “operadores” que serán el proveedor, el usuario, el representante autorizado, el importador y el distribuidor. Por un lado, el proveedor, a pequeña o gran escala, consiste en “una persona física o jurídica, administración pública, agencia u otro organismo que desarrolla un sistema de IA o que hace desarrollar un sistema de IA con vistas a comercializarlo a través de una marca, o ponerlo en servicio propio, a título oneroso o gratuito.” Por otro lado, encontramos al usuario, que abarca “cualquier persona física o jurídica, administración pública, agencia u otro organismo que utilice un sistema de IA bajo su autoridad.” El “representante autorizado" es toda persona física o jurídica establecida en la Unión que haya recibido un mandato escrito de un proveedor de un sistema de IA para respectivamente, ejecutar y llevar a cabo en su nombre las obligaciones y procedimientos establecidos en la presente propuesta. El “importador” es todo aquel que comercialice o ponga en servicio un sistema de IA que lleve el nombre o la marca de una persona física o jurídica establecida fuera de la Unión.

En caso de comisión de delitos por, o a través de sistemas IA, hemos de tener en cuenta a estos sujetos a la hora de individualizar la responsabilidad penal, en caso de que alguno de ellos obrara con dolo o incumpliendo algún deber de cuidado.

La evaluación positiva del sistema y su acreditación oficial dependerán del cumplimiento de una serie de requisitos que expondré a continuación.

> Observamos un primer requisito de acción y supervisión humana. En lo que a la acción humana respecta, los usuarios deberían ser capaces de tomar decisiones autónomas con conocimiento de causa en relación con los sistemas de IA. Se les deberían proporcionar los conocimientos y herramientas necesarios para comprender los sistemas de IA e interactuar con ellos de manera satisfactoria y, siempre que resulte posible, permitírseles evaluar por sí mismos o cuestionar el sistema. Del mismo modo, fortalecer el derecho a no ser sometidos a ninguna decisión basada exclusivamente en procesos automatizados cuando tal decisión produzca efectos adversos sobre los usuarios o terceros. (Grupo de expertos de alto nivel en inteligencia artificial (AI HLEG), 2019, pág. 16)
> En cuanto a la supervisión humana implica todo procedimiento que ayude a garantizar que un sistema de IA no socave la autonomía humana o provoque otros efectos adversos. Encontramos tres tipos de supervisión. En primer lugar, la participación humana,

consistente en la capacidad de que intervengan seres humanos en todos los ciclos de decisión. También implica el control humano sobre el sistema, la capacidad de que intervengan seres humanos durante el ciclo de diseño del sistema y en el seguimiento de su funcionamiento. Finalmente, el mando humano es la capacidad de supervisar la actividad global del sistema (incluidos sus efectos económicos, sociales, jurídicos y éticos), así como la capacidad de decidir cómo y cuándo utilizar el sistema en una situación determinada. (Grupo de expertos de alto nivel en inteligencia artificial (AI HLEG), 2019, pág. 16)

El segundo requisito implica que el sistema sea sólido técnicamente y seguro. Para tal evaluación será necesario la participación de expertos técnicos que pongan a prueba la resistencia a los ataques y su respectiva seguridad. Deberá protegerse frente a las vulnerabilidades, un ataque a una IA puede tener efectos mucho más perniciosos. Es preciso tener en cuenta las posibles aplicaciones imprevistas de la IA, así como el abuso potencial de un sistema IA por parte de agentes malintencionados. Para conseguir tales fines será necesario la elaboración de un plan de repliegue y seguridad general. Será necesario para la elaboración de una IA fiable, sólida y segura, que su proceso de desarrollo y evaluación se elabore de la forma más precisa posible. A mayor precisión en la evaluación de escenarios de la conducta de la IA, mayor previsión de fallos del sistema, pudiendo mitigar, respaldar y corregir riesgos imprevistos asociados a predicciones incorrectas del sistema. Cuando no sea posible evitar este tipo de predicciones, es importante que el sistema pueda indicar la probabilidad de que se produzcan esos errores. (Grupo de expertos de alto nivel en inteligencia artificial (AI HLEG), 2019, págs. 16-17). Es interesante la posible elaboración de una “memoria” de la conducta de la IA, que podría facilitar el proceso de ensayo y reproducción de comportamientos.

El tercer requisito consiste en la gestión de la privacidad y de los datos. Los sistemas de IA deben garantizar la protección de la intimidad y de los datos a lo largo de todo el ciclo de vida de un sistema (tanto la información inicialmente facilitada como la información generada). Del mismo modo se debe salvaguardar la calidad e integridad de estos datos. Cuando se recopilan datos, estos pueden contener sesgos sociales, imprecisiones y errores. La introducción de datos malintencionados en un sistema de IA puede alterar su comportamiento. Deberían establecerse protocolos que rijan el acceso a los datos. Quién puede acceder a los datos y en qué circunstancias. Solamente debería permitirse acceder a

> los datos personales a personal debidamente cualificado, poseedor de las competencias adecuadas y que necesite acceder a la información pertinente. (Grupo de expertos de alto nivel en inteligencia artificial (AI HLEG), 2019, pág. 17)

> Finalmente, un último requisito consistente en la transparencia. El conjunto de datos y los procesos que dan lugar a la decisión del sistema de IA, incluidos los relativos a la recopilación y etiquetado de los datos, así como a los algoritmos utilizados, deberían documentarse. Esto también es aplicable a las decisiones que adopte la IA. Esto permitirá identificar los motivos de una decisión errónea por parte del sistema, lo que a su vez podría ayudar a prevenir futuros errores. Este requisito también implica la capacidad de explicar tanto los procesos técnicos de un sistema de IA como las decisiones humanas asociadas. Requiere que las decisiones que adopte un sistema de IA sean comprensibles para los seres humanos y estos tengan la posibilidad de rastrearlas. Cuando un sistema de IA tenga un impacto significativo en la vida de las personas, debería ser posible reclamar una explicación adecuada y adaptada del proceso de toma de decisiones del sistema de IA. (Grupo de expertos de alto nivel en inteligencia artificial (AI HLEG), 2019, pág. 18)

Este modelo de *"Compliance"* para las personas implicadas en el desarrollo y fabricación del sistema, ha de basarse en las siguientes directrices o criterios para la fiabilidad de la IA, los cuales seguirá la agencia externa encargada de acreditar el sistema. A continuación, me dispongo a enumerar los criterios a seguir. Un primer grupo de directrices serán las que llamaremos sustanciales, el segundo grupo lo denominamos directrices técnicas.

> En lo que al primer grupo respecta, encontramos en primer lugar, la fabricación de una IA lícita, lo que implica que ha de cumplir todas las leyes y reglamentos aplicables, así como la carta de Derechos Humanos. Ha de ser ética, de modo que se garantice el respeto de los principios y valores éticos de nuestra sociedad. Debe ser "robusta", desde el punto de vista técnico y también social, lo que implica que el sistema no tenga fallos internos que den lugar a daños accidentales. (Grupo de expertos de alto nivel en inteligencia artificial (AI HLEG), 2019, págs. 9-10)

> En cuanto al segundo grupo de directrices, uno de los criterios esenciales consistirá en el de prevención del daño, que implica que los sistemas de IA no deberían provocar daños (o agravar los existentes) ni perjudicar de cualquier otro modo a los seres humanos. Todos

> los sistemas IA deben ser seguros y robustos desde el punto de vista técnico, debería garantizarse que no puedan destinarse a usos malintencionados. La prevención del daño implica asimismo tener en cuenta el entorno natural y a todos los seres vivos. Encontramos también el criterio de equidad, que consistirá en que los profesionales de la IA deberían estudiar cuidadosamente cómo alcanzar un equilibrio entre los diferentes intereses y objetivos que puedan confrontarse. Del mismo modo, la equidad conlleva la capacidad de oponerse a las decisiones adoptadas por los sistemas de IA y por las personas que los manejan, así como de tratar de obtener compensaciones adecuadas frente a ellas, se debe poder identificar a la entidad responsable de la decisión y explicar los procesos de adopción de decisiones. Finalmente, un criterio descriptivo que será crucial para conseguir que los usuarios confíen en los sistemas de IA y para mantener dicha confianza. Esto implica también que los procesos han de ser transparentes. La empresa fabricante en cuestión deberá también comunicar abiertamente las capacidades y la finalidad de los sistemas de IA. Las decisiones deben poder explicarse a las partes que se vean afectadas por ellas de manera directa. Los algoritmos de "caja negra", requieren especial atención, pude ser necesario adoptar otras medidas relacionadas con este criterio (por ejemplo, la trazabilidad, la auditabilidad y la comunicación transparente sobre las prestaciones del sistema). (Martín, 2019) (Grupo de expertos de alto nivel en inteligencia artificial (AI HLEG), 2019, págs. 10-12)

Tal y como he expuesto anteriormente, otra de las cuestiones fundamentales a determinar para la implementación de un modelo de "Compliance" para los sistemas IA, que permita un control humano significativo eficiente es la calificación de niveles de riesgo. Me remito a la categorización del riesgo de los sistemas IA establecida por la Comisión Europea, y expuesta anteriormente. Los criterios que han de seguirse son los niveles de riesgo potenciales, y la complejidad de la variabilidad de la conducta del sistema. Ingenieros de Software cualificados tendrán que analizar el diseño para ver si hay herramientas del sistema que permitan su reprogramación y aprendizaje autónomo que conlleve un resultado lesivo. Para ello la agencia externa acreditada emitirá un informe motivado y revalidado posteriormente por las autoridades correspondientes, el cual será preceptivo a efectos del desarrollo del sistema.

La razón de ser de este modelo de *"compliance"* penal para los sistemas IA, radica en la imputación de la responsabilidad penal a aquellas personas, físicas o jurídicas, a las cuales se

les atribuye la conducta delictiva del sistema IA como suya, siempre que no hubieran seguido este modelo. Por lo contrario, de haber seguido el modelo, superando todas las fases de control y supervisión, con posterior acreditación oficial para el uso de la agencia externa, y si aun así el sistema produce un resultado lesivo, la responsabilidad recaería sobre la agencia externa que otorga la acreditación. Esto se fundamenta en el principio de confianza. Este principio consiste en que una persona participa en una determinada actividad durante sus relaciones sociales y el transcurso de su vida, con la confianza de que los otros intervinientes en una relación determinada actuarán cumpliendo sus respectivos deberes de cuidado (por ejemplo, cuando se conduce un vehículo motor). Es decir, no hemos de comprobar en cada acción que realicemos, si los demás intervinientes en dicha acción están actuando con la diligencia debida y respetando los deberes de cuidado, a menos de que tengamos indicios de lo contrario. Por lo tanto, observamos que este principio es de vital importancia para la individualización de la responsabilidad penal, puesto que nos servirá para identificar quién ha incumplido sus respectivos deberes de cuidado, en función de si el sistema inteligente autónomo se encuentra homologado o no.

Conclusiones.

El avance vertiginoso del desarrollo de sistemas IA, cada vez con mayor autonomía, junto con los enormes beneficios que suponen, conlleva un correlativo riesgo implícito, la comisión de delitos por o a través de estos sistemas. La gran mayoría de organismos internacionales coinciden en la necesidad de abordar esta problemática a través de un marco legislativo sólido y eficaz.

Este marco legislativo debe establecer garantías sólidas para evitar cualquier daño o lesión, así como la adopción de medidas de carácter preventivo. Nuestro Ordenamiento Jurídico ostenta mecanismos suficientes para atribuir a una persona física o jurídica aquellos delitos perpetrados por robots y sistemas IA, siempre que hayan sido cometidos a través de un uso instrumentalizado de estos sistemas.

Supone una mayor complejidad para aquellos sistemas autónomos inteligentes dotados de una IA “fuerte” cuya conducta, completamente autónoma, deriva en la comisión de un delito.

En este supuesto hemos de abordar una política legislativa eficaz que integre un modelo "*Compliance*" en la elaboración estos sistemas. Esto tiene una doble virtualidad, en primer lugar, tiene un carácter preventivo y serán eliminados riesgos inherentes a estos sistemas. Por otra parte, nos dará luz para la individualización de la responsabilidad entre los agentes intervinientes en la creación de la máquina que comete el delito en cuestión. La legislación penal debe ir encaminada a materializar este modelo basado en el de las personas jurídicas. Del mismo modo, será necesario establecer un etiquetado para categorizar estos sistemas en función de su riesgo.

Por último, la implementación de un nuevo tipo penal, que consistirá en un delito de riesgo en el que se castigue la mera elaboración de aquellos sistemas que la Comisión europea considera como inadmisibles o prohibidos.

Con todo esto se pretende que el Derecho Penal cumpla la función de asegurar la vida social y pacífica de las personas ante la necesidad imperiosa de establecer un marco regulativo amplio entorno a los sistemas de Inteligencia Artificial.

REFERENCIAS

Artículo de revista electrónica (sin DOI)

- Santos González, M.J. (2017). Regulación legal de la robótica y la inteligencia artificial. *Revista jurídica de la Universidad de León.* http://revpubli.unileon.es/index.php/juridica/article/view/5285
- Ramón Fernández, F. (2019). Robótica, inteligencia artificial y seguridad: ¿Cómo encajar la responsabilidad civil? Universitat Politècnica de Valencia. https://riunet.upv.es/bitstream/handle/10251/117875/Rob%C3%B3tica.pdf?sequence=1&isAllowed=y
- Sillick, T. J. y Schutte, N. S. (2006). Emotional intelligence and self-esteem mediate between perceived early parental love and adult happiness. *E-Journal of Applied Psychology*, *2*(2), 28-48. Recuperado de http://ojs.lib.swin.edu.au/index.php/ejap
- Domínguez, Martha C., & García-Vallejo, Felipe (2009). La sexta revolución tecnológica: El camino hacia la singularidad en el siglo XXI. El Hombre y la Máquina, (33),8-21. [fecha de Consulta 9 de Abril de 2021]. ISSN: 0121-0777. Disponible en: https://www.redalyc.org/articulo.oa?id=47812225002
- Romeo, Carlos M. (2020). Criminal Responsabiliity of Robots and Autonomous Artificial Intelligent Systems
- Almansa, E. (2020). ¿Para qué se utilizan los drones con inteligencia artificial? https://revistadigital.inesem.es/informatica-y-tics/drones-con-inteligencia-artificial/
- Sarabia, D. (2017) Sky Net: el Gran Hermano chino que vigila con 20 millones de cámaras inteligentes https://www.eldiario.es/tecnologia/sky-net-gran-hermano-china_1_3173654.html
- https://www.wsj.com/articles/facebook-really-is-spying-on-you-just-not-through-your-phones-mic-1520448644

Artículo de periódico

- Fernández, J. (29 de enero de 2021). Drones y bombas que “hablan”: La IA es la gran revolución militar, y nadie está al mando. El Confidencial.

- Malvar, A. (12 de agosto de 2017). ¿Qué fue de Tay, la robot de Microsofto que se volvió nazi y machista? Público.
- Embury-Dennis T. (7 de diciembre de 2020). Científico nuclear iraní fue asesinado con una "metralleta controlada por satélite," según Irán. Independent en Español.
- Zamarreñio, A. (29 de enero de 2021). La voz de la ciudadanía, la ausente en el debate de la Inteligencia Artificial. Cadena Ser.
- Carabantes López, M. (2014). Inteligencia Artificial: Condiciones de posibilidad técnicas y sociales para la creación de máquinas pensantes. *Universidad Complutense de Madrid.* https://eprints.ucm.es/id/eprint/24630/1/T35134.pdf

Página web

- *La década de los 90, Eduardo Maura y las mujeres y la ciencia*. (9 de febrero de 2021). Hoy por hoy. Recuperado de https://www.youtube.com/watch?v=1RN6WQHKbTc&t=1946s

Datos de investigación

- Sánchez-Élez Martín, M. (2019). Ética, Legislación y Profesión. Directrices éticas para una IA fiable.

BIBLIOGRAFÍA

Libros

- Romeo Casabona, C.M., Guanarteme Sánchez Lazaro, F. y Armaza Armaza E.J. 2010. Adaptación del Derecho Penal al desarrollo social y tecnológico.
- Hallevy, G. 2015. Liability for crimes involving Artificial Intelligence Systems

Capítulo de libro

- Eduardo Aboso, G. 2017. Derecho Penal Cibernético. La cibercriminalidad y el Derecho penal en la moderna sociedad de la información y la tecnología de la comunicación.

- Núñez Zorrila, M.C. 2019. Inteligencia artificial y responsabilidad civil. Régimen jurídico de los daños causados por robots autónomos con inteligencia artificial.
- Mir Puig, S. y Luzón Peña, D.M. 1996. Responsabilidad Penal de las empresas y sus órganos y responsabilidad por el producto.
- Kiefer, M. 2016. Cibercrimen. Aspectos de Derecho Penal y procesal penal.
- Provolo, D; Riondato, S; y Yenisey, F. 2014. Genetics, Robotics, Law, Punishment.
- Hallevy, G. 2015. Liability for crimes involving Artificial Intelligence Systems.
- Amador Hidalgo, L. 1996. Inteligencia Artificial y sistemas expertos.

NORMATIVA

- Ley de la Iniciativa Nacional de la Inteligencia Artificial del Congreso de los Estados Unidos (DIVISION E, SEC. 5001)

INSTITUCIONES Y ÓRGANOS

- Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, Al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Inteligencia Artificial para Europa. COM/2018/237 final
- Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de Inteligencia Artificial y se modifican determinados actos legislativos de la Unión. Comisión Europea. Bruselas, 21.4.2021 COM (2021) 206 final 2021/0106 (COD)
- Conclusiones de la Reunión del Consejo Europeo (19 October 2017). Brussels, EUCO 14/17.
- Conclusiones de la Reunión Extraordinaria del Consejo Europeo (2 de octubre de 2020) Brussels, EUCO 13/20.
- Conclusiones del Consejo de la Unión Europea sobre el plan coordinado de desarrollo y uso de la Inteligencia Artificial. (11 de febrero de 2019)
- Conclusiones de la Presidencia del Consejo de la Unión Europea sobre la Carta de Derechos Fundamentales en el contexto de la Inteligencia Artificial y Cambio Digital. (21 de octubre de 2020). 11481/20