VOLUMEN Nº4 JULIO 2024

Revista de Responsabilidad Penal de Personas Jurídicas y Compliance

RESPONSABILIDAD PENAL Y EXTRAPENAL DE LAS PERSONAS JURÍDICAS POR DELITOS RELACIONADOS CON LA INTELIGENCIA ARTIFICIAL: VÍAS DE INTERVENCIÓN LEGAL Y PRINCIPALES OBSTÁCULOS

Prof. Dr. Vincenzo Mongillo

ÍNDICE

RESPONSABILIDAD PENAL Y EXTRAPENAL DE LAS PERSONAS JURÍDICAS POR DELITOS RELACIONADOS CON LA INTELIGENCIA ARTIFICIAL: VÍAS DE INTERVENCIÓN LEGAL Y PRINCIPALES OBSTÁCULOS

Prof. Dr. Vincenzo Mongillo *

I. Introducción

Algoritmos de alta complejidad, vehículos sin conductor, robots "humanoides", sistemas de armas autónomos y letales, y mucho más, ya no son sólo material para guionistas y autores de libros de ciencia ficción. Sin duda alguna, hemos entrado en la era de la inteligencia artificial (en adelante también denominada "IA"), que probablemente se convertirá en la tecnología más importante del siglo XXI, gracias a los extraordinarios avances impulsados por el aumento exponencial de los datos digitales y las capacidades computacionales[1]. El carácter histórico de tales logros técnicos-científicos explica que observadores atentos vislumbren una cuarta revolución industrial[2] y, en perspectiva, "el mayor acontecimiento de la historia de nuestra civilización"[3]. Sus posibles repercusiones afectarán – y, en cierta medida, ya han afectado – a todos los ámbitos de la vida social, tanto en contextos de paz como de guerra, de trabajo o de ocio: medicina, industria, finanzas, tráfico rodado, asistencia, recaudación de impuestos, entretenimiento, funcionalidad de los hogares (la llamada domótica), conflictos armados, etc.[4].

---

* Catedrático (Professore ordinario) de Derecho Penal en la Universidad de Roma "UnitelmaSapienza".

[1] Véase, en la doctrina penal, por ejemplo, F. Basile, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, en *Diritto penale e intelligenza artificiale. "Nuovi scenari"*, editado por G. Balbi, F. De Simone, A. Esposito y S. Manacorda, Torino, 2022, 3 y s. Simplificando al extremo, podemos definir la IA como cualquier máquina capaz de realizar tareas específicas típicamente asociadas al intelecto humano, mediante algoritmos, programas informáticos y sistemas electrónicos.

[2] Schwab, *La cuarta revolución industrial*, Barcelona, 2016, *passim.*

[3] Según la opinión del gran físico, cosmólogo y matemático Stephen Hawking, fallecido en 2018: ver *AI will either 'transform or destroy' society, says Prof Stephen Hawking at intelligence centre launch*, en *{www.cambridge-news.co.uk}.*

[4] Sobre algunas de estas aplicaciones y sus repercusiones jurídicas, véanse las contribuciones en *Automazione, Diritto e Responsabilità,* editado por L. Picotti, Napoli, 2023, espec. 213 y ss., 293 y ss.

Por lo tanto, los ordenamientos jurídicos están llamados a hacer frente a las ambivalentes repercusiones de las llamadas máquinas inteligentes de última generación, que, por un lado, representan un factor de mejora de las condiciones de vida de naciones y pueblos enteros, al ser capaces de fomentar el crecimiento humano y social, y, por otro, constituyen una fuente de riesgos potencialmente desmesurados.

Por estas razones, la IA se ha convertido en un campo de investigación muy estimulante también para los estudiosos del Derecho penal, como revela un considerable número de investigaciones y estudios en toda Europa y en todo el mundo. En este contexto, destacan las investigaciones sobre la responsabilidad penal o administrativa de las personas jurídicas u otras entidades colectivas y, en particular, de las sociedades mercantiles, que figuran entre los mayores usuarios de las nuevas tecnologías. Esto se debe a dos razones, que afectan tanto al aspecto relativo a la prevención como a la represión de delitos.

La primera razón es que las técnicas de inteligencia artificial, en combinación con la denominada *blockchain*, probablemente cambiarán radicalmente los métodos y prácticas de cumplimientopenal y el diseño de sistemas de control interno en las empresas. En el contexto general de lo que puede designarse como el "mundo RegTech"[5], las empresas, utilizando sofisticados sistemas informáticos, pueden disparar señales de alarma (*red flags*) que de otro modo no se hubieran podido identificar a través de las técnicas clásicas de análisis y supervisión. Además, pueden construir sistemas de toma de decisiones transparentes y fiables, en los que sea más complicado ocultar actividades ilícitas[6].

---

[5] El término hace referencia a la "tecnología reguladora", es decir, el uso de la tecnología digital para optimizar las actividades de gestión reguladora y ayudar a las empresas en los procesos de cumplimiento normativo y los controles operativos.

[6] Sobre este tema, incluido el uso de la tecnología no sólo como medio de prevención del delito, sino también como herramienta para la comisión de delitos, véase L. Picotti, *New Technologies as Tools for and Means Against Crime: Substantial Aspects*, en *Revue Internationale de Droit Pénal*, 2020, 2, 183. En la literatura española, véase el número especial de la *Revista Estudios Penales y Criminológicos* sobre "Inteligencia artificial y sistema penal", editado por C. Fernández Bessa y X. Ferreiro Baamonde, 2023, disponible en el siguiente enlace: *{https://revistas.usc.gal/index.php/epc/issue/view/556}*. Para una visión general reciente y una revisión más amplia de la literatura relevante, véase Birritteri, *Corporate Criminal Liability and New Technologies: Digital Compliance Strategies in the Fight against Economic Crimes*, en AA.VV., *The Role of Technology in Preventing and Combating Organised Crime, Financial Crimes and Corruption - Book of Proceedings*, OSCE, 2023, 11. Véase también, en general, Sabia, *Artificial Intelligence and Environmental Criminal Compliance*, en *Revue Internationale de Droit Pénal,* 2020, 1, 179 y ss.; Severino, *The Importance of Corporate Compliance in the Digital Era*, en *Revue Internationale de Droit Pénal, 2021*, 2, 435 y ss.; Mongillo, *Presente e futuro della* compliance *penale*, en *{www.sistemapenale.it}*, 11 de enero de 2022; Gullo, voce *Compliance*, en *Studi in onore di Carlo Enrico Paliero*, editado por C. Piergallini, G. Mannozzi, C. Sotis, C. Perini, M. Scoletta y F. Consulich, Milano, 2022,

Obviamente, todo lo dicho hasta aquí pertenece al mundo de las aspiraciones; la realidad podría ser menos idílica, ya que el nuevo hito de la *digital compliance* también presenta grandes incógnitas y riesgos que hay que gestionar. Entre las cuestiones que exigen atención, podemos mencionar, por ejemplo: la calidad de los datos en los que se basan los sistemas de inteligencia artificial[7]; el impacto sobre la protección de los datos personales y los derechos de los trabajadores, con innegables implicaciones de carácter ético; el papel de la ciberseguridad en la salvaguarda de los datos sensibles objeto de los procesos de cumplimiento digital; las disparidades en el acceso a los recursos digitales; o los efectos macroeconómicos sobre el empleo y la organización del trabajo, que requieren una cuidadosa consideración de las implicaciones sociales y políticas. De todo ello se deduce que, incluso desde la perspectiva de la "conformidad" legal, la regulación y gestión de las herramientas de IA aconsejan la adopción de un enfoque holístico que tenga en cuenta los aspectos tecnológicos, éticos, jurídicos, de seguridad y sociales.

La segunda razón del especial interés que los instrumentos de IA hoy suscitan entre los juristas, incluidos los penalistas, es aún más relevante desde la óptica del impacto social: su uso en el contexto de una empresa puede ocasionar peligros o daños de diversa índole, significativos desde el punto de vista de la comisión de delitos (dolosos o culposos). Piénsese en los sectores, ya de gran importancia y actualidad, de los coches de conducción autónoma (*self-driving cars*)[8] [9], la robótica y

---

1289 ss.

[7] En una conferencia celebrada el 22 de junio en la Universidad de Roma UnitelmaSapienza, el Director de la Unidad de Información Financiera (UIF) italiana, Enzo Serata, señaló algunas cuestiones críticas, observando cómo la completa dependencia, en el sector privado, de la supervisión de las transacciones a través de algoritmos inteligentes podría llevar a una disminución en la calidad general de los informes de operaciones sospechosas (S.O.S.), haciendo que la evaluación del informante individual siga siendo crucial para garantizar una información de calidad y evitar distorsiones que compliquen la realización de las tareas asignadas a la UIF. Sobre la *compliance* predictiva, también en el ámbito de la lucha contra el blanqueo de capitales, véase A. Esposito, *Note sparse sull'intelligenza artificiale,* en *Diritto penale e intelligenza artificiale,* cit., 42-47; J. Gimeno, *Instrumentos actuales de policía y justicia predictiva en el proceso penal español: análisis crítico y reflexiones* de lege ferenda *ante aplicaciones futuras*, 27 noviembre 2023, en *Inteligencia artificial y sistema penal*, cit.,

[8] A nivel monográfico, exhaustivamente, M. Lanzi, Self-driving cars *e responsabilità penale. La gestione del "rischio stradale" nell'era dell'intelligenza artificiale,* Torino, 2023.

[9] Pensemos en los robots que se utilizan desde hace tiempo en operaciones quirúrgicas, algoritmos de diagnóstico-terapéutico, etc. Sobre esta dimensión inédita del riesgo, véase, más recientemente, el estudio de N. Amore y E. Rossero, *Robotica e intelligenza artificiale nell'attività medica. Organizzazione, autonomia, responsabilità. Una ricerca sociologica e giuridico penale,* Bolonia, 2023, y en particular los capítulos sobre cuestiones de Derecho penal escritos por N. Amore, ivi, 101 y ss., 145 y ss. y 177 y ss.

los sistemas de diagnóstico médico, el *trading* financiero o la gestión logística mediante algoritmos.

Como ya se ha dicho, el problema se ve amplificado por el siguiente dato fáctico: gran parte de los sistemas de IA son producidos o utilizados por personas, en particular, con forma societaria. A este respecto, la cuestión jurídica clave es si las personas jurídicas involucradas pueden *de iure condito* o deben *de iure condendo* ser llamadas a responder en procedimientos penales o para-penales (por ejemplo: en el ordenamiento jurídico italiano, en virtud del Decreto Legislativo nº 231 de 8 de junio de 2001; en sistema penal español, en virtud de los artículos 31-*bis* y siguientes del Código penal) por delitos relacionados con el uso de IA.

En este artículo pretendemos centrarnos en esta última perspectiva, explorando, en primer lugar, la posibilidad de que una organización pueda ser considerada responsable de la comisión de delitos de diversa índole vinculados al uso de dispositivos o algoritmos de IA. Para ello examinaremos el marco normativo actual[10], exploraremos las técnicas jurídicas que parecen en abstracto viables para responsabilizar a las personas jurídicas en estas peculiares esferas de riesgo, y abordaremos los posibles obstáculos y argumentos que se oponen a tal enfoque jurídico[11]. Por último, extraeremos conclusiones de las consideraciones realizadas[12].

II. Los regímenes de responsabilidad penal (o para-penal) de las empresas a nivel internacional y los delitos relacionados con la IA: el escenario actual

Ante el dilema de si una persona jurídica puede ser considerada responsable de un delito caracterizado, en la fase de preparación o de ejecución, por el uso de herramientas de IA, hay que distinguir entre los sistemas informáticos no autónomos y los sistemas que pueden tomar decisiones totalmente autónomas, que los propios programadores o usuarios finales son incapaces, en todo o en parte, de prever[13].

[10] Véase el párrafo 2.

[11] Véase el párrafo 3.

[12] Véase el párrafo 4.

[13] Para una visión general véase. S. Beck, *The Problem of Ascribing Legal Responsibility in the Case of Robotics*, en 31 *AI & Soc.*, 2016, 473 y ss.

1. Sistemas de IA no autónomos y sistemas autónomos utilizados intencionalmente para cometer delitos

Partiendo de la primera constelación de casos, los sistemas *no autónomos* no parecen plantear problemas adicionales a los abordados tradicionalmente en el Derecho penal de las personas físicas o en el Derecho punitivo de las entidades pluripersonales.

Tales dispositivos, de hecho, se limitan a ejecutar instrucciones recibidas del programador humano o, en cualquier caso, operan bajo el control directo de una o varias personas físicas. En consecuencia, pueden activarse las vías "ordinarias" o usuales de *enforcement* e imputación de responsabilidad penal a la persona física y, eventualmente, también a la persona jurídica[14].

En efecto, tanto en los casos en que la herramienta tecnológica se utilice con el propósito específico de cometer un delito contra la vida, la seguridad, la privacidad personal, el patrimonio ajeno, etc., como en las situaciones en que el uso del sistema cause un daño involuntario (es decir, imputable a título de imprudencia) y susceptible de constituir una infracción penal[15], la responsabilidad podrá ser invocada sobre la base de los principios y regulas habituales en los diferentes ordenamientos jurídicos. Podrán ser considerados responsables del delito, según el caso, el productor, el distribuidor o el usuario final del sistema[16], así como la persona jurídica o las personas jurídicas según corresponda. Respecto a la entidad colectiva, los modelos de imputación varían desde las formas basadas en la responsabilidad vicaria (*vicarious liability*) o en la responsabilidad objetiva (*strict liability*), hasta las centradas en la culpa de organización (*organizational fault*) o en la falta de prevención de delitos específicos (*failure to prevent model*)[17].

Obviamente, incluso entre estos casos más "sencillos", pueden surgir, en la práctica, cuestiones delicadas de atribución de responsabilidad, tanto desde el punto de vista del autor o cómplice individual como de la organización a la que pertenecen. Como ya se ha mencionado, estas

---

[14] Sobre este tema, véase también B. Panattoni, *AI and Criminal Law: the Myth of 'Control' in a Data-Driven Society*, en *Revue Internationale de Droit Pénal*, 2021, 1, 125 y ss.

[15] Caldwell *et al.*, *AI-enabled future crime*, en 9 *Crime Science*, 2020, 14, revisaron los posibles usos de los sistemas de IA en la comisión de delitos. En la doctrina italiana, véase, entre otros, B. Magro, *A..I.: la responsabilità penale per la progettazione, la costruzione e l'uso dei robot,* en *il Quot. giur.*, 12 de junio de 2018.

[16] En la literatura italiana, véase C. Piergallini, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, en *Riv. it. dir. proc. pen.*, 2020, 1753.

[17] Sobre este punto véase también Fe. Mazzacuva, *The Impact of AI on Corporate Criminal Liability: Algorithmic Misconduct in the Prism of Derivative and Holistic Theories*, en *Revue Internationale de Droit Pénal*, 2021, 1, 143 y ss.

dificultades aplicativas no suelen diferir de las que se plantean normalmente en los casos de responsabilidad por productos defectuosos[18]. Sin embargo, los problemas tradicionales de imputación se ven agravados, en el contexto que nos concierne, por la incidencia en la dinámica fáctica de tecnologías complejas como las de la IA, especialmente en lo que respecta a los métodos utilizados en la fabricación y la marcada complejidad técnica del *output* de la producción.

En particular, desde la perspectiva de la responsabilidad individual, se amplían – para mencionar sólo las más evidentes – las cuestiones dogmáticas relacionadas con la identificación de los responsables y las causas penalmente relevantes del resultado, la distinción entre acción y omisión, las fuentes legales y la delimitación de las posiciones de garantía de los distintos actores en el proceso de producción, programación y comercialización de la máquina, la previsibilidad *ex ante* del eventual carácter defectuoso del producto, los requisitos de la cooperación imprudente y la acumulación de culpas, el papel del principio de confianza en Derecho penal.

También tiene trascendencia la tendencia a distribuir el proceso productivo en cadenas de suministro extremadamente fragmentadas: un aspecto que no sólo caracteriza a la industria de la IA, sino que aquí se eleva a la máxima potencia. La figura del "productor" se descompone en una miríada de operadores económicos, distribuidos nacional y globalmente, para cada componente del producto final. Esto complica la reconstrucción de las posiciones de garantía y los eslabones causales, y agrava las dificultades de previsión de los efectos del conjunto por parte de cada uno de los participantes en la cadena causal. Pero incluso cuando el proceso de producción se concentra en una única organización, la imagen monolítica del "fabricante" o "diseñador" corresponde a un uso sincopado del lenguaje, dado que en los procesos de construcción de máquinas inteligentes intervienen multitud de técnicos que aportan su contribución al proyecto unitario.

Además, pueden surgir problemas de jurisdicción, debido a la división nacional e incluso mundial de los diferentes segmentos productivos[19]. Desde este último punto de vista, la relación entre vendedor y

[18] C. Piergallini, *Inteligencia artificial,* cit., 1753. En idioma español, en general y por todos: W. Hassemer y F. Munoz Conde, *La responsabilidad por el producto en derecho penal*, Barcelona, 1995; S. Escobar Vélez, *La responsabilidad penal por productos defectuosos*, Valencia, 2012; J.M. Paredes Castañón y T. Rodríguez Montañés, *El caso de la colza. Responsabilidad penal por productos adulterados o defectuosos*, Valencia, 1995.

[19] Sobre el impacto de la globalización económica en la aplicación de la ley penal en el espacio, remitimos a V. Mongillo, Forced labour *e sfruttamento lavorativo nella catena di fornitura delle imprese: strategie globali di prevenzione e repressione*, en *Riv. trim. dir. pen. econ.*, 2019, 3-4, 630 y ss., párr. 6 espec.; Id., *Criminalità di impresa transfrontaliera e giurisdizione penale "sconfinata":*

consumidor también puede estar profundamente fragmentada geográficamente. Baste considerar que hoy en día cualquiera puede comprar en línea herramientas de IA – drones, por ejemplo – a vendedores que residen en terceros países.

En cuanto a la concreta cuestión de la posible imputación de responsabilidad a una entidad colectiva, también deben tenerse en cuenta las peculiaridades de los distintos regímenes normativos. Sólo a título de ejemplo, considérese el caso en que la puesta en circulación de un vehículo autónomo bajo la supervisión de un conductor humano lleve a la comisión de un homicidio imprudente y, sin embargo, las normas aplicables en materia de responsabilidad de las empresas no incluyan este delito entre los que pueden dar lugar a la responsabilidad de la *societas*. Evidentemente, este problema de aplicación sólo puede plantearse en los sistemas legales nacionales basados en un "catálogo cerrado" (*numerus clausus*) de delitos de los que puedan responder las personas jurídicas[20].

Los resultados de esta primera fase de nuestro análisis también pueden reproducirse sin demasiada dificultad con respecto a los *sistemas de IA que son completamente autónomos* pero que han sido diseñados desde el principio con el *objetivo de cometer delitos* y causar intencionadamente daños a terceros, por ejemplo, con fines terroristas o de desestabilización de gobiernos legítimos.

En estos casos, las aplicaciones de IA actúan como una especie de *longa manus* de quienes pretenden perpetrar hechos delictivos[21]. Por consiguiente, la responsabilidad penal podrá atribuirse – como se ha subrayado anteriormente – a las personas físicas que utilicen la herramienta con intenciones criminales, así como a la persona jurídica a la que pertenezcan, cuando la legislación vigente lo permita.

Además, en caso de condena de la persona física y/o jurídica, el dispositivo lesivo puede, por regla general, ser decomisado como *instrumentum sceleris*, es decir,como instrumento utilizado para cometer el delito, en el ordenamiento jurídico español según el art. 127 del código penal y en el ordenamiento jurídico italiano en virtud del art. 240 del código penal.

---

il difficile equilibrio tra efficienza e garanzie, en *Riv. it. dir. proc. pen.*, 2023, 1, 111 y ss.

[20] Como es el caso de nuestro Decreto Legislativo nº 231 de 8 de junio de 2001: para una visión general (en inglés) véase C. De Maglie, Societas Delinquere Potest? *The Italian Solution*, en *Corporate Criminal Liability. Emergence, Convergence and Risk*, editado por M. Pieth y R. Ivory, Dordrecht, 2011, 255 y ss.

[21] Sobre este tema, véase también I. Salvadori, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, en *Riv. it. dir. proc. pen.*, 2021, 1, 83 y ss.

2. Sistemas de IA completamente autónomos y no programados/utilizados para cometer delitos

Como se ha anticipado, los que se acaban de reseñar son los casos más sencillos, los *easy cases*, como diría Herbert Hart.

La cuestión se vuelve mucho más nebulosa y compleja en relación con los sistemas de IA *completamente autónomos* que *no se programan* ni se *utilizan para cometer delitos*. Estos son los verdaderos *hard cases*, los que más ocuparán a los tribunales tan pronto como los escenarios que hoy tememos comiencen a convertirse en una posibilidad más realista. Nos referimos a herramientas digitales estructuradas *by design* para aprender automáticamente y actuar "*solas*", es decir, capaces de percibir su entorno, interactuar con él, analizar datos, hacer previsiones, tomar decisiones y provocar modificaciones en la realidad externa con total independencia tanto del productor del sistema como del usuario[22].

En la literatura científica se habla de *machine learning* (aprendizaje automático), cuya última evolución son las sofisticadas técnicas de *deep learning* (aprendizaje profundo)[23]. La máquina, en este caso, es un sistema abierto[24] que aprende de forma continua y automática, con la consecuente modificación de las conexiones entre neuronas artificiales, por lo que desde este prisma lo que se pretende es imitar, en la medida de lo posible, la "plasticidad" del cerebro humano y el cambio incesante de las redes neuronales que lo componen[25].

Desde la perspectiva del Derecho penal, el problema más delicado que suscitan estos avances tecnológicos es precisamente la imposibilidad de predecir, al menos en su totalidad, el funcionamiento futuro del sistema, *ergo* todas las decisiones que, en las infinitas situaciones de la vida real, el dispositivo de IA podrá tomar independientemente de la instrucción o autorización de una guía humana.

Si las máquinas inteligentes estuvieran totalmente *automatizadas*, este problema no se plantearía, ya que por definición se puede automatizar todo lo que es predecible. Por el contrario, los dispositivos de última generación capaces de autoaprendizaje toman decisiones a

---

[22] Sobre este punto, desde la perspectiva específica de las armas autónomas y sus implicaciones para el Derecho penal, véase R. Crootof, *War Torts: Accountability for Autonomous Weapons*, en 164 *University of Pennsylvania Law Review*, 2016, 1347 y ss.

[23] Cf. T. Sejnowski, *The Deep Learning Revolution*, Cambridge, 2018.

[24] Véase también B. Magro, Robot, cyborg *e intelligenze artificiali*, en *Cybercrime*, en *Omnia. Trattati giuridici*, editado por A. Cadoppi, S. Canestrari, A. Manna y M. Papa, Torino, 2019, 1191.

[25] Las redes neuronales artificiales están "compuestas por elementos interconectados que funcionan sincrónicamente sobre el modelo de las neuronas biológicas y sus sinapsis": así, R. Bodei, *Dominio e sottomissione. Schiavi, animali, macchine e Intelligenza Artificiale*, Bolonia, 2019, 317.

través del contacto con el entorno externo y los datos almacenados en el *cloud*, "un potente *hub* computacional capaz de almacenar, procesar y proporcionar enormes cantidades de datos[26], de los cuales extraer continuamente para las actualizaciones, los *upgrades*"[27]. La consecuencia es que ni el diseñador, ni el programador, ni el usuario final pueden conocer exactamente de antemano el *pattern* decomportamiento que la máquina elegirá, una y otra vez, al interpretar las infinitas situaciones de la vida real[28].

En algunos aspectos, una dosis de imprevisibilidad está incluso "preordenada", ya que el objetivo del fabricante de estos dispositivos avanzados no es instruir y regular de antemano cualquier elección de la máquina, sino hacer que la tecnología "pensante" funcione y tome sus decisiones de una manera que se espera que sea lo más eficaz posible. Ésta es también la única declinación concebible del principio de confianza en la relación hombre-máquina inteligente. Pero la principal novedad, frente a las versiones clásicas del principio de confianza en la teoría del delito imprudente, es que en este caso no se trata de confiar en una persona humana de la que se tienen razones para creer que tiene la suficiente experiencia, formación y prudencia para realizar determinadas tareas, sino en una máquina que aprende, hipotéticamente dispuesta a un continuo proceso de aprendizaje basado en la experiencia y perfeccionamiento de sus "habilidades". Aquí, como se ha dicho, reside el *punctum dolens* penal de las tecnologías en cuestión.

Es evidente entonces cómo, a través de esta inédita dimensión empírica, se pasa del *peligro previsible* y en todo o en parte cuantificable, que representa el dominio de la prevención, al *riesgo desconocido*, que se sitúa, electivamente, en el dominio de la llamada precaución. De hecho, la idea de *prevención* está marcada por el conocimiento científicamente corroborado. El *principio de precaución*, en cambio,a se refiere a los ámbitos de riesgo caracterizados por una considerable incertidumbre científica, de modo que se refiere a lo que se supone que pueda ocurrir, pero no se sabe si – y en qué términos – ocurrirá.

En una sociedad que tiende a rechazar la idea de lo "fortuito", es fácil predecir una enérgica demanda de justicia por parte de las víctimas ante cualquier fallo del algoritmo presuntamente "inteligente". Sin embargo, según la opinión doctrinal dominante, en la fundamentación

[26] Se habla, en este sentido, de *big data.*

[27] R. Bodei, *Dominio e sottomissione,* cit., 315.

[28] Sobre la cuestión de la imprevisibilidad tecnológica de las "máquinas inteligentes", véase, por ejemplo, A. Cappellini, *Reati colposi e tecnologie dell'intelligenza artificiale*, en *Diritto penale e intelligenza artificiale*, cit., 23-25.

de la responsabilidad por imprudencia no cabe recurrir al principio de precaución como fuente jurídica supletoria del deber de cuidado.

En puridad, este criterio de imputación subjetiva "requiere, ante todo, la violación de reglas de cautela con fundamento nomológico, orientadas a la prevención de resultados previsibles (y no meramente hipotéticos) *ex ante*: sobre la base, por tanto, del patrimonio cognitivo disponible (para el agente modelo de referencia o, al menos, para el agente concreto que por azar posee conocimientos superiores) en el momento de la conducta, cuya naturaleza contraria a deber no puede ser afirmada de manera retroactiva"[29].

La *perspectiva orientada por el principio de precaución* se convierte así en un método de buena gestión administrativa del riesgo hipotético, que puede incluso aconsejar, en casos límite, la prohibición absoluta de actividades o del uso de tecnologías de las que se teme que puedan generar peligros aún no corroborados científicamente, pero que parecen previsiblemente insostenibles en cuanto a su gravedad y potencial difusión.

Una vez definido el marco conceptual de nuestra reflexión, no se puede excluir que las decisiones tomadas autónomamente por un sistema de IA puedan constituir, al menos desde un punto de vista material-objetivo, conductas penalmente típicas. Nos vienen a la mente ejemplos de abusos de mercado llevados a cabo mediante *software* de IA capaz de gestionar de forma autónoma transacciones bursátiles o financieras[30], o daños a la integridad física – homicidio o lesiones imprudentes – causados por robots/sistemas de IA dotados igualmente de autonomía operativa[31].

A veces, la experimentación sobre el terreno de una determinada máquina (por ejemplo, un dispositivo de diagnóstico o uno utilizado en intervenciones quirúrgicas) o los *feedback* de los clientes pueden

---

[29] D. Castronuovo, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato,* Roma, 2012, 47. En la doctrina española, sobre el principio de precaución y las muy diversas definiciones que se han otorgado al mismo, véase A. Galán Muñoz, *La problemática utilización del principio de precaución como referente de la política criminal del moderno derecho penal. ¿Hacia un derecho penal del miedo a lo desconocido o hacia uno realmente preventivo?*, en *Revista de estudios de la justicia*, 2015, n. 22, 69 y ss.

[30] Véase F. Consulich, *Il nastro di Mobius. Intelligenza artificiale e imputazione penale nelle nuove forme di abuso del mercato,* en *Banca, borsa e titoli di credito*, 2018, 2, 195 y ss.; M. Palmisano, *L'abuso di mercato nell'era delle nuove tecnologie.* Trading *algoritmico e principio di personalità dell'illecito penale*, en *Dir. pen. cont.,* 2019, 2, 129 y ss.; A.F. Tripodi, *Abusi di mercato e* trading *algoritmico*, en *Il diritto nell'era digitale. Persona, Mercato, Amministrazione,* editado por R. Giordano, A. Panzarola, A. Police, S. Preziosi y M. Proto, Milano, 2022, 745 ss.

[31] Véase, por ejemplo, R. Crootof, *War Torts*, cit., 1347 y ss.; así como, T. King, N. Aggarwal, M. Taddeo y L. Floridi, *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*, en 26 *Science and Engineering Ethics*, 2020, 89 y ss.

permitir al fabricante y/o al usuario tomar conciencia de que un determinado tipo de IA se desviará, en un cierto porcentaje de casos (aunque no sea exactamente cuantificable), de las funciones programadas, desencadenando cursos causales potencialmente dañinos que no pueden neutralizarse con medidas preventivas adecuadas, en el estado de los conocimientos científicos. En estos casos, podemos hablar de un mínimo *riesgo conocido*, que puede llevar a la autoridad estatal a la decisión jurídico-política de tolerarlo, cuando el efecto secundario adverso parezca de poca relevancia, o, por el contrario, a prohibirlo como riesgo jurídicamente inadmisible y por esto no permitido.

Sin embargo, en el escenario característico del *machine learning* y, más aún, del *deep learning*, el conocimiento científico-experimental, por regla general, ni permite afirmar que el uso de un determinado sistema de IA pueda causar peligros o daños específicos incontrolables, ni excluirlos categóricamente.

En este contexto, la responsabilidad penal individual está expuesta a grandes incertidumbres y dudas potencialmente irresolubles. Pero lo mismo sucede con la cuestión relativa a la responsabilidad penal de las organizaciones empresariales implicadas, que sitúa al jurista ante un dilema difícil de resolver de modo unívoco. En todo caso, antes de responder a esta última cuestión, parece necesario abordar una hipótesis que, no obstante su carácter fantasioso, surge recurrentemente en el debate científico: la idea de atribuir *personalidad jurídica* a los sistemas de inteligencia artificial como tales y *castigar directamente a la máquina*[32], en presencia de un mal funcionamiento que genere perjuicios a intereses jurídicamente protegidos, penalmente trascendentes. Esta sugerencia evoca los juicios medievales contra los animales o

[32] Véase, por ejemplo, el pensamiento del penalista israelí G. Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems*, Dordrecth, 2015, que aboga por la responsabilidad penal directa de la IA; Id., *The Criminal Liability of Artificial Intelligence Entities - From Science Fiction to Legal Social Control*, en *Akron Intellectual Property Journal*, 2010, vol. IV, 171 y ss. IV, 171 y ss. *Contra*, G. Quintero Olivares, *La robótica ante el derecho penal: el vacío de respuesta jurídica a las desviaciones incontroladas*, en *Revista Electrónica de Estudios Penales y de la Seguridad*, {www.ejc.reeps}*.com,* 2017, 9, quien – en contra de la opinión expresada por A. Sánchez del Campo Redonet, *Cuestiones jurídicas que plantean los robots*, en *Revista de privacidad y derecho digital*, 2016, 2 – observa que es imposible considerar a los robots capaces de cometer delitos y a la destrucción física de la máquina como equivalente a la pena de muerte, pues la condición ontológica de la ley penal es que sus posibles reacciones sean conocidas o conocibles *ex ante* por hipotéticos futuros delincuentes, del mismo modo que los mandatos y prohibiciones penales (c.d. conocibilidad de la ley). El autor concluye señalando cómo rechazar la responsabilidad penal de la máquina "no equivale a la irrelevancia de lo que "haga" una máquina" (p. 10). Sobre esta cuestión, véase también F Basile, *La responsabilità penale dei sistemi di intelligenza artificiale: scienza o fantascienza?*, en *Automazione,* cit., 103 y ss.; S. Preziosi, *La responsabilità penale per eventi generati da sistemi di AI o da processi automatizzati*, en *Il diritto nell'era digitale,* cit., 722 y ss.

incluso el *animismo* de los pueblos primitivos[33], que pasa por atribuir propiedades espirituales a realidades físico-materiales.

Es evidente, sin embargo, que la solución a los problemas jurídicos aquí abordados no puede venir, en el estado actual de modernidad hipertecnológica que nos rodea, de la exhumación del pensamiento arcaico más irracional y, por tanto, de una regresión – por decir lo menos grotesco – a las fases primitivas del desarrollo social; de hecho, hasta la fecha, ninguna legislación (al menos que sepamos) prevé medidas o sanciones penales directamente aplicables a las herramientas informáticas o dispositivos de IA[34].

En el estado actual de los conocimientos técnico-científicos, hay consenso unánime en que estos instrumentos supuestamente inteligentes carecen de capacidad de autodeterminación y, por tanto, de verdadera *autoconciencia* o identidad personal, entendida – al menos a partir del padre del empirismo moderno John Locke – como la conciencia que una persona tiene de su permanencia a través del tiempo y de las fracturas de la experiencia[35]. La autoconciencia es indispensable para fundar la responsabilidad penal *stricto sensu,* y de ella depende también la posibilidad de que cualquier medida punitiva (independientemente

[33] Cf. H. Kelsen, *Teoría pura del derecho* (1960), trad. esp., México, 1966, 45 y ss., 98 y ss.; Id., *Teoría general del derecho y del estado* (1945), trad. esp., México, 1995, 109, 127 y ss., 227, sobre la interpretación animista de la naturaleza de los pueblos primitivos (según esta interpretación se cree que "cada objeto del mundo sensible es considerado como lo morada de un espíritu invisible, amo del objeto, y que "tiene" a éste en la misma forma en que la substancia tiene sus cualidades, y el sujeto gramatical sus predicados"); el fenómeno antropológico anímico también ha sido cuidadosamente investigado, en clave psicoanalítica, por S. Freud, *Tótem y tabú*, en *Obras completas*, XIII, trad. esp., Buenos Aires, 1976, 79 ss.

[34] Véase el *General Report* de L. Picotti, en *Traditional Criminal Law Categories and AI,* cit., 11 y ss., basado también en los "informes especiales" y en el análisis de las respuestas de los grupos nacionales al cuestionario elaborado por la AIDP, en el que la A. señala que "In all the countries of the national reports collected, AI systems do not have legal personhood or legal capacity"; por lo tanto, «they cannot be considered as subject of criminal law either". Este problema se viene abordando en la literatura científica pertinente desde hace varios años, empezando por los estudios pioneros de L.B. Solum, *Legal Personhood for Artificial Intelligences*, en 70 *North Carolina Law Review*, 1992, 1231 y ss.

[35] De hecho, tal enfoque de la "personalidad", en el pensamiento filosófico, se remonta al menos a J. Locke, *An Essay Concerning Human Understanding,* London, 1690, *Ensayo sobre el entendimiento humano*, Mexico, 2013, lib. II, cap. XXVII, espec. n. 9 ("Porque, como el tener conciencia siempre acompaña al pensamiento, y eso es lo que hace que cada uno sea lo que llama sí mismo, y de ese modo se distingue a sí mismo de todas las demás cosas pensantes, en eso solamente consiste la identidad personal, es decir, la mismidad de un ser racional. Y hasta el punto que ese tener conciencia pueda alargarse hacia atrás para comprender cualquier acción o cualquier pensamiento pasados, hasta ese punto alcanza la identidad de esa persona"); n. 18 ("Es en esta identidad personal en lo que están fundados el derecho y la justicia de las recompensas y de los castigos [...]").Y ya Séneca, en *De la ira*, II, 26, decía: "¿No es demencia irritarse contra cosas que no pueden merecer ni sentir nuestra cólera?" (*His irasci quam stultum est, quae iram nostram nec meruerunt, nec sentiunt*).

de su calificación) pueda motivar psicológicamente al destinatario a cumplir la norma. En definitiva, un dispositivo de IA, "aun siendo "inteligente", no deja de ser siempre una máquina"[36]. El propio uso del término "inteligencia" refleja un lenguaje metafórico con el que atribuimos al dispositivo cualidades de las que, en realidad, carece[37].

En contra podría decirse que: muchos ordenamientos jurídicos nacionales ya admiten la responsabilidad penal de las personas jurídicas, en sí mismas carentes de consistencia psicofísica[38]. Sin embargo, la comparación no se sostiene: un algoritmo no sólo carece de *self-consciousness*,sino también del sustrato propio de una colectividad humana. En cambio, una *legal person* (persona jurídica) no es sólo una abstracción jurídica, sino una organización de personas y recursos, y por tanto también una colectividad de individuos de carne y hueso (como los directivos, empleados y colaboradores de una sociedad mercantil), motivados por un precepto normativo (mandato o prohibición) y la amenaza de sanciones por su incumplimiento.

Queda entonces por abordar la cuestión de si, frente a delitos en cuya dinámica causal hayan intervenido herramientas de IA completamente autónomas, la persona jurídica que – a través de sus miembros – haya hecho uso de esa tecnología, o que haya diseñado, producido, distribuido o vendido el dispositivo, también puede ser castigada.

Para ello son necesarias algunas aclaraciones preliminares. En primer lugar, conviene considerar que, a nivel internacional, la responsabilidad penal de las empresas está, por regla general, indisolublemente vinculada a la comisión de un delito por parte de una persona física (el denominado "hecho de conexión"; en alemán *Anknüpfungstat*)[39].

---

[36] Mensaje del Santo Padre Francisco para la 57ª Jornada Mundial de la Paz (1 de enero de 2024), 14.12.2023, en el que observa que los llamados sistemas inteligentes "son, a fin de cuentas, "fragmentarios", en el sentido de que sólo pueden imitar o reproducir algunas funciones de la inteligencia humana", y, con referencia específica a los sistemas autónomos de armas, se señala que "nunca sujetos moralmente responsables. La exclusiva capacidad humana de juicio moral y de decisión ética es más que un complejo conjunto de algoritmos, y dicha capacidad no puede reducirse a la programación de una máquina [...]".

[37] Bodei, *Dominio e sottomissione,* cit., 300.

[38] Así, en efecto, S. Gless, E. Silverman y T. Weigend, *If Robots Cause Harm, Who Is To Blame? Self-Driving Cars and Criminal Liability*, en 19 *New Criminal Law Review*, 2016, 3, 412 y ss.

[39] Véase, entre otros, *European Developments in Corporate Criminal Liability*, editado por J. Gobert y A.-M. Pascal, Oxon-New York, 2011 y C. De Maglie, *Models of Corporate Criminal Liability in Comparative Law*, en 4 *Wash. U. Global Stud. L. Rev.*, 2005, 547. Para una visión de conjunto de los principales modelos, también fuera de Europa, véase M. Pieth y R. Ivory, *Emergence and Convergence: Corporate Criminal Liability Principles in Overview,* en *Corporate Criminal Liability,* cit., 3 ss. En la bibliografía italiana, véase V. Mongillo, *La responsabilità penale tra individuo ed ente collettivo,* Torino, 2018, 175-262; Id., *The Allocation of Responsibility for Criminal Offences between Individuals and Legal Entities in Europe*, en *Corporate Criminal Liability and Compliance Programs*, editado por A. Fiorella, Napoli, 2012, vol. II, *Towards a Common Model in the European Union*,

Sin embargo, en los casos mencionados anteriormente, es difícil – y la mayoría de las veces imposible – afirmar la responsabilidad penal de una persona física.

Para empezar, puede resultar imposible la demostración judicial de la existencia del tipo objetivo del delito (en inglés, *actus reus*), por ejemplo, en lo que respecta al nexo causal entre el resultado final ofensivo y un error de diseño cometido por un individuo, teniendo en cuenta además que muchas personas – y a veces variadas galaxias empresariales y multiempresariales – cooperan para crear y lanzar determinados productos al mercado.

Pero aparte de ello, el verdadero e insuperable obstáculo suele ser la prueba del tipo subjetivo (*mens rea*), ya que, en los casos en cuestión, no es posible imputar la responsabilidad penal a un individuo del que resulta descartada la intención de cometer un delito y que ni siquiera podía prever el comportamiento del sistema informático, salvo la simple y abstracta posibilidad de que algo negativo pudiera salir mal: pero en este caso estaríamos volviendo no al campo de la prevención, sino al de la precaución[40] o, a lo sumo, al de una genérica e inconsistente *previsibilidad de la imprevisibilidad.* Como ya hemos señalado, de hecho, estos dispositivos inteligentes toman decisiones autónomas e impredecibles, a la vez que los mecanismos por los que aprenden y toman determinadas decisiones son a menudo desconocidos (el nodo de la llamada *"black-box"*)[41].

En resumen, la imposibilidad – ya sea teórica o práctica – de responsabilizar penalmente a un ser humano (por la vía del dolo, la imprudencia o hipótesis intermedias como la *recklessness* inglesa), significa que tampoco se podrá responsabilizar y castigar a una *corporation.*

---

121 y ss.; G. De Simone, *Profili di diritto comparato*, en *Responsabilità da reato degli enti,* editado por G. Lattanzi y P. Severino, Torino, 2020, vol. I, *Diritto sostanziale*, 3 y ss.

[40] Sobre la imposibilidad de aplicar el principio de precaución para integrar normativamente el delito imprudente, con referencia a la cuestión que nos ocupa, véase G. Quintero Olivares, *La robótica ante el derecho penal*, cit., 1 y ss.

[41] Sobre estos aspectos, véase C. Grandi, *Positive obligations (*Garantestellung*) grounding the criminal responsibility for not having avoided an illegal result connected to the AI functioning,* en *Traditional Criminal Law Categories and AI,* cit., 67 y ss. Véase también A. Moraiti, *AI Crimes and Misdemeanors: Debating the Boundaries of Criminal Liability and Imputation*, en *Revue Internationale de Droit Pénal*, 2021, 1, 109 y ss.; U. Pagallo, *From Automation to Autonomous Systems: A Legal Phenomenology with Problems of Accountability*, en *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, IJCAI, 2017; en la literatura italiana, B. Magro, *Decisione umana e decisione robotica. Un ipotesi di responsabilità da procreazione robotica,* en *Leg. pen., 2020,* 1 y ss.; B. Fragasso, *La responsabilità penale del produttore di sistemi di intelligenza artificiale*, en *Dir. pen. cont.*, 2023, 1, 31 y ss. Para una visión general de los vínculos entre la IA y el derecho, véanse las diversas contribuciones publicadas en *Regulating Artificial Intelligence*, editado por T. Wischmeyer y T. Rademacher, Cham, 2020.

Esto es indudablemente cierto en el caso de los *modelos de heterorresponsabilidad*, es decir, aquellos en los que la persona jurídica responde indirectamente por el delito cometido por una persona física dentro del marco de la relación que lo une a la organización y en el interés de esta. El tradicional mecanismo estadounidense de imputación, conocido como *vicarious liability* (responsabilidad vicaria)[42], implica, de hecho, que las corporaciones sólo pueden ser consideradas culpables y castigadas cuando sea posible identificar a una persona física que ha realizado una conducta que cumple todos los requisitos objetivos y subjetivos del delito relevante[43].

Algunos estudiosos han reflexionado sobre la posible extensión de la *corporate mind* – en la que fundamentar una imputación de responsabilidad – también a los fallos algorítmicos que contribuyen a causar daños a terceros, argumentando la posibilidad de imaginar ilícitos "cometidos" por sistemas digitales y considerarlos ilícitos corporativos (*corporate wrongs*), similares a los perpetrados por empleados[44].

Se trata de construcciones claramente artificiosas porque comparan datos empíricos objetivamente incomparables. Partiendo de esta incongruente asimilación, este planteamiento consideraría al algoritmo como un *agent* de la organización y, sobre todo, le atribuiría una culpabilidad/*mens rea* que luego imputaría a la persona jurídica. El planteamiento es inasumible, debido a la mencionada imposibilidad de captar un elemento de *Gewissen* (conciencia moral), o más sencillamente autoconciencia, en la máquina.

La respuesta a tal cuestión sólo puede ser negativa, y ello es así aun asumiendo el recurso a un modelo diferente de responsabilidad,

---

[42] Observa que los modelos vicariales anglosajones como el estadounidense tienen, en todo caso, "una larga tradición de complementar el punto de partida vicarial con una visión más organizativa en el ámbito de la determinación de la pena", B. Feijoo Sánchez, *La función de la responsabilidad penal de las personas jurídicas en el derecho penal español*, en *REDEPEC*, vol. nº 1, 2023, 17, del cual véanse también las agudas consideraciones sobre las ventajas del modelo latino frente a modelos vicariales como el alemán (p. 37 y ss.)

[43] Véase V.P. Nanda, *Corporate Criminal Liability in the United States: Is a New Approach Warranted?*, en *Corporate Criminal*, cit., 63 y ss.; C. Wells, *Corporations and Criminal Responsibility*, New York, 2018, *passim.* Para un examen de las prácticas de *enforcement* en los Estados Unidos en relación con este modelo de responsabilidad penal corporativa, véase, en una bibliografía interminable, J. Arlen, *Corporate Criminal Enforcement in the United States: Using Negotiated Settlements to Turn Corporate Criminals into Corporate Cops,* en 17.12 *NYU School of Law Public Law Research Paper*, 2017, 1 y ss.; B. Garrett, *Too Big to Jail. How Prosecutors Compromise with Corporations*, Belknap, 2014.

[44] Se hace referencia en particular al trabajo de M.E. Diamantis, *The Extended Corporate Mind: When Corporations Use AI to Break the Law*, en 98 *North Carolina Law Review*, 2020, 893. Sobre este tema, véase también el análisis y las soluciones propuestas por R. Abbot y A.F. Sarch, *Punishing Artificial Intelligence: Legal Fiction or Science Fiction*, en 53 *UC Davis Law Review*, 2019, 323.

basado en la teoría de la identificación, también de origen anglosajón (*identification doctrine*), aunque también con ascendientes en la teoría del órgano propia del derecho público de matriz continental[45]. En su caso, conllevaría el problema adicional de cómo concebir el algoritmo como la "*directing mind and will of the company*"[46].

Por último, los resultados no cambian al dirigir nuestra atención a los ordenamientos jurídicos en los que la responsabilidad de la persona jurídica se basa en un requisito de "culpa de organización"[47] o, de forma similar, se construye en términos de no prevención del delito *(failure to prevent).* La idea que subyace a estas disciplinas es que las organizaciones pueden ser consideradas responsables por no haber implementado *compliance programs* o procedimientos adecuados para prevenir la comisión de delitos específicos.

No se ignora cómo, en determinadas condiciones, muchos de estos regímenes de responsabilidad empresarial admiten la posibilidad de declararla de forma "autónoma", es decir, prescindiendo de la identificación material del autor del delito (la llamada "culpabilidad anónima"o *anonymous guilt*). En este paradigma "autonomista" podemos situar, sin duda, el art. 8 del Decreto Legislativo italiano nº 231/2001[48] o el art. 31-*ter*, párrafo 1, del Código penal español. Sin embargo, incluso estos modelos de responsabilidad siguen anclados en la necesidad de determinar la comisión de un delito en todos sus elementos esenciales, objetivos y subjetivos, o al menos los factores objetivos de la conducta humana y del resultado, permitiendo como mucho que la autoridad judicial prescinda de la identificación del autor individual concreto[49].

---

[45] V. Mongillo, *La responsabilità penale tra individuo ed ente collettivo,* cit., espec. 123 y ss.

[46] Véanse: C. Wells, *Corporate Criminal Liability in England and Wales: Past, Present and Future,* en *Corporate Criminal Liability,* cit., 91 y ss.; J. Gobert, *Corporate Criminality: Four Models of Fault*, en 14 *Legal Stud.*, 1994, 393.

[47] Con referencia a la experiencia italiana, véase V. Mongillo, *El defecto de organización: enigma y esencia de la responsabilidad «penal» de las personas jurídicas en la experiencia de la aplicación italiana*, in *La Ley Compliance penal*, n. 14, junio-septiembre, 2023, 1 y ss.

[48] Sobre los problemas de aplicación que plantea esta disposición, véase V. Mongillo, *La responsabilità penale tra individuo ed ente collettivo,* cit., 310.

[49] Sobre las normas que rigen la responsabilidad penal de las empresas, véase, por lo que respecta al marco jurídico italiano y para una visión general, C. De Maglie, *Societas Delinquere Potest?*, cit., 255. Con referencia al modelo británico del *failure to prevent*, véase, entre otros, C. Wells, *Corporate Responsibility and Compliance Programs in the United Kingdom,* en *Preventing Corporate Corruption. The Anti-Bribery Compliance Model*, editado por S. Manacorda, F. Centonze y G. Forti, Cham, 2014; L. Campbell, *Corporate Liability and the Criminalization of Failure,* en 12 *Law and Financial Markets Review*, 2018, 2, 57 y ss.; G.R. Sullivan, *The Bribery Act 2010: An Overview*, en 2 *Criminal Law Review*, 2011, 87 y ss.

III. Las perspectivas de futuro: ¿hacer responsables a las personas jurídicas por delitos relacionados con la IA?

Ha llegado el momento de reflexionar, de manera concisa, sobre las técnicas legislativas que podrían adoptarse, en el futuro próximo, para fundamentar la responsabilidad de una empresa por delitos determinados por el uso de sistemas de IA, y en particular, de aquellos completamente autónomos. Además, es necesario entender si las diversas estrategias político-criminales son adecuadas y justas o no.

Nuestra primera tesis es que este problema no debe abordarse desde una perspectiva unilateral y monista. Especialmente en relación con las nuevas tecnologías de IA, cualquier decisión de política criminal debería ser el último peldaño de un proceso más amplio y articulado de regulación pública, dentro del cual el legislador debería encargarse también de establecer las "reglas del juego" y los límites del denominado *erlaubtes Risiko* ("riesgo permitido")[50].

Las instituciones de la UE, recientemente, han adoptado importantes medidas en la dirección auspiciada, entre las que destaca el Reglamento sobre inteligencia artificial (la llamada "Ley europea de Inteligencia Artificial"), publicado en el Diario Oficial de la Unión Europea el 12 de julio 2024[51]. Las Organizaciones Internacionales también podrían adoptar iniciativas normativas similares, a escala global, mediante la celebración de acuerdos multilaterales y la coordinación de su aplicación e implementación[52].

El instrumento europeo que acaba de ser adoptado tiene como objetivo introducir normas comunes en el mercado único para garantizar la circulación de herramientas de IA seguras, prohibiendo las prácticas más peligrosas – incluyendo algunas formas muy controvertidas de *predictive policing* [53]– y estableciendo medidas específicas de cumplim-

[50] Sobre la importancia de este aspecto, véase también A. Fiorella, *Responsabilità penale del Tutor e dominabilità dell'Intelligenza Artificiale. Rischio permesso e limiti di autonomia dell'Intelligenza Artificiale*, en *Il diritto nell'era digitale,* cit., 656 y ss.

[51] Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.° 300/2008, (UE) n.° 167/2013, (UE) n.° 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial). El texto definitivo del Reglamento ha introducido importantes cambios respecto a las versiones anteriores, algunos de los cuales – como veremos – afectan precisamente al marco de la *predictive policing.*

[52] Como también pide el Santo Padre Francisco en su Mensaje para la 57ª Jornada Mundial de la Paz, cit.

[53] La versión final de la *Ley de IA* incluye entre las prácticas prohibidas las herramientas policiales predictivas que identifican a delincuentes potenciales basándose únicamente en la elaboración del perfil de una persona física o en la evaluación de los rasgos y características de

iento normativo, con inclusión de la supervisión del producto tras su comercialización, o la obligación de adoptar todas las medidas correctivas adecuadas para garantizar el buen funcionamiento del sistema de IA, su eventual retirada del mercado o su posible reintroducción en el mismo dentro del plazo que pueda prescribir la autoridad de supervisión del mercado, según la lógica de la *reactive fault* o del ilícito por omisión de reacción (véase, por ejemplo, los arts. 20 y 79, párrafo 3, del Reglamento)[54]. Los operadores que deseen comercializar *software* de IAde alto riesgo deberán cumplir de manera efectiva con estas prescripciones, y se impondrán obligaciones similares a los operadores para la distribución de los diversos dispositivos. El instrumento normativo de la UE también prevé que los Estados miembros (EM) establezcan o designen autoridades nacionales responsables de garantizar la aplicación de estas normas y que establezcan (véase el art. 99) "el régimen de sanciones" – que deberán ser "efectivas, proporcionadas y disuasorias" – "y otras medidas de ejecución, como advertencias o medidas no pecuniarias, aplicable a las infracciones del presente Reglamento que cometan los operadores [...]". Por último, se especifica que los Estados miembros deberán adoptar "todas las medidas necesarias" para garantizar una correcta y eficaz aplicación del referido Reglamento. En consecuencia, el texto deja un amplio margen de apreciación y decisión a los Estados miembros sobre los tipos de medidas sancionadoras a adoptar (así como sobre su contenido), más allá de los requisitos esenciales establecidos por el Reglamento[55].

---

su personalidad, aunque también especifica que "esta prohibición no se aplicará a los sistemas de IA utilizados para apoyar la evaluación humana de la implicación de una persona en una actividad delictiva que ya se base en hechos objetivos y verificables directamente relacionados con una actividad delictiva", es decir en una sospecha razonable (véase el artículo 5, apartado 1, letra d) y considerando 42). Esta disposición de compromiso atenuó la anterior prohibición general que había propuesto el Parlamento Europeo con respecto a estos sistemas en junio de 2023. Véase, para un análisis extenso y preciso, E. Pietrocarlo, Predictive policing: *criticità e prospettive dei sistemi di identificazione dei potenziali criminali*", en *Dir. pen. cont.*, 2023, 2, 145 y ss. En la literatura extranjera sobre el tema, véase, por todos, A.G. Ferguson, *Policing Predictive Policing*, en 94 *Washington Law Review*, 2017, 5, 1109 ss.

[54] Sobre el concepto de *reactive fault*, véase, en la pionera doctrina australiana, B. Fisse, *Reconstructing Corporate Criminal Law: Deterrence, Retribution, Fault, and Sanctions*, 1982-1983, in 56 *S. Cal. L. Rev.*, 1983, 1141 ss., spec. 1195 ss.; B. FISSE y J. Braithwaite, *The Allocation of Responsibility for Corporate Crime: Individualism, Collectivism and Accountability*, en 11 *Sydney L. Rev.*, 1988, 468 ss. En la doctrina italiana, véase V. Mongillo, *La responsabilità penale tra individuo ed ente collettivo*, cit., 377 ss. y, recientemente, el amplio volumen de A. Orsina, *La responsabilità da reato dell'ente tra colpa di organizzazione e colpa di reazione*, Torino, 2024, espec. 145 ss., 620 ss.

[55] De hecho, el artículo 99 del Reglamento también establece lo siguiente: "[...]3. El incumplimiento de la prohibición de las prácticas de IA a que se refiere el artículo 5 se sancionará con multas administrativas de hasta 35 000 000 EUR o, si el infractor es una empresa, de hasta el 7 % de su volumen de negocios total anual a escala mundial correspondiente al ejercicio anterior,

En nuestra opinión, antes de considerar cualquier medida penal o punitiva de aplicación directa a las empresas que producen o utilizan herramientas de IA, los Estados – incluso fuera de la UE y, por tanto, no sujetos a disposiciones europeas directamente aplicables – deberían contar con una legislación que regule la producción y venta de tales dispositivos tecnológicos y otras actividades en el sector. Dicha legislación

---

si esta cifra es superior. 4. El incumplimiento de cualquiera de las siguientes disposiciones relativas a los operadores u organismos notificados, distintas de las establecidas en los artículos 5, se sancionará con multas administrativas de hasta 15 000 000 EUR o, si el infractor es una empresa, de hasta el 3 % de su volumen de negocios total anual a escala mundial correspondiente al ejercicio anterior, si esta cifra es superior: (a) Obligaciones de los prestadores con arreglo al artículo 16; (b) obligaciones de los representantes autorizados con arreglo al artículo 22; (c) las obligaciones de los importadores con arreglo al artículo 23; (d) obligaciones de los distribuidores con arreglo al artículo 24; (e) las obligaciones de los responsables del despliegue en virtud del artículo 26; (f) los requisitos y obligaciones de los organismos notificados con arreglo al artículo 31, a los apartados 1, 3 y 4 del artículo 33 o al artículo 34; (g) obligaciones de transparencia para los proveedores y los implantadores de conformidad con el artículo 50. 5. El suministro de información incorrecta, incompleta o engañosa a los organismos notificados o a las autoridades nacionales competentes en respuesta a una solicitud se sancionará con multas administrativas de hasta 7 500 000 EUR o, si el infractor es una empresa, de hasta el 1 % de su volumen de negocios total anual a escala mundial correspondiente al ejercicio anterior, si esta cifra es superior. 6. En el caso de las PYME, incluidas las de nueva creación, cada multa a que se refiere el presente artículo será de hasta los porcentajes o el importe a que se refieren los apartados 3, 4 y 5, si éste fuera inferior. 7. Al decidir sobre la imposición de una multa administrativa y sobre su cuantía en cada caso concreto, se tendrán en cuenta todas las circunstancias pertinentes de la situación específica y, en su caso, se tendrá en cuenta lo siguiente: (a) La naturaleza, gravedad y duración de la infracción y de sus consecuencias, teniendo en cuenta la finalidad del sistema de IA, así como, en su caso, el número de personas afectadas y el nivel de perjuicio sufrido por éstas; (b) si otras autoridades de vigilancia del mercado ya han aplicado multas administrativas al mismo operador por la misma infracción; (c) si otras autoridades ya han aplicado multas administrativas al mismo operador por infracciones de otro Derecho de la Unión o nacional, cuando tales infracciones se deriven de la misma actividad u omisión que constituya una infracción pertinente del presente Reglamento; (d) el tamaño, el volumen de negocios anual y la cuota de mercado del operador que comete la infracción; (e) cualquier otro factor agravante o atenuante aplicable a las circunstancias del caso, como los beneficios financieros obtenidos o las pérdidas evitadas, directa o indirectamente, por la infracción; (f) el grado de cooperación con las autoridades nacionales competentes, con el fin de remediar la infracción y mitigar los posibles efectos adversos de la misma; (g) el grado de responsabilidad del operador, teniendo en cuenta las medidas técnicas y organizativas que haya aplicado; (h) la forma en que la infracción llegó a conocimiento de las autoridades nacionales competentes, en particular si el operador notificó la infracción y, en caso afirmativo, en qué medida; (i) el carácter intencionado o negligente de la infracción; (j) cualquier medida adoptada por el operador para mitigar el daño sufrido por las personas afectadas. 8. Cada Estado miembro establecerá normas sobre la medida en que podrán imponerse multas administrativas a las autoridades y organismos públicos establecidos en dicho Estado miembro [...]". Sobre este tema véase también C. Minelli, *La responsabilità "penale" tra persona fisica e* corporation alla *luce della Proposta di Regolamento sull'Intelligenza Artificiale,* en *Dir. pen. cont., 2022*, 2, 50 ss.; A. Giannini, *Intelligenza artificiale,* human oversight *e responsabilità penale: prove d'impatto a livello europeo*, en *Criminalia*, 2021, 249 ss.

debería, como mínimo, identificar qué productos están permitidos, establecer las reglas de *compliance* que deben observarse para introducir estos dispositivos en el mercado de forma segura e imponer formas de seguimiento posterior a la comercialización, indicando también la autoridad pública encargada de los controles.

El segundo paso que hay que dar consiste en reflexionar detenidamente sobre el papel que el recurso penal puede desempeñar en esta materia, en particular, en lo que se refiere a las acciones de *enforcement* dirigidas contra las empresas en relación con los delitos vinculados a la utilización de la IA.

Teóricamente, podrían considerarse y desarrollarse tres estrategias sancionadoras.

1) La primera estrategia consiste en la previsión de *sanciones penales/punitivas aplicables al sistema de IA como tal*, es decir, como autor directo de un delito, a condición, por supuesto, del reconocimiento de su personalidad jurídica. Desde este enfoque, también debería considerarse la posibilidad de activar las normas ordinarias sobre la responsabilidad penal o *ex crimine* de las personas jurídicas, vigentes en las distintas jurisdicciones, en caso de que las infracciones penales relacionadas con la IA se cometan en el contexto de una organización empresarial.

2) El segundo modelo de intervención podría contemplar la adopción de *sanciones penales/punitivas directas contra la empresa que haya producido o utilizado un sistema de IA*, causando daños a terceros y/o realizando los elementos constitutivos de un delito. Según este mecanismo de imputación, la responsabilidad autónoma y directa de la entidad empresarial podría basarse en la falta de adopción de medidas de cumplimiento normativo de carácter preventivo, con anterioridad a la producción del resultado dañoso derivado del mal funcionamiento de la IA. Se trataría, por tanto, de un supuesto de responsabilidad empresarial basado en la culpa o el defecto de organización de la empresa.

3) La tercera solución consiste en la imposición de *sanciones punitivas a las empresas*, independientemente de la constatación de daños y/o delitos, en consideración a la mera *violación de los requisitos y obligaciones de cumplimiento normativo* que deben respetarse en la introducción de herramientas de IA en el mercado y la supervisión posterior a su distribución y venta. Como ya se ha mencionado, esta estrategia presupone la previa adopción de una regulación pública que aborde de forma integral el fenómeno en cuestión.

Dicho esto, los tres modelos no parecen igualmente plausibles.

En cuanto a la *primera hipótesis* normativa (responsabilidad directa

de los algoritmos), ya hemos expresado las razones por las que la idea de que *machina delinquere potest* no es convincente ni practicable. Como se ha explicado, el estado actual de los conocimientos científicos hace imposible captar en el funcionamiento de las herramientas de IA una *free will* o autoconciencia y capacidad reales de ser motivadas por la amenaza de un mal futuro[56]. Legitimar una responsabilidad de este tenor sería una *fictio* insensata, sin perjuicio de la posibilidad de "atacar" materialmente al instrumento de IA en función de su objetiva peligrosidad comprobada.

La *segunda solución* es, en principio, viable. No obstante, también requiere algunas aclaraciones y advertencias. Introducir una disposición por la que se castigue la no prevención del delito-resultado derivado del mal funcionamiento del sistema de IA, en ausencia de una regulación pública que establezca, de manera precisa, los estándares aplicables a la producción y venta de tales instrumentos tecnológicos, equivaldría a confiar – de manera irracional y, por tanto, injusta – a las empresas la tarea de gestionar todos los riesgos (y las cuestiones de responsabilidad correlativas) asociados con la producción y utilización de estos dispositivos.

Por lo tanto, como ya hemos señalado anteriormente, de acuerdo con un principio general de política jurídica racional, la introducción de sanciones – formalmente o sustancialmente – penales contra las empresas debería constituir el último recurso en el contexto de una estrategia más amplia y articulada de regulación pública.

Además, al menos en determinados escenarios que no son en absoluto remotos, es cuestionable que pueda legitimarse – en términos dogmáticos y político-criminales – una disposición que introduzca la responsabilidad penal de las empresas por la causación material de resultados negativos imprevisibles, generados por sistemas capaces de tomar decisiones de forma completamente autónoma[57]. En otras

[56] Sobre este punto, véase, en la literatura italiana: F. Basile, *Intelligenze artificiali e diritto penale: quattro possibili percorsi di indagine*, en *Diritto penale e uomo*, 2019, 2 ss.; Cappellini, Machina delinquere non potest*? Brevi appunti su intelligenza artificiale e responsabilità penale* en *Discrimen*, 2019; R. Borsari, *Intelligenza artificiale e responsabilità penale: prime considerazioni*, en *Medialaws*, 2019, 3, 262 ss.; D. Piva, *Machina discere, (deinde) delinquere et punire potest*, en *Il diritto nell'era digitale*, cit, 681 y ss.; P. Severino, *Intelligenza artificiale e diritto penale,* en *Intelligenza artificiale: il diritto, i diritti, l'etica*, editado por U. Ruffolo, Milano, 2020, 533 y ss.; A.F. Tripodi, *Uomo,* societas*,* machina, en *Leg. pen.*, 2023, 12. En la literatura internacional véanse, además de los autores ya mencionados en las notas anteriores, F. Lagioia y G. Sartor, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective, en* 33 *Philosophy & Technology*, 2020, 433 y ss.; D. Lima, *Could AI Agents Be Hold Criminally Liable? Artificial Intelligence and The Challenges for Criminal Law*, en 69 *South Carolina Law Review*, 2018, 677 y ss.

[57] Véase también B. Panattoni, *AI and Criminal Law,* cit., 125 y ss. Para una visión general, véase E. Gruodytė y P. Čerka, *Artificial Intelligence as a Subject of Criminal Law: A Corporate Liabil-*

palabras, cabe dudar mucho de la posibilidad de probar en juicio que una organización es culpable si ni siquiera con toda la diligencia abstractamente concebible habría podido prever y, por lo tanto, evitar el daño causado por el "fallo" de un sistema de IA cuyo funcionamiento resulte – total o al menos en parte – inescrutable[58]. Es evidente que, en este caso, la única posibilidad real de evitar el acontecimiento imprevisible sería renunciar por completo a la producción y/o uso del sistema de IA, acogiéndose a una regla de abstención basada en el principio de precaución.

Teniendo en cuenta el conjunto de estos aspectos, la *tercera solución* podría consistir, al menos en las fases iniciales del proceso de regulación normativa que la materia reclama, en un compromiso equilibrado para evitar, por un lado, injustas imputaciones de responsabilidad y, por otro, una sustancial paralización de la investigación y la producción en el campo de la IA, con un indeseable *chilling effect.* Como ya hemos subrayado, para desarrollar mejor esta política, el primer paso debe ser la introducción de una legislación que, en línea con el nuevo Reglamento europeo antes mencionado, establezca las "reglas del juego" y todos los requisitos de conformidad asociados a la comercialización y supervisión de los instrumentos de IA. Esta opción también parece más respetuosa con el principio de *ultima ratio*, según el cual el Derecho penal debería activarse sólo tras haber verificado la capacidad de otras formas de responsabilidad, empezando por la responsabilidad civil y administrativa – para garantizar una protección adecuada de los intereses en juego, teniendo en cuenta asimismo la novedad del fenómeno que debe gestionarse y los riesgos a enfrentar. A este respecto, cabe mencionar también la Resolución del Parlamento de la UE de 16 de febrero de 2017, que contiene recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica, la cual, entre otras cosas, había sugerido el establecimiento de un "régimen de seguro obligatorio" para los posibles daños producidos en tales contextos, así como un "fondo para garantizar la compensación de los daños y perjuicios en los supuestos en los que no exista una cobertura de seguro"[59].

---

ity Model Perspective, en *Smart Technologies and Fundamental Rights,* editado por J.S. Gordon, Leiden, 2020, 260 y ss. Sobre algunas de estas perspectivas futuras, véase también *Feasibility study on a future Council of Europe instrument on artificial intelligence and criminal law*, publicado por el European Committee on Crime Problems (CDPC) del Consejo de Europa en 2020, disponible en *{www.coe.int/cdpc}.*

[58] Sobre la evaluación del defecto de organización de las empresas en el contexto de la nueva era digital, véase también A. Nisco, *Riflessi della* compliance *digitale in ambito 231*, en *{www.sistemapenale.it}*, 14 de marzo de 2022.

[59] El texto de la Resolución está disponible en el siguiente enlace: *{https://eur-lex.europa.eu/legal-*

Por lo tanto, sólo después de haber elaborado un corpus normativo exhaustivo, capaz de ordenar de manera adecuada el fenómeno examinado, sería razonable prever sanciones dirigidas a la entidad colectiva, que desplacen el *focus* de las actividades de *enforcement* relativas a las personas jurídicas desde la causación de daños a terceros a la violación de los requisitos de *compliance* relativos a la introducción segura en el mercado y la supervisión postventa de las herramientas de IA.

De este modo, tales medidas sancionatorias permitirían garantizar la correcta aplicación de la normativa pública por parte de los operadores económicos y se orientarían a proteger el respeto de tales (esenciales) disposiciones. En aplicación del principio de precaución, la regulación pública del sector también podría establecer prohibiciones absolutas de producir y comercializar técnicas específicas que no ofrezcan unas garantías mínimas de fiabilidad y que se consideren demasiado arriesgadas por los daños que podrían causar.

En este contexto, sería deseable una estrategia sancionadora articulada, compuesta de medidas punitivas clásicas y medidas "programáticas" dirigidas a la adopción de específicas acciones preventivas o correctivas por parte de la empresa.

Desde el primer punto de vista, deberían introducirse sanciones pecuniarias por infringir las normas de producción/comercialización/uso de productos de IA o por incumplir el deber de informar a las autoridades en caso de incidentes nocivos o peligrosos. Desde el segundo punto de vista, debería otorgarse al juez el poder de imponer la adopción de medidas para garantizar el cumplimiento adecuado de sus deberes por parte de las empresas y, en particular, la implantación o mejora de los mecanismos de cumplimiento empresarial y los sistemas de control interno[60], a los que podría añadirse un período de supervisión pública (administrativa o judicial, en todo caso por parte de una institución pública), para comprobar escrupulosamente que cumplen con los estándares impuestas por la normativa del sector[61].

Por otra parte, al menos en una primera fase de "experimentación" de la nueva legislación, es deseable un *enforcement* puramente san-

---

content/ES/TXT/PDF/?uri=CELEX:52017IP0051&from=EN}. Sobre este tema, véase también la Propuesta de directiva del Parlamento europeo y del Consejo relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial (Directiva sobre responsabilidad en materia de IA), COM/2022/496 final.

[60] En el mismo sentido, véase F: Consulich, Flash offenders. *Le prospettive di* accountability *penale nel contrasto alle intelligenze artificiali devianti*, en *Riv. it. dir. proc. pen.*, 2022, 3, 1051 ss.

[61] Sobre las posibles reformas del sistema de sanciones contra entidades colectivas en el ordenamiento jurídico italiano, véase, también para una revisión más amplia de la literatura, V. Mongillo, *Il sistema delle sanzioni applicabili all'ente collettivo tra prevenzione e riparazione. Prospettive* de iure condendo, en *Riv. trim. dir. pen. ec.*, 2022, 3-4, 559 ss.

cionatorio administrativo en casos de incumplimiento de la regulación específica en materia de IA. Esto también permitiría, con el tiempo, comprender si un modelo de regulación basado en disposiciones no penales es suficiente para contrarrestar los daños que potencialmente pueden generar los dispositivos de IA.

IV. Conclusiones

En esta contribución, hemos intentado explorar las oportunidades y los retos que debe enfrentar la posible responsabilización de las empresas en relación con los delitos vinculados al uso de sistemas de IA.

En referencia a los dispositivos completamente autónomos, se ha destacado cómo actualmente existe un vacío de responsabilidad, dados los modelos de *corporate criminal liability* más difundidos a nivel internacional.

También se ha señalado cómo la previsión de sanciones penales dirigidas directamente contra los sistemas digitales de IA y el mismo castigo "directo" de las entidades colectivas que los utilizan no encuentra una base adecuada en el estado actual de los conocimientos científicos.

En este contexto, las únicas alternativas plausibles parecen reducirse a la introducción de sanciones punitivas contra:

(a) la organización que negligentemente no implemente medidas destinadas a contener el riesgo de producción de resultados lesivos debidos a la IA, en los casos y en la medida en que los conocimientos técnico-científicos hoy disponibles permitan identificar estrategias y acciones para prevenir un riesgo previsible: en este caso, la persona jurídica respondería por la no prevención de los daños producidos como consecuencia del uso del dispositivo de IA; o

(b) la *corporation* que omita adoptar las medidas de cumplimientonormativo o los estándaresestablecidos para producir y comercializar legalmente dichos productos digitales, con independencia de la producción de un delito o resultado lesivo o peligroso.

Así pues, hemos tratado de señalar los puntos fuertes y débiles de ambas opciones, cuya exposición revela lo difícil que resulta regular un universo tecnológico que está creciendo a una velocidad inimaginable hace tan sólo unos años, así como las complejas implicaciones que deben tenerse en cuenta a la hora de adoptar necesarias medidas de política legislativa[62].

[62] Véase también C. Burchard, *L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmica della società*, en *Riv. it. dir. proc. pen.*, 2019, 4, 1909; W.S. Laufer, *The Missing Account of Progressive Corporate Criminal Law*, en 14 *New York University Journal of Law & Business*, 2017, 71.

Frente a estos desafíos, las estrategias para contener eficazmente los riesgos inherentes al uso de las nuevas formas de IA deben tener como objetivo el desarrollo de tecnologías al servicio de la dignidad humana, los derechos fundamentales y el progreso social, sin sacrificar – en el ámbito disciplinar en el que nos hemos centrado – los principios fundamentales del Derecho penal, como logros irrenunciables de la civilización jurídica.