REVISTA
DE DERECHO, EMPRESA Y SOCIEDAD
(REDS)

Número 24, Época II, 2024

ISSN: 2340-4647

IURE LICET ABOGADOS

Dykinson, S.L.

REVISTA
DE DERECHO, EMPRESA Y SOCIEDAD (REDS)

Número 24, Época II, 2024

ISSN: 2340-4647

Dykinson, S.L.

ADQUISICIÓN Y SUSCRIPCIONES

Dykinson, S.L.

Suscripción versión electrónica (Revista en PDF)

Compra directa a través de nuestra web:

www.dykinson.com/derechoempresaysociedad

COMPOSICIÓN DEL COMITÉ CIENTÍFICO INTERNACIONAL

Miembros del Comité:

Presidente

Dr. Dº BORJA MAPELLI CAFFARENA

Catedrático emérito de Derecho penal y Ciencias Criminales. Universidad de Sevilla

Vocales:

PIERRE LUIGI MARIA DELL´OSSO
Fiscal Antimafia de la República de Italia.
(Procurador Nacional Antimafia de Italia)

CAMILO CELA CONDE
Director del Laboratorio de Sistemática Humana
Universidad de las Islas Baleares

ANTONIO FLAMINI
Doctor y ex Decano de la Facultad de Derecho de la Universidad de Camerino, Catedrático de Derecho Civil y miembro de la "Escuela Civilística" que agrupa a los más prestigiosos catedráticos de derecho civil italiano.
Universidad de Camerino (Italia)

LUCIA RUGGERI
Profesora de Derecho civil y Decana de la Facultad de Derecho de la Universidad de Camerino

ANGEL REBOLLLEDO VARELA
Catedrático de Derecho Civil
Universidad de Santiago de Compostela

JEAN-BERNARD AUBY
Ex decano de la Facultad de Derecho de la Universidad de París XII. Profesor de Derecho Público en la Universidad de Sciences Po Paris y director de la Acción mutaciones de l'Publique Pública Droit et du (cambios en el Gobierno y Derecho Público, MADP) de Sciences Po Paris.
Universidad de Sciences Po Paris

LORENZO MEZZASOMA
Juez Honorario en el Tribunal de Perugia.
Catedrático Derecho Civil y coordinador de actividades de investigación de Derecho civil de la Universidad de Perugia.
Universidad de Perugia

JOSE PABLO ALZINA DE AGUILAR
Cónsul General de España en Brasil

MIGUEL OLMEDO CARDENETE
Catedrático Derecho Penal
Universidad de Granada

IGNACIO BENÍTEZ ORTÚZAR
Catedrático Derecho Penal
Universidad de Jaén

SECCIONES PERMANENTES EN LA REVISTA: Derecho, Empresa y Sociedad

Coordinadora de Derecho Privado, Bioderecho, IA y Transformación digital

Dra. Dª MARÍA LUDOMIRA KUBICA

Profesora Ayudante Doctora de Derecho Civil

Universidad Complutense de Madrid

Coordinadora de Nuevas formas de criminalidad y lucha contra la corrupción

Dra Dª DEMELSA BENITO SÁNCHEZ

Profesora de Derecho penal

Universidad de Deusto

Coordinador de Economía, Empresa, Estudios Financieros y Negocios

Dr. Dº JONATHAN TÉLLEZ TORRES

Profesor Ayudante Doctor de Derecho penal y Ciencias Criminales

Universidad de Sevilla

ÍNDICE

RESPONSABILIDAD DE LA INTELIGENCIA ARTIFICIAL Y LA NECESIDAD DE INTEGRAR LA RESPONSABILIDAD CORPORATIVA

Carlo Piparo
Universidad de Sevilla/ Universidad de Udine
Edgar Iván Colina Ramírez
Profesor Contratado. Doctor. Universidad de Sevilla

Fecha de recepción: 06/09/2024
Fecha de aceptación: 11/10/2024

RESUMEN: La rápida progresión y la integración generalizada de las tecnologías de la información y las comunicaciones (TIC) han marcado el comienzo de una nueva era de transformaciones sociales y jurídicas radicales. Entre los muchos avances revolucionarios, la Inteligencia Artificial (IA) se ha convertido en una fuerza fundamental que impregna casi todas las facetas de nuestra vida cotidiana. Desde los ámbitos del comercio y la industria hasta la atención médica, el transporte y el entretenimiento, las tecnologías de IA se han convertido en herramientas indispensables que dan forma a la forma en que interactuamos, trabajamos y navegamos por el mundo que nos rodea. Con sus notables capacidades y su alcance en constante expansión, la IA es un testimonio de la incesante búsqueda de la innovación por parte de la humanidad y del potencial ilimitado de la tecnología para revolucionar la sociedad.

Mientras completan todas las tareas para las que están programados, los sistemas de Inteligencia Artificial pueden realizar acciones, que podrían resultar en delitos si los cometen los humanos. Pero los delitos siguen la reserva de la ley, por lo tanto, puede ser difícil penalizar tales delitos debido a la falta de una ley escrita.

En los sistemas jurídicos modernos, la estructura de los delitos no solo requiere la comisión de un hecho típico, sino también la determinación de hacerlo. En este escenario, al ser la IA una entidad no humana, la reconstrucción de la responsabilidad penal es particularmente difícil de teorizar. Sin embargo, hemos observado que la imposición de responsabilidad penal por comportamientos dañinos y erráticos de los robots a menudo nos lleva a un círculo vicioso debido a la dificultad de atribuir responsabilidad a individuos individuales.

Este artículo pretende evaluar la naturaleza de la IA y sus relaciones con el derecho penal, deconstruir tres posibles modelos de responsabilidad de la IA, para evaluar si es necesario o no aplicar la responsabilidad de la empresa.

ABSTRACT: The rapid progression and widespread integration of Information and Communication Technology (ICT) have ushered in a new era of sweeping social and legal transformations. Among the many groundbreaking advancements, Artificial Intelligence (AI) has emerged as a pivotal force, permeating nearly every facet of our daily lives. From the realms of commerce and industry to healthcare, transportation, and entertainment, AI technologies have become indispensable tools shaping the way we interact, work, and navigate the world around us. With its remarkable capabilities and ever-expanding reach, AI stands as a testament to humanity's relentless pursuit of innovation and the boundless potential of technology to revolutionize society.

While completing all the tasks they are programmed for, Artificial Intelligence systems can perform actions, which could result in crimes if committed by humans. But crimes follow the reserve of law, therefore can be difficult to criminalize such crimes because of the lack of written law.

In modern legal systems, the structure of crimes doesn't only require the commission of a typical fact, but also the determination to do it. In this scenario, being AI a non-human entity, the reconstruction of criminal responsibility is particularly difficult to theorize. Nevertheless, we have observed that imposing criminal liability for harmful and erratic behaviors of robots often leads us in a vicious circle due to the difficulty in attributing responsibility to individual individuals.

This paper wants to assess the nature of AI and its relationships with criminal law, to deconstruct three possible AI liability models, to assess wether or not it is necessary to apply company responsibility.

PALABRAS CLAVE: Inteligencia Artificial, Derecho Penal, Responsabilidad Empresarial.

KEYWORDS: Criminal Law, Artificial Intelligence, Company Responsibility.

1. INTRODUCCIÓN.

Hoy en día, la Inteligencia Artificial (IA) es una fuerza dominante y omnipresente, impulsada por técnicas rápidamente adaptables, como algoritmos de aprendizaje automático, *data mining* y sistemas predictivos[1]. Estas técnicas son testigos de un nivel extraordinario, y tal vez desconcertante, de integración de la Inteligencia Artificial en nuestras vidas y sociedades[2]. Actualmente, estas técnicas encuentran aplicaciones en la mayoría de los sectores: desde navegadores de Internet y aplicaciones para teléfonos inteligentes hasta videojuegos, desde proyectos de ingeniería y gráficos animados hasta hospitales e investigación[3]. Eminencias como Stephen Hawking predijeron que dentro de un siglo la

[1] BODEN, M.A. *Intelligenza artificiale*, in J. l-Khalili (editor), *Il futuro che verrà*, Bollati Boringhieri, 2018, p. 133.

[2] Vid. Al respecto PIPARO, C., *Machina delinquere potest? A modern criminalization challenge due to lack of text,* in *Text, context, and subtext in law,* Timisoara, 2023, p. 900.

[3] ITALIANO G.F., *Intelligenza artificiale: passato, presente, futuro*, in F. Pizzetto (editor), *Intelligenza artificiale, protezione dei dati personali e regolazione*, 2018, p. 216.; KAPLAN, J. *Intelligenza artificiale. Guida al futuro prossimo*, Luiss University Press, II ed., 2018, pp. 81 ss., y pp. 193 ss. y FLORISI, L.*What the Near Future of Artificial Intelligence Could Be*, in *Philosophy & Technology,* n. 32, 2019, pp. 3 ss. (*online* at https://doi.org/10.1007/s13347-019-00345-yp).

inteligencia computacional superará las capacidades humanas[4]. La omnipresencia de la inteligencia artificial requiere una respuesta legal[5].

El derecho penal debe estar preparado para la que ya se ha revelado como una verdadera revolución tecnológica y enfrentarse a los desafíos que esta transformación pueda plantear[6]. Abordar esto requiere evaluar la adaptabilidad de las normas existentes para acomodar tecnologías novedosas, deliberar sobre la conveniencia de formular nuevas regulaciones adaptadas, o, alternativamente, perseverar con las leyes existentes, aunque con posibles tensiones, posiblemente respaldadas por el derecho precedente[7].

Este esfuerzo debe tender a armonizar estos nuevos resultados legales con los derechos fundamentales como el debido proceso, la privacidad y la igualdad[8].

2. EL CONCEPTO DE INTELIGENCIA ARTIFICIAL

John McCarthy, un científico informático estadounidense, utilizó por primera vez el término Inteligencia Artificial en 1955. Aproximadamente tres décadas después, en un ensayo de 1987, Roger Schank, un destacado teórico de la Inteligencia Artificial y una figura fundamental en la lingüística computacional, atribuyó cinco atributos a la IA: la capacidad de comunicación, autoconciencia, comprensión de la realidad externa, acción intencionada guiada por objetivos y un notable grado de creatividad, que abarca la capacidad de tomar decisiones alternativas cuando el curso de acción inicial resulta infructuoso o inviable[9].

Este conjunto de características ofrece dos ideas clave. En primer lugar, la Inteligencia Artificial no puede limitarse al ámbito de humanoides inteligentes o cíborgs: estos últimos pueden manifestarse como una aplicación de Inteligencia Artificial. En segundo lugar, los sistemas de Inteligencia Artificial no pueden replicar los mecanismos cognitivos de la mente humana. En consecuencia, es más apropiado considerar la Inteligencia Artificial como una disciplina computacional en lugar de una emulación del intrincado sistema biológico

[4] WALKER, L., *Stephen Hawking warns artificial intelligence could end humanity*, *Newsweek*, 14 May 2015, where the Author quotes the Speaking of S. Hawking during Zeitgeist Conference, London, May 2015.

[5] Como ya se ha destacado en PIPARO, C. *Ibidem: "The European Parliament Resolution of 16 February 2017, providing recommendations to the European Commission on civil law rules on robotics (2015/2103(INL)), es un documento que ofrece orientación y sugerencias a la Comisión Europea sobre la necesidad de elaborar normas de derecho civil específicas para el ámbito de la robótica. El documento representa un paso importante para abordar las consecuencias jurídicas y sociales asociadas con el avance de la tecnología robótica. En la resolución se destaca la importancia de crear un marco jurídico claro y coherente que aborde las cuestiones relacionadas con la responsabilidad y la seguridad en el ámbito de la robótica. Reconoce que la creciente presencia de robots e inteligencia artificial plantea una serie de retos, entre ellos determinar la responsabilidad en caso de daños causados por un robot, proteger los datos personales y garantizar la seguridad de los propios robots. En esta resolución, el Parlamento Europeo pide a la Comisión Europea que considere la adopción de un marco jurídico específico para la robótica que tenga en cuenta los principios éticos y los derechos fundamentales de las personas. También hace hincapié en la necesidad de promover la investigación e innovación en el ámbito de la robótica para garantizar que Europa siga siendo competitiva en este sector que evoluciona rápidamente. En resumen, la Resolución del Parlamento Europeo de 16 de febrero de 2017 es un documento importante que plantea la cuestión de las normas de derecho civil sobre robótica e insta a la Comisión Europea a que considere este desafío y adopte medidas apropiadas para abordarlo".*

[6] PIPARO, C.*Machina delinquere potest?, op. cit.,* p. 901.

[7] STELLA, F., *Giustizia e modernità. La protezione dell'innocente e la tutela delle vittime*, Giuffré, 2003, pp. 292 ss.

[8] BASSINI, M., LIGUORI, M., POLLICINO, L. *Sistemi di Intelligenza Artificiale, responsabilità e accountability. Verso nuovi paradigmi?*, en PIZZETTI, F. (ed.), *Intelligenza artificiale, protezione dei dati personali e regolazione*, 2018, p. 334.

[9] SCHANK, R. C. *What's IA, Anyway?*, in *IA Magazine*, Winter 8(4), 1987, pp. 59 ss.

humano[10]. Los expertos en Inteligencia Artificial tienden a preferir el término "racionalidad" sobre "inteligencia[11]", denotando la capacidad de seleccionar el recorrido óptimo para lograr objetivos específicos, guiados por criterios de optimización de recursos[12].

Hoy en día, el uso de la expresión Inteligencia Artificial[13] asume diversas connotaciones e interpretaciones dependiendo de la disciplina específica o el contexto de referencia, dejando así sin definiciones universales[14]. *«Otras definiciones van más allá en la explicación [...] de habilidades y tareas. Por ejemplo, el científico informático Nils John Nilsson describe una tecnología que "funciona apropiadamente y con previsión en su entorno". Otros hablan de la capacidad de percibir, perseguir objetivos, iniciar acciones y aprender de un bucle de retroalimentación. Una definición similar ha sido propuesta por el Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial (AI HLEG) de la Comisión Europea (CE): "Sistemas que muestran comportamiento inteligente al analizar su entorno y tomar acciones -con cierto grado de autonomía- para lograr objetivos específicos". Estas definiciones basadas en tareas ayudan en cierta medida a comprender mejor qué es la IA. Pero aún tienen limitaciones. Conceptos como "cierto grado de autonomía" siguen siendo algo vagos. Además, estas definiciones aún parecen demasiado amplias en el sentido de que describen fenómenos que la mayoría de nosotros no estaríamos inclinados a agrupar bajo el término IA. Por ejemplo, la definición de Nilsson también se aplica a un termostato clásico. Este dispositivo también es capaz de percibir (medir la temperatura de la habitación), perseguir objetivos (la temperatura programada), iniciar acciones (regular el termostato) y aprender de un bucle de retroalimentación (detenerse una vez que se ha alcanzado la temperatura programada). Aun así, la mayoría de las personas no estarían inclinadas a considerar un termostato como IA[15]».*

[10] KAPLAN J., *Intelligenza artificiale*, cit., p. 41.

[11] En RUSSEL S., NORVIG, P. *Artificial Intelligence: A Modern Approach, Prentice Hall,* 3rd edition, 2009, pp. 36 ss. Los autores discuten sobre el concepto de "agente racional", definiéndolo como el agente que constantemente hace elecciones que maximizan su rendimiento esperado, dada su secuencia de percepción (entrada sensorial histórica), la medida de rendimiento (un criterio para el éxito), y cualquier conocimiento incorporado sobre su entorno. En esencia, la racionalidad significa seleccionar las acciones que son susceptibles de conducir a los mejores resultados globales, teniendo en cuenta la información disponible para el agente hasta ese punto. Según los autores, la racionalidad no requiere omnisciencia; en cambio, opera basándose en la percepción del agente del mundo y sus experiencias pasadas. Incluye la capacidad de recopilar información, explorar entornos desconocidos y adaptar el comportamiento basado en el aprendizaje de estas experiencias. Los agentes racionales son autónomos en el sentido de que pueden tomar decisiones independientes del conocimiento previo completo, confiando en su percepción y el aprendizaje pasado para hacer decisiones informadas.En resumen, la racionalidad, tal como se define en el documento, representa un principio fundamental en la inteligencia artificial, guiando a los agentes a tomar decisiones que optimicen su rendimiento basándose en la información a su disposición y su comprensión del medio ambiente.

[12] RUSSEL S., NORVIG, P., *Ibidem.*

[13] El conceptode Inteligencia Artificial fue inventado por un científico de la computación estadounidense John McCarthy in 1955. Vid. Ampliamente a PIPARO, C., *Machina delinquere potest?, cit.*, donde el Autor afirma que «El término es usado públicamente por el estudioso durante un seminario celebrado en Dartmouth College. El estudioso continuó sus estudios en el campo de la inteligencia artificial, lo que le llevó a ganar el Premio Turing en 1971 por sus importantes contribuciones en esta area».

[14] Vid. Ampliamente GUTIERREZ C. I., AGUIRRE A., UUK, R. C. BOINE, C.M. Franklin, *A Proposal for a Definition of General Purpose Artificial Intelligence Systems*, *2022*. donde el Autor afirma que: «En la actualidad, no hay directrices que expliquen los criterios de inclusión de los sistemas de IA que se clasifican como Sistemas de Inteligencia Artificial de Propósito General (GPAIS). En el contexto de la Ley de inteligencia artificial, la definición existente tiene muchas oportunidades de mejora. La Presidencia eslovena de la UE definió GPAIS como un "sistema de IA... capaz de realizar funciones generalmente aplicables como el reconocimiento de imagen/habla, la generación de audio/vídeo, la detección de patrones, la respuesta a preguntas, la traducción, etc".] La Presidencia francesa de la UE subraya además que el GPAIS: "puede ser utilizado en una pluralidad de contextos e integrado en una multitud de otros sistemas de IA". Fuera del contexto de la UE, el término GPAIS se utiliza raramente y aleatoriamente para describir sistemas de IA que varían considerablemente en términos de autonomía, agencia, modalidad y métodos de capacitación».

[15] SHEIKH H., PRINS, C. SCHRIJVERS, E.*Artificial Intelligence: Definition and Background, In: Mission AI. Research for Policy. Springer,* 2023

La Carta Ética Europea, adoptada por la Comisión Europea para la Eficacia de la Justicia en 2018, enmarca la Inteligencia Artificial como «*el conjunto de métodos científicos, teorías y técnicas destinadas a reproducir mediante máquinas las habilidades cognitivas de los seres humanos*», enfatizando el objetivo actual de delegar tareas complejas hasta ahora realizadas por humanos a máquinas.

En contraste, la Comunicación de la Comisión Europea sobre Inteligencia Artificial en Europa de 2018 caracteriza la Inteligencia Artificial como *«sistemas que exhiben comportamiento inteligente al analizar su entorno y tomar acciones de forma autónoma para lograr objetivos específicos»*. Esta definición abarca sistemas de Inteligencia Artificial tanto en el ámbito virtual, como asistentes de voz, software de análisis de imágenes y motores de búsqueda, como en aquellos integrados en hardware físico, incluidos robots avanzados, vehículos autónomos, drones y aplicaciones de Internet de las cosas[16].

Un examen escrupuloso realizado por el Grupo de Expertos de Alto Nivel Independiente, designado por la Comisión Europea con fines de asesoramiento sobre Inteligencia Artificial, se alinea con las definiciones anteriores[17]. Según este grupo, el concepto de Inteligencia Artificial denota *«software (y potencialmente hardware) diseñado por humanos que, cuando se le presenta un objetivo complejo, opera en el ámbito físico o digital al percibir su entorno mediante la adquisición de datos, interpretar datos estructurados o no estructurados, razonar basándose en el conocimiento o información derivada de estos datos, y determinar el curso de acción óptimo para alcanzar el objetivo especificado».* Los sistemas de IA pueden adoptar reglas simbólicas o adquirir un modelo numérico, y también son capaces de adaptar su comportamiento mediante el análisis de las consecuencias de sus acciones anteriores sobre el entorno. Como disciplina científica, la IA abarca una variedad de enfoques y técnicas, que incluyen aprendizaje automático, razonamiento mecánico y robótica, integrando varios métodos dentro de sistemas ciber-físicos. En consecuencia, aunque la comunidad científica emplea diversas definiciones de Inteligencia Artificial, ciertas características comunes emergen. En resumen, la Inteligencia Artificial generalmente alude a un compendio de metodologías, teorías y técnicas científicas con el objetivo de replicar las habilidades cognitivas humanas a través de medios mecanizados[18].

---

[16] Dentro de Europa, el acrónimo CEPEJ significa la Comisión Europea para el Mejoramiento de la Eficiencia de la Justicia. Surgiendo como una entidad operativa bajo el paraguas del Consejo de Europa, CEPEJ se dedica a mejorar continuamente los sistemas de justicia dentro de los estados miembros que abarca. Su inicio fue impulsado por la misión de fomentar un mayor acceso a la justicia, elevar el estándar de los servicios judiciales y mantener los principios de equidad en los procedimientos legales. Un aspecto fundamental del rol de CEPEJ gira en torno al desarrollo e implementación de herramientas estandarizadas, metodologías y criterios de referencia. Estos elementos sirven como catalizadores para la transformación progresiva de los sistemas de justicia en toda Europa. CEPEJ ofrece su experiencia, proporcionando orientación valiosa a los estados miembros, participando en esfuerzos de investigación exhaustivos y recopilando datos invaluables para evaluar la funcionalidad de los sistemas judiciales. A través de estos esfuerzos multifacéticos, CEPEJ tiene como objetivo discernir y promover las mejores prácticas, estimular esfuerzos colaborativos y facilitar diálogos constructivos entre profesionales judiciales, formuladores de políticas y partes interesadas pertinentes. El mandato de CEPEJ abarca diversas dimensiones de la eficiencia judicial. Incluye áreas como la gestión de casos, la administración judicial, el cumplimiento de los plazos legales, la calidad de los pronunciamientos legales y la integración de tecnologías de información de vanguardia en el ámbito de la justicia. Además, CEPEJ aborda temas relacionados con el acceso a la justicia, programas de capacitación judicial y la evaluación crítica de los sistemas judiciales existentes.
Al abogar por principios fundamentales, como la eficiencia, el acceso y la imparcialidad en la administración de justicia, CEPEJ desempeña un papel fundamental en el fomento de la efectividad general de los marcos legales dentro de Europa, fortaleciendo consecuentemente los cimientos del Estado de Derecho.

[17] ALGERI,L., *Intelligenza artificiale e polizia predittiva*, in Dir. Pen. e Processo, vol. 6, 2021, p. 724.

[18] KOF, J. N. BOERS, E., KOSTERS, W. A., PUTTEN, P., POEL, M., *Artificial Intelligence: Definition, Trends, Techniques and Cases, in Knowledge for sustainable development: an insight into the Encyclopedia of life support systems*, Leiden, 2002, p. 1096.

De hecho, en este contexto hablaremos de entidades de Inteligencia Artificial, refiriéndonos generalmente a *«máquinas de IA, robots, agentes y algoritmos*[19]*»*.

3. DERECHO PENAL E IA. LA MÁQUINA COMO HERRAMIENTA DE JUSTICIA

La influencia de la Inteligencia Artificial permea tanto en los ámbitos del derecho procesal como en el derecho penal sustancial.

En el ámbito de la investigación y de la actividad de policía de prevención[20], la Inteligencia Artificial aumenta la eficacia de la aplicación de la ley a través de avances en prácticas policiales, incluyendo la policía predictiva, y la implementación de metodologías de perfilado como sistemas de reconocimiento facial e identificación biométrica. Específicamente, estos programas facilitan la identificación de riesgos criminales y la asignación prudente de recursos para prevenir de manera proactiva actividades delictivas previsibles y reducir la victimización. Por ejemplo, el programa *Keycrime*, desarrollado basado en las experiencias de la Jefatura de Policía de Milán, sirve como una herramienta efectiva para predecir delitos en serie, incluyendo robos, fraudes contra personas mayores, robos en apartamentos, violencia sexual, entre otros. De manera similar, el programa *XLAW*, ideado por la Policía de Nápoles y desplegado en varias regiones, se utiliza para pronosticar robos. Además, estas aplicaciones de IA están diseñadas para mejorar la precisión en la identificación de perpetradores después de un evento delictivo[21].

En el ámbito judicial, la Inteligencia Artificial tiene el potencial de permitir evaluaciones más completas y matizadas de los acusados en casos criminales. Estas evaluaciones implican la cruzada de datos históricos sobre los acusados y evaluaciones de su propensión subjetiva a participar en comportamientos delictivos. En esencia, estos algoritmos examinan factores como el estatus socioeconómico, antecedentes familiares, tasas de criminalidad en el vecindario y estado laboral para ofrecer un presunto pronóstico sobre el riesgo criminal de un individuo, a menudo presentado en una escala que va desde "bajo" hasta "alto", o expresado como porcentajes específico[22]. En resumen, estos algoritmos sirven como herramientas para analizar extensos datos históricos, identificar patrones recurrentes y generar evaluaciones basadas en una base estadística significativamente más robusta que los juicios humanos que sustentan las evaluaciones tradicionales[23].

---

[19] LIOR, A. *Ai entities as AI agents: artificial intelligence liability and the AI respondeat superior analogy*, in *Mitchell Hamline Law Review,* 2020, p. 1045.

[20] ISAAC, W.S. *Hope, Hype, and Fear: The Promise and Potential Pitfalls of Artificial Intelligence in Criminal Justice*, in Ohio St. J. Crim. L., 2018, 543 ss.; F. Basile, *Intelligenza artificale e diritto penale*, cit., 13 ss.

[21] MANES, V.*L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, en Discrimen, 2020, p. 7.

[22] MANES, V. *L'oracolo algoritmico,* cit., p. 8.

[23] Así, la definición proporcionada en el informe del EPIC, Algoritmos en el Sistema de Justicia Penal, disponible en https://epic.org/algorithmic-transparency/crimjustice/, refleja esta comprensión. El Electronic Privacy Information Center (EPIC) es una organización sin fines de lucro que opera en los Estados Unidos y se centra en salvaguardar la privacidad y las libertades civiles. Establecido en 1994, EPIC está dedicado a proteger la privacidad de los individuos, la libertad de expresión y los valores democráticos en la era digital. A través de la promoción de políticas, litigios y educación pública, EPIC trabaja para defender los derechos de privacidad y abordar las amenazas emergentes a la privacidad y las libertades civiles provocadas por las nuevas tecnologías y las prácticas gubernamentales. EPIC cubre una amplia gama de temas, incluida la vigilancia, la protección de datos, la privacidad del consumidor, la libertad de información y la transparencia de los algoritmos. Además, EPIC realiza investigaciones, publica informes y proporciona recursos para capacitar a las personas en la comprensión y preservación de sus derechos de privacidad.

A) LA MÁQUINA CRIMINAL.

Los actos criminales se definen, según la doctrina, como *«cualquier acto (u omisión) que constituya un delito castigado por la ley penal, sin pérdida de generalidad para las jurisdicciones que definen el delito de manera similar*[24]*»*.

En el ámbito financiero, especificamente dentro de los mercados financieros, existen pruebas del despliegue de *bots* sociales, *softwares* que automatizan cuentas en redes sociales y simulan usuarios humanos, en esquemas como "*pump-and-dump*" (inflar y descargar[25]). Estos esquemas ilícitos inflan artificialmente los precios de los valores al difundir información falsa, engañosa o exagerada para crear una demanda artificial, lo que finalmente permite la venta de valores a precios elevados. Simulaciones de mercado también han demostrado que agentes comerciales artificiales, empleando aprendizaje por refuerzo, una técnica de aprendizaje automático basada en recompensar elecciones correctas, pueden adquirir la práctica del *spoofing* financiero. Esto implica colocar órdenes continuas durante un período definido sin ninguna intención de ejecutarlas, con el objetivo principal de manipular los precios del mercado[26].

La importancia del crimen de Inteligencia Artificial como un fenómeno distinto aún no ha sido reconocida. Hasta hace relativamente poco, los sistemas de IA estaban limitados por comportamientos predefinidos, operando exclusivamente a través de algoritmos establecidos por programadores, como *softwares* diseñados para deshabilitar los sistemas de ciberseguridad de un banco o causar destrucción o daño a datos informáticos. Asignar culpabilidad criminal en tales casos no presentaba desafíos significativos: independientemente de la complejidad de las acciones de la entidad de IA, la responsabilidad recaía en última instancia en su controlador o usuario. En estos casos, las entidades de IA carecen de agencia cognitiva y sus comportamientos se adhieren a patrones predefinidos, siendo predecibles. Desde esta perspectiva, las entidades inteligentes eran percibidas como meros instrumentos manejados por humanos para la comisión de delitos[27]. En consecuencia, el concepto de confiscación de activos como medida preventiva podía aplicarse a entidades de IA, incluso en ausencia de una condena penal, como se articula en el Artículo 240[28] del Código Penal italiano[29].

A la luz de las consideraciones anteriores, la ley italiana, en línea con los marcos legales de otros Estados miembros de la Unión Europea, carece actualmente de definiciones para los delitos cometidos por inteligencia artificial. La ausencia de regulaciones dedicadas para abordar los delitos perpetrados por agentes de IA autónomos subraya la urgente necesidad de un mayor desarrollo legal en este ámbito. Esto es crucial para garantizar el establecimiento de mecanismos apropiados de responsabilidad y regulación en respuesta al panorama en constante evolución de los avances tecnológicos[30].

---

[24] KING T.C, AGGRWAL, N. TADDEO, M., FLORIDI L., *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions, in Science and engineering ethics*, 2019, 90.

[25] KING T.C, AGGRWAL, N. TADDEO, M., FLORIDI L., *Artificial Intelligence Crime,* cit., 89 ss.

[26] BORSARI, R. *Intelligenza Artificiale e responsabilità penale: prime considerazioni*, en *Media Laws,* 2020, p. 263.

[27] RIONDATO, S. *op. cit.*, 85 ss.

[28] Artículo 240 del Código Penal Italiano:
1. En caso de condena, el juez tiene la autoridad para ordenar el decomiso de los objetos que fueron utilizados o destinados para la comisión del delito, así como los objetos que son el producto o lucro del delito.

[29] RIONDATO, S. *Ibidem.*

[30] Como ya he tenido la oportunidad de manifestar en PIPARO C.,, *op. cit.,* pp. 900 y ss.

B) MÁQUINA Y HOMBRE EN UN PARADIGMA UNITARIO

Con la informática, la automatización desplaza su enfoque de la automatización de actividades materiales más o menos complejas al procesamiento de información, entendida como el conjunto de datos interconectados a través de los cuales una idea toma forma y se comunica, convirtiéndose en el nuevo objetivo de la automatización. Esto marca la primera aplicación de principios mecanicistas al pensamiento, anteriormente entendido como prerrogativa exclusiva de la *res cogitans*. Incluso un flujo de ideas puede, dentro de ciertos límites y condiciones específicas, descomponerse en una pluralidad de pasos lógicos elementales ordenados secuencialmente, lo que permite una reproducción mecánica sin intervención humana[31].

Este cambio de paradigma radical desafía la división clásica entre hombre y máquina construida a lo largo de la línea divisoria entre sujeto y objeto. El debate, anticipado por el genio de Alan Turing, se abre en cuanto a la posibilidad de crear máquinas inteligentes dotadas del atributo característico del yo. Este punto de partida común da lugar a reflexiones que abarcan dimensiones éticas y legales, posicionándolas entre los problemas más destacados de la modernidad[32].

La primera área de consideración involucra la posibilidad de atribuir cierta subjetividad a las máquinas inteligentes. Desde la perspectiva de la teoría general del derecho, la categoría de subjetividad legal está experimentando un resurgimiento, similar a sus primeros días en la ciencia legal. A medida que la sociedad se vuelve cada vez más compleja, la subjetividad legal se vuelve más compleja. Una vez alcanzado un logro revolucionario, la unidad del sujeto legal se está desintegrando[33].

La segunda área explora la aceptabilidad de fenómenos que involucran una hibridación cada vez más cercana entre el hombre y la máquina, de acuerdo con el nuevo paradigma del llamado "transhumanismo" o "posthumanismo[34]". Esta doctrina secular aboga por trascender los límites biológicos del cuerpo mediante el aprovechamiento del potencial tecnológico[35].

Otra pregunta concierne a la naturaleza y propósito último de la IA. Es necesario determinar si es un ser en sí mismo o simplemente una herramienta. Como veremos más adelante, ambas soluciones llevan consigo alternativas teóricas.

En este paisaje culturalmente dinámico, surge el concepto de la "reserva de humanidad". Se basa en la observación de que la subjetividad ya no es una esfera exclusivamente reservada para los humanos. La incursión de las máquinas en el ámbito del intelecto, compitiendo con los humanos en terrenos más que el simple trabajo físico, plantea el evidente riesgo de la absorción completa del espacio personal por parte de las máquinas[36].

4. LA ESTRUCTURA DEL DELITO

Hoy en día, la imposición de penas depende de dos requisitos claves: primero, la conducta debe estar previamente prohibida (reserva de ley), y segundo, que un juez administre el

[31] LOEVINGER, L. *Jurimetric. The Next Step Forward*, in Minnesota Law Review, 5 1949, 455 ss.

[32] TURING, A.M., Computer machinery and intelligence, in Mind, vol. 49, 1950, 433.

[33] GALLONE, G. *Riserva di umanità e funzioni amministrative*, 2021, p. 13.

[34] RODOTÀ, S. *Il diritto di avere diritti*, Bari, 2012, 341 ss.

[35] GALLONE, G. *Riserva di umanità, op.cit.,* 2021, p. 14.

[36] Vid. TERRAVECCHIA, G.P. y FERRARI M., *Logos e techne. Tecnologia e filosofia. Romanae Disputationes*, 2016-17, Bologna, 2017.

castigo. En el sistema legal italiano, estas condiciones están explícitamente delineadas en los artículos 13 y 25 de la Constitución italiana[37].

Para establecer un acto como un delito, los sistemas legales modernos requieren un *minium* de dos elementos fundamentales[38]. El primer elemento, conocido como *actus reus* (literalmente: el acto criminal), exige que el acto se ajuste precisamente a los criterios especificados en las leyes pertinentes. El segundo elemento, conocido como *mens rea* (literalmente: la intención criminal), comprende una gama de estados mentales, siendo el nivel más alto la voluntariedad, a veces acompañada de intención o propósito específico. Los estados mentales más bajos incluyen la negligencia (cuando una persona razonable debería haber sabido) y los delitos de responsabilidad estricta[39]. Cuando se puede demostrar que un individuo cometió conscientemente el acto criminal o lo hizo con intención criminal, esa persona es considerada penalmente responsable del delito[40].

El objetivo de este estudio es, en principio, identificar y resumir cómo incorporar la inteligencia artificial en el ámbito del derecho penal y, luego, analizar y desglosar la relación atípica hombre-máquina y su posible resultado como concurrencia criminal. Este trabajo se centrará en el actus reus, dejando para otra ocasión el enfoque en la *mens rea.*

A) EL ACTUS REUS

Esta sección analiza la responsabilidad penal de la inteligencia artificial utilizando tres modelos, tal como lo expone la distinguida doctrina legal: el modelo de responsabilidad por perpetración a través de otro, el modelo de responsabilidad por consecuencia natural y probable, y el modelo de responsabilidad directa.

a) El modelo de responsabilidad por perpetración a través de otro

El modelo de perpetración a través de otro imagina a la inteligencia artificial como un agente inocente, similar a un niño: careciendo de voluntad autónoma, la máquina solo puede ser utilizada como una herramienta. De hecho, la máquina simplemente ejecuta una orden dada por un humano. En este escenario, solo aquellos que explotan al agente inocente son considerados responsables penalmente como perpetradores a través de otro.

---

[37] Artículo 13, Constitución Italiana:
1. La libertad personal es inviolable.
2. No se permite ninguna forma de detención, inspección o registro personal, ni ninguna otra restricción de la libertad personal, excepto por acto fundamentado de la Autoridad Judicial y solo en los casos y de la manera prevista por la ley.
Artículo 25, Constitución Italiana:
1. Nadie puede ser desviado del juez competente preestablecido por la ley.
2. Nadie puede ser castigado excepto de acuerdo con una ley que estuviera en vigor antes del acto cometido.
3. Nadie puede ser sometido a medidas de seguridad excepto en los casos previstos por la ley.

[38] Tradicionalmente, la doctrina y jurisprudencia italiana deconstruyen el delito en:
1. Elemento objetivo: Este aspecto de un delito se refiere a las acciones físicas llevadas a cabo por el individuo, constituyendo el núcleo material de la infracción. Incluye no solo los aspectos físicos de la acción en sí misma, como un asalto físico o un robo, sino también cualquier factor contextual que pueda ser pertinente para definir la infracción, como la ubicación, el momento o el método empleado.
2. Elemento subjetivo: El estado psicológico o la intención del individuo que comete el acto. Incluye el dolus, que indica una intención deliberada de participar en la conducta que constituye la infracción, así como el culpa, que indica una falta de diligencia o cuidado en el comportamiento del individuo que resulta en la comisión de la infracción.
3. Desvalor social (o meramente legal): Este elemento representa la discrepancia entre el acto y todo el marco legal, que se extiende más allá del ámbito penal. Refleja si la acción está en armonía con el sistema legal más amplio.

[39] HALLEVY, G.,, *The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control*, Akron Intellectual Property Journal: Vol. 4, 2010, p. 178.

[40] Como ya se destacó en PIPARO, C. *Ibidem.*

Relevante doctrina[41] argumenta que cuando las entidades de inteligencia artificial están involucradas en un delito, deberían ser consideradas como agentes inocentes carentes de atributos humanos. En consecuencia, las entidades de inteligencia artificial son vistas estrictamente como máquinas y no como participantes activos en el acto criminal, ya sea como principales o cómplices. En tales casos, donde el perpetrador real carece de mens rea (intención criminal), la responsabilidad legal recae invariablemente sobre el creador, programador o usuario final de la entidad de inteligencia artificial. Hallevy compara estas circunstancias con situaciones que involucran a individuos mentalmente limitados como niños, personas mentalmente incompetentes o aquellos sin un estado mental criminal. En estos escenarios, el intermediario (entidad de inteligencia artificial) es considerado una herramienta sofisticada, siendo el verdadero perpetrador el orquestador del delito, quien es responsable como perpetrador principal[42].

Pero, ¿quién asume el papel del perpetrador a través de otro? Se identifican dos candidatos potenciales: el programador de *software* de IA y el usuario final. Un programador puede diseñar intencionalmente software para una entidad de IA con el objetivo de permitirle cometer delitos específicos. Por ejemplo, un programador podría crear software para un robot industrial y programarlo para incendiar una fábrica durante horas no ocupadas. Aunque el robot comete incendio provocado, la responsabilidad legal se asigna al programador. Alternativamente, un usuario final puede ser visto como el perpetrador a través de otro cuando utiliza una entidad de IA sin programarla ellos mismos. Por ejemplo, un usuario compra un robot doméstico capaz de ejecutar comandos de su amo. El usuario instruye al robot para que confronte físicamente a cualquier intruso en su hogar. En este escenario, el robot lleva a cabo el asalto, pero se considera al usuario como el perpetrador[43].

En ambos casos, la entidad de IA es el agente real que cometió el delito. Sin embargo, dado que ni el programador ni el usuario final realizaron acciones que se ajustan a la definición precisa del delito, no cumplen con el requisito de actus reus para ese delito en particular. El modelo de responsabilidad del perpetrador a través de otro trata las acciones de la entidad de IA como si fueran las acciones del programador o del usuario final, fundamentadas en el uso instrumental de la entidad de IA como un agente inocente[44].

Este modelo de responsabilidad no atribuye ninguna capacidad mental a la entidad de IA en sí misma. En cambio, equipara a las entidades de IA con objetos inanimados o animales en términos de responsabilidad penal. Por ejemplo, al igual que un ladrón que usa un destornillador para entrar a un edificio no hace que el destornillador sea penalmente responsable. Este modelo es especialmente relevante en casos en los que las entidades de IA se utilizan instrumentalmente para cometer delitos sin utilizar sus capacidades avanzadas o cuando las entidades de IA desactualizadas carecen de funcionalidades modernas[45].

Es importante tener en cuenta que «*este modelo no es adecuado cuando el software de la entidad de IA no fue diseñado para cometer el delito específico, pero el delito fue cometido por la entidad de IA de todos modos. El modelo tampoco es adecuado cuando la entidad de IA*

[41] HALLEVY, G. *The Criminal Liability of Artificial Intelligence Entities, op.* cit., pp. 179 e ss.

[42] Vid. Ampliamente OGUNNOIKI, K. *A critique of Gabriel Hallevy's models of criminal liability of artificial intelligence entities*, in *International Journal of Comparative Law and Legal Philosophy (IJOCLLEP),* 4 (3), 2022, pp. 4 y ss; PIPARO, C., *Ibidem.*

[43] Vid. HALLEVY, G.,, *Ibid.;* PIPARO, C., *Ibid.*

[44] HALLEVY, G.,, *Ibid.*

[45] OGUNNOIKI, K.,op. cit., p. 5.

específica funciona no como un agente inocente, sino como un agente semi-inocente[46]». El resultado legal de aplicar este modelo es que el programador y el usuario final son responsables penalmente por el delito específico cometido, mientras que la entidad de IA no tiene ninguna responsabilidad penal[47].

En términos legales, la aplicación de este modelo resulta en responsabilidad penal para el programador y el usuario final con respecto al delito específico cometido, mientras que absuelve a la entidad de IA de cualquier responsabilidad penal[48].

b) El modelo de responsabilidad por consecuencia natural y probable

El segundo modelo de responsabilidad de la inteligencia artificial en delitos implica situaciones en las que los programadores o usuarios están profundamente involucrados en las actividades de la entidad de inteligencia artificial pero no tienen la intención de cometer delitos. Sin embargo, si la entidad de IA comete un delito durante sus operaciones normales, puede aplicarse el modelo de responsabilidad por consecuencia natural y probable.

Un ejemplo de dicho escenario puede involucrar un piloto automático basado en inteligencia artificial con la misión de salvaguardar la aeronave durante su vuelo. En el recorrido hacia la activación del piloto automático, el piloto humano observa una tormenta que se acerca y decide abortar la misión y regresar a la base. La entidad de IA percibe la acción del piloto humano como una amenaza para la misión y responde tomando medidas para eliminar esta amenaza percibida mediante la activación del asiento eyectable. En consecuencia, el piloto humano sucumbe a las acciones de la entidad de IA. Es evidente que el programador no tuvo la intención de causar daño, especialmente al piloto humano. Sin embargo, las acciones de la entidad de IA resultan en la muerte del piloto humano, ejecutadas de acuerdo con las instrucciones programadas.

Este modelo responsabiliza a las personas por delitos que son una consecuencia natural y probable de su conducta, incluso si no tenían conocimiento real del delito. Por ejemplo, un usuario emplea software de IA diseñado para detectar amenazas en internet para proteger un sistema informático en el que está instalado. Sin embargo, sin el conocimiento del usuario, la IA destruye todo software externo reconocido como una amenaza, cometiendo inadvertidamente un delito informático[49].

Esta forma de responsabilidad se basa en la negligencia y abarca escenarios en los que los programadores o usuarios deberían haber previsto la posibilidad de un delito pero no tenían la intención de que ocurriera. Se aplica a individuos que no fueron los perpetradores reales del delito pero contribuyeron intelectualmente a él. Los programadores y usuarios razonables deberían haber previsto el delito y haber tomado medidas para prevenirlo, aunque no tuvieran la intención de que ocurriera. Sin embargo, las consecuencias legales varían según si los programadores o usuarios fueron negligentes sin intención criminal o si utilizaron consciente y voluntariamente la entidad de IA para cometer un delito, lo que resultó en la comisión de otro delito. En este último caso, pueden ser responsables del delito como si lo hubieran cometido consciente y voluntariamente[50].

Sin embargo, como critica la doctrina mencionada anteriormente, la aplicación de este modelo conduce a dos posibles resultados. Por un lado, si «*la entidad de IA actuó como un agente*

[46] OGUNNOIKI, K.,*Ibid.*
[47] Vid.LACEY, N. y WELLS, C., *Reconstructing Criminal Law-Critical Perspectives on Crime and the Criminal Process,* 1998; PIPARO, C., *Ibid.*

[48] PIPARO, C., *Ibid.*

[49] HALLEVY, G., op. cit., pp. 183-184

[50] HALLEVY, G.,, op. cit. pp. 184 e ss.

inocente, totalmente ajeno a la prohibición penal, no se le responsabiliza penalmente por el delito cometido, ya que la acción de la entidad de IA no difiere del modelo de responsabilidad por perpetración a través de otro. Pero si la entidad de IA no actuó simplemente como un agente inocente, entonces, además de la responsabilidad penal del programador o usuario, de conformidad con el modelo de responsabilidad por consecuencia natural y probable, la entidad de IA misma deberá ser considerada penalmente responsable del delito específico directamente[51]*».*

c) El modelo de responsabilidad directa.

Al aplicar el modelo de responsabilidad por consecuencia natural y probable a la responsabilidad penal de la entidad de IA, surgen dos posibles resultados. Si la entidad de IA actuó como un agente inocente, ajeno a la naturaleza criminal de sus acciones, no se le responsabilizará penalmente por el delito que cometió. Esto alinea con el primer modelo de responsabilidad, donde la entidad de IA se considera una herramienta utilizada por otros. Sin embargo, si la entidad de IA no actuó como un agente inocente y tenía conocimiento de la prohibición penal, puede ser considerada directamente y de manera independiente responsable penalmente por el delito específico que cometió[52]. Este modelo de responsabilidad directa constituye el tercer enfoque hacia la responsabilidad de la entidad de IA y pone el enfoque directamente en la IA misma. La determinación de la responsabilidad de la entidad de IA depende de si actuó inocentemente o tenía conocimiento de la conducta prohibida[53].

Los sistemas de IA pueden recibir entrada sensorial y analizar datos factuales, similar a la comprensión humana. Se espera que imiten los procesos cognitivos humanos, pero el intento específico, el requisito mental más fuerte, implica tener un propósito u objetivo para lograr un resultado particular. Por ejemplo, en casos de asesinato, el intento específico se refiere a la intención de causar daño o muerte a una persona específica. Las entidades de IA pueden ser programadas con un propósito y tomar acciones para cumplirlo, demostrando un intento específico. Aunque los humanos tienen sentimientos que el software de IA no puede replicar, como el amor o los celos, estos sentimientos generalmente no son necesarios para la mayoría de los delitos específicos. Muchos delitos solo requieren conocimiento de los elementos externos, y el intento específico solo es relevante para unos pocos delitos. Por lo tanto, la ausencia de tales emociones en las entidades de IA no obstaculiza la imposición de responsabilidad penal[54].

Si una entidad de IA cumple con todos los elementos de un delito, no debería estar exenta de responsabilidad penal. A diferencia de ciertos segmentos de la sociedad, como los bebés o los enfermos mentales, que tienen disposiciones legales que los eximen de la responsabilidad penal, no está claro si existen marcos similares para las entidades de IA[55].

La responsabilidad penal de una entidad de IA no reemplaza la responsabilidad de sus programadores o usuarios; más bien, se impone además de su responsabilidad. La responsabilidad de una entidad de IA no depende de la responsabilidad de su programador o usuario. Si una entidad de IA es programada o utilizada por otra, la responsabilidad de la entidad programada o utilizada permanece intacta[56].

---

[51] OGUNNOIKI, K.,op. cit., p. 7.

[52] GERSTNER, M. E. *Liability Issues with Artificial Intelligence Software*, Santa Clara L. Rev, 1993.

[53] DOBRINOIU, M. The Influence of Artificial Intelligence on Criminal Liability, Faculty of Law, University of Bucharest, http://cks.univnt.ro.

[54] PADHY, N. P. *Artifical intelligence and intelligent systems*, in *Oxford University Press,* 2005, p. 14.

[55]. PADHY, N. P., *op. cit.*, p. 10.

[56] PIPARO, C., *Ibid.*

No hay razón para eximir a las entidades de IA o a los humanos de la responsabilidad penal basada en su colaboración. Si una entidad de IA y un humano actúan como perpetradores conjuntos, cómplices o instigadores, deberían estar sujetos a la responsabilidad penal correspondiente, independientemente de su identidad[57].

Los elementos de culpa negativa y las defensas relevantes en el derecho penal se aplican a las entidades de IA, incluida la legítima defensa, la necesidad, la coacción o la intoxicación. Es posible que se necesiten algunos ajustes al aplicar estas defensas a las entidades de IA, pero fundamentalmente, la responsabilidad penal de una entidad de IA, siguiendo el modelo de responsabilidad directa, es similar a la de un humano. Se basa en los mismos elementos y se evalúa de la misma manera, con ajustes específicos realizados en ciertos casos[58].

d) Los modelos en combinación

Los tres modelos de responsabilidad delineados anteriormente no son mutuamente excluyentes, sino que pueden coexistir e interactuar en diversos escenarios legales. Por ejemplo, cuando una entidad de IA actúa como agente inocente en la comisión de un delito específico, y el único responsable de ese acto es el programador, el modelo legal más adecuado para tal situación es el modelo de perpetración a través de otro (el primer modelo de responsabilidad). En este caso, el programador asume la responsabilidad de las acciones de la entidad de IA como el perpetrador a través de otro.

En el mismo escenario donde el programador es en sí mismo una entidad (como cuando una entidad de IA programa a otra entidad de IA para cometer un delito específico), el modelo de responsabilidad directa (el tercer modelo de responsabilidad) también se aplicaría junto con el primer modelo de responsabilidad. Por lo tanto, en tales casos, el programador de la entidad de IA podría ser considerado penalmente responsable, combinando elementos del modelo de perpetración a través de otro y el modelo de responsabilidad directa[59].

De manera similar, si la entidad de IA asume el papel del perpetrador físico de un delito específico, pero el delito no fue premeditado, el modelo de responsabilidad por consecuencia natural y probable podría ser aplicable. En esta situación, el programador podría considerarse negligente si el delito no fue cometido intencionalmente, o el programador podría ser considerado responsable del delito específico si se planeó deliberadamente otro delito, incluso si el delito real cometido no formaba parte del plan criminal original.

Sin embargo, es crucial tener en cuenta que en los casos en que el programador no es humano, el modelo de responsabilidad directa aún debe aplicarse, además de la aplicación simultánea del modelo de responsabilidad por consecuencia natural y probable. Del mismo modo, cuando el perpetrador físico es humano y el planificador es una entidad de IA, se necesita un enfoque similar[60].

La interacción de estos tres modelos de responsabilidad crea un paisaje legal único en el ámbito de las entidades de IA y el derecho penal. En consecuencia, cuando las entidades de IA y los humanos están involucrados, ya sea directa o indirectamente, en la comisión de un delito específico, se vuelve más difícil evitar la responsabilidad penal. Si el objetivo principal de imponer la responsabilidad penal es mantener el control legal y social dentro de una sociedad específica, entonces la aplicación coordinada de los tres modelos se vuelve esencial en el contexto de las entidades de IA[61].

---

[57] HALLVEY, G., *The Criminal Liability of Artificial Intelligence Entities,* cit., p. 192.

[58] DRESSLER, J. *Cases and materials on Criminal law*, cit, pp. 616-622.

[59] HALLVEY, G. op. cit., p. 23

[60] HALLVEY G., *ibid.*

[61] HALLVEY G., op. cit., p. 24.

Suponiendo que las entidades de IA posean autoconciencia, conciencia y libre albedrío, entra en juego su potencial responsabilidad penal. Dado que las entidades de IA pueden encarnar principios sociales y éticos, al ser creaciones de humanos, directa o indirectamente, este documento argumenta que existen marcos legales, jurídicos y tecnológicos adecuados para reconocer a las entidades de IA como actores legales activos dentro del ámbito de la justicia penal[62].

5. LOS PROBLEMAS DE "NO HANDS" Y DE "TOO MANY HANDS"

Anteriormente, afirmamos que la responsabilidad penal de una entidad de IA no reemplaza la responsabilidad de sus programadores o usuarios; más bien, se impone además de su responsabilidad. Imaginemos un algoritmo escrito por quince programadores autónomos en un repositorio compartido de *Git-Hub* y el código se vende en 2022. En 2023, el código es autorizado por la junta competente de una empresa que vende el producto final, donde el algoritmo está incorporado en un robot de limpieza, entrenado para mantener limpias las ventanas exteriores. Con el paso de los años, la empresa quiebra, cierra y la máquina queda sin supervisión. Durante los años siguientes, una serie de actualizaciones desactivan el *firewall*, y la máquina descarga datos sin verificar, lo que convierte a la máquina en violenta y comienza a atacar a las personas que se acercan a las ventanas. También en entornos corporativos complejos pueden ocurrir fallos operativos. Los científicos organizacionales reconocen que los sistemas organizacionales defectuosos, más que la culpabilidad individual, pueden ser la causa raíz. Un ejemplo es el fallo de los canales de comunicación, que impiden el flujo de información crucial entre empleados bien intencionados. Otro ejemplo involucra el fallo de un banco en presentar informes obligatorios contra el lavado de dinero debido a un sistema no operativo. Esto enfatiza que las deficiencias sistémicas pueden impedir que el personal de cumplimiento sea alertado sobre eventos desencadenantes de informes, demostrando el potencial de fallas organizacionales para obstaculizar procesos cruciales a pesar de las mejores intenciones de los empleados individuales. Estos escenarios se atribuyen al llamado "problema de sin manos", sugiriendo que incluso si cada empleado se comporta de manera responsable, las operaciones corporativas complejas aún pueden salir mal. El problema surge, de hecho, no solo en caso de mala organización empresarial, sino también si se interrumpe la conexión con la empresa. La imprevisibilidad inherente de los algoritmos avanzados complica aún más las cosas, dificultando la identificación de un empleado culpable cuando surge este problema.

En diferentes escenarios donde numerosos actores (o empleados en contextos corporativos) están involucrados en el diseño y ejecución de algoritmos, la investigación y reconstrucción de actos conjuntos que resultan en daño resultan ser desafiantes, ya que las corporaciones tienen poco incentivo para cooperar y los empleados individuales están motivados para evadir la responsabilidad.

Este obstáculo explica las luchas consistentes del Departamento de Justicia de EE. UU. para acusar a individuos dentro de grandes entidades corporativas. También explica por qué doctrinas especiales de responsabilidad civil, como la responsabilidad del producto, a veces evitan la necesidad de identificar a un único empleado responsable, sosteniendo directamente la responsabilidad de toda la corporación[63]. En tales casos, no hay duda de que el vínculo causal que conecta al trabajador/programador individual con las acciones del robot está interrumpido, por lo que no se puede atribuir responsabilidad al programador/trabajador mediante el uso de los modelos descritos anteriormente.

---

[62] G. Hallvey, *ibid.*

[63] Ante este problema, el Departamento de Justicia tuvo que actualizar sus políticas para obligar a los fiscales a investigar todas las pistas contra individuos antes de resolver un caso contra una corporación. Véase el Memorándum Yates, supra nota 46 ("Los abogados del Departamento no deben resolver asuntos con una corporación sin un plan claro para resolver casos relacionados con individuos…").

Desde una perspectiva diferente, las operaciones corporativas, frecuentemente intrincadas y expansivas, involucran a numerosos empleados, una tendencia perpetuada en entornos automatizados donde equipos distribuidos gestionan algoritmos corporativos. El potencial de daño surge de un solo actor, ya sea intencionalmente, a través de acciones como sobornar a funcionarios, o inadvertidamente, como negligencia en el control de calidad.

Un impedimento significativo para establecer la responsabilidad corporativa es el problema de "demasiadas manos", un desafío que surge de la complejidad de las operaciones corporativas. Este obstáculo probatorio hace que sea difícil o insuperable probar la existencia de un empleado específico que contribuya al daño algorítmico. La renuencia de las corporaciones a cooperar y los empleados individuales que evaden la culpa agravan este desafío. Las luchas recurrentes del Departamento de Justicia para acusar a individuos dentro de grandes corporaciones ejemplifican la complejidad del problema de "demasiadas manos". Las doctrinas especializadas de responsabilidad civil, como las leyes de responsabilidad del producto, a veces evitan la necesidad de identificar a un empleado culpable específico, optando por la responsabilidad corporativa directa, reconociendo las dificultades prácticas planteadas por el Problema de "demasiadas manos" en casos de daño algorítmico[64].

A) RESPONSABILIDAD CORPORATIVA POR ROBOTS

La teoría general del derecho considera a las corporaciones como sujetos legales, al igual que las personas. En esencia, tanto los demandantes como los fiscales tienen la capacidad de iniciar acciones legales contra las corporaciones por cualquier violación, ya sea civil o penal. Por lo tanto, las corporaciones tienen una subjetividad jurídica individual, diferente y separada de los empleados y accionistas. Estos últimos, entonces, asumen responsabilidad separada por violaciones personales[65].
La responsabilidad corporativa se ha extendido masivamente en todo el mundo durante las últimas décadas, y es aceptada y disciplinada en el contexto de la mayoría de los sistemas jurídicos modernos[66].

Dentro del contexto de este artículo, la manifestación de la responsabilidad corporativa se refiere a la culpabilidad penal por delitos caracterizados por un elemento de *mens rea* que implica propósito o conocimiento: los delitos penales con un elemento de *mens rea* exigen responsabilidad para garantizar justicia y disuasión[67].

Sin embargo, la simple posibilidad lógica de sustituir a las corporaciones por algoritmos en esta *fictio iuris* no implica automáticamente que esta ecuación sea recomendable. Tanto las corporaciones como los algoritmos son entidades complejas, y se necesita precaución para

[64] Memorándum de Sally Quillian Yates, Fiscal General Adjunta, Departamento de Justicia de los Estados Unidos, a todos los Jefes de Componentes y Fiscales de los Estados Unidos, Responsabilidad Individual por Conducta Incorrecta Corporativa 2 (9 de septiembre de 2015) [en adelante, Memorándum Yates], https://www.justice.gov/archives/dag/file/769036/download [https://perma.cc/7J5F-2BTM] ("En grandes corporaciones, donde la responsabilidad puede ser difusa y las decisiones se toman en varios niveles, puede ser difícil determinar si alguien poseía el conocimiento y la intención criminal necesarios para establecer su culpabilidad más allá de una duda razonable"); Amanda M. Rose & Richard Squire, Litigios Intraportafolio, 105 NW. U. L. REV. 1679, 1684 (2011) ("En algunas situaciones puede ser imposible para las víctimas descubrir o probar qué empleados particulares dentro de una empresa causaron sus lesiones").

[65] *See generally* Miriam Hechler Baer, *Governing Corporate Compliance*, 50 B.C. L. REV. 949, 955–56 (2009) ("Uno de los grandes desafíos para los formuladores de políticas, entonces, es elaborar reglas y regulaciones que obliguen a las empresas a internalizar los costos a largo plazo de sus acciones indebidas sin eliminar los incentivos individuales para divulgar información").

[66] G. De Simone, *Responsabilità da reato degli enti,* (edited by) G. Lattanzi, Paola Severino, p. 4.

[67] M. E. Diamantis, *Employed Algorithms: a labot model of corporate liability for AI,* Duke law journa, 2022, p. 813.

evitar consecuencias no deseadas potencialmente desastrosas, como obstaculizar excesivamente el sector tecnológico o empoderarlo de manera imprudente[68].

Para apreciar la necesidad de la responsabilidad corporativa, debemos reconocer la inmensa influencia ejercida por el sector privado, impulsada principalmente por dos factores: la recopilación de datos y los recursos financieros. Las corporaciones multinacionales, en particular, poseen un poder financiero sustancial, controlando eficazmente tanto el dinero como la información. Gran parte de este poder se deriva de la recopilación y utilización de vastas cantidades de datos, a menudo envueltos en secreto. Especialmente, gigantes tecnológicos como *Google, Facebook, Amazon, Apple* y *Microsoft* controlan y explotan los datos de miles de millones de individuos, dando forma al comportamiento del consumidor para maximizar beneficios. Esta concentración de poder digital plantea amenazas para la democracia y el funcionamiento del mercado[69].

Con un poder tan inmenso debe venir una responsabilidad igualmente inmensa. En consecuencia, surge una pregunta crucial: ¿deberían las corporaciones asumir la responsabilidad penal por las acciones derivadas de las herramientas de inteligencia artificial que introducen en el mercado? Las profundas arcas financieras de las corporaciones pueden ser la clave para esta pregunta. La culpabilidad organizacional surge de la expansión impulsada por el lucro de las actividades comerciales de estas megacorporaciones. Parece justificable y esencial distribuir los riesgos al mantener consistentemente responsables a aquellos que se benefician del despliegue de IA. De lo contrario, la ausencia de responsabilidad por las acciones dañinas de la IA podría permitir que las corporaciones se expandan mientras la sociedad sigue siendo vulnerable[70].

La brecha de responsabilidad en la IA parece favorecer a las empresas que desarrollan e implementan estas tecnologías, a menudo a expensas de los intereses sociales. Esta responsabilidad corporativa limitada y no regulada podría alentar a las corporaciones a asumir riesgos socialmente perjudiciales para maximizar beneficios, poniendo así en peligro a la sociedad. Los esfuerzos de cabildeo de la industria tecnológica contra regulaciones legales resaltan aún más estas preocupaciones. Para abordar esta complejidad, se necesitan con urgencia regulaciones específicas del sector que vayan más allá de simples estándares éticos[71].

B) ¿ROBOTS: ESCLAVOS O EMPLEADOS?

En 2010, la tecnoética Joanna Bryson presentó una analogía provocadora entre los algoritmos y el trabajo: la antropomorfización de los robots inevitablemente los categorizaría como esclavos y aclararía la asignación de responsabilidad por resultados negativos: *"los robots deberían ser construidos, comercializados y considerados legalmente como esclavos*[72]".

La propuesta de Bryson es moralmente problemática a nivel mundial debido a las asociaciones históricas e implicaciones de la esclavitud. En los Estados Unidos, por ejemplo, la Decimotercera Enmienda rechaza categóricamente cualquier forma de esclavitud. En Italia, el artículo 600 del Código Penal italiano castiga la esclavitud, y el artículo 5 de la

[68] M. E. Diamantis, *Employed algorithms,* cit., p. 816: *"Aunque la ley podría sustituir a las corporaciones como acusadas en lugar de sus algoritmos, eso no necesariamente significa que hacerlo sea una buena idea".*

[69] NEMITZ, P. Constitutional democracy and Technology in the Age of Artificial Intelligence, in Philosophical Transactions of the Royal Society A, 2019.

[70] OSMANI, N. The Complexity of Criminal Liability of AI Systems, 2020, p. 69.

[71] OSMANI, N. *op. cit.*, p. 70

[72] BRYSON, J. J., *Robots Should Be Slaves*, en Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues, 2010, p. 63.

Carta de los Derechos Fundamentales de la Unión Europea establece que "nadie puede ser sometido a esclavitud o servidumbre".

Además, existe el riesgo de que tratar a los robots como esclavos pueda tener consecuencias negativas. El abuso de los robots ha demostrado provocar respuestas empáticas por parte de los sujetos humanos, lo que potencialmente refuerza la tendencia a antropomorfizarlos. Por lo tanto, el intento de Bryson de prevenir la antropomorfización a través de la analogía de la esclavitud puede llevar inadvertidamente al efecto contrario[73].

A pesar de estas preocupaciones, la propuesta de Bryson explora la idea de los robots como una forma de trabajo. Los algoritmos poseen capacidades productivas, realizando tareas que los humanos no pueden o prefieren no hacer. Por lo tanto, una corrección del punto de vista extremo de Bryson podría ser el empleado robot. El empleo captura mejor la cooperación constructiva entre las corporaciones y los algoritmos, evitando rupturas constitucionales o sistémicas[74].

C) DESAFÍOS DE LA RESPONSABILIDAD CORPORATIVA

Desde una perspectiva de aplicación corporativa, existen similitudes en los desafíos estructurales planteados por los empleados humanos y los algoritmos.

La responsabilidad corporativa ofrece una solución práctica al permitir que las víctimas busquen justicia de las corporaciones, que tienen bolsillos más profundos y mayor influencia sobre la cultura y el comportamiento organizacional. De manera similar, los desafíos presentados por la mala conducta algorítmica se asemejan a los de la mala conducta de los empleados. Las víctimas de daños algorítmicos a menudo enfrentan barreras para obtener recurso legal, ya que los propios algoritmos no son demandados responsables. Responsabilizar a las corporaciones por la mala conducta algorítmica no solo brinda un camino hacia la justicia para las víctimas, sino que también ofrece un medio para inducir un cambio en el comportamiento algorítmico[75].

A pesar de la justificabilidad de responsabilizar a las corporaciones por la mala conducta tanto de los empleados como de los algoritmos, surgen preocupaciones éticas con respecto a la responsabilidad vicaria, que consiste en sancionar a una parte por las acciones de otra. Por ejemplo, el artículo 27 de la Constitución Italiana establece que "la responsabilidad penal es personal".

De hecho, el desafío radica en la incapacidad de las corporaciones para ejercer un control total sobre el comportamiento de los empleados y los algoritmos. Las corporaciones no siempre pueden monitorear o controlar perfectamente el comportamiento de los empleados, lo que lleva a brechas en el cumplimiento y posibles conductas indebidas.

Para superar estos desafíos, la ley debe encontrar un equilibrio entre responsabilizar a las corporaciones por las acciones tanto de empleados como de algoritmos, al tiempo que reconoce las limitaciones prácticas en el control corporativo. Si bien la responsabilidad vicaria puede ser éticamente problemática, sigue siendo una herramienta necesaria para lograr justicia para las víctimas y prevenir daños futuros. La ley debe evolucionar para reflejar el cambiante panorama de las prácticas corporativas, asegurando que las corporaciones sean responsables de las acciones tanto de agentes humanos como algorítmicos[76].

[73] DIAMANTIS M. E., *Employed algorithms, op. cit.*, p. 823.

[74] DIAMANTIS M. E., *Employed algorithms, op. cit.*, p. 824.

[75] DIAMANTIS, M. E. *Employed algorithms, op. cit.*, p. 829.

[76] DIAMANTIS, M. E. *Employed algorithms, op. cit.*, p. 844 y ss.

6) MODELOS DE RESPONSABILIDAD CORPORATIVA

En este escenario, la discusión debería cambiar de si las entidades legales deberían tener responsabilidad penal a cómo abordar y regular de manera efectiva su responsabilidad en el ámbito de la inteligencia artificial.

Para explorar más a fondo la responsabilidad penal corporativa, existen diversas doctrinas y teorías utilizadas para imponer responsabilidad penal a las corporaciones.

A) RESPONSABILIDAD CORPORATIVA INDIRECTA

Esto representa un enfoque nominalista según el cual la responsabilidad de la corporación se deriva de la responsabilidad individual de sus representantes. En estos casos, un demandante tendría que demostrar que hay un agente o grupo de agentes cuyas conductas erráticas pueden establecer la responsabilidad de la empresa. La responsabilidad de una empresa, por lo tanto, no representa una conducta indebida independiente, sino que surge debido a su relación legal con estos individuos.

La búsqueda del actor rebelde y la culpabilidad individual pone de manifiesto las complicaciones inherentes en la evaluación de la culpa. Como se mencionó anteriormente, en respuesta al argumento de que los actos de una máquina inteligente son impredecibles, sigue siendo un desafío establecer la intención individual. En consecuencia, sería una tarea difícil para los tribunales establecer la culpa y buscar justicia, lo que conduciría a la impunidad absoluta[77].

Este modelo, por ejemplo, podría ser particularmente factible dentro del escenario de *no hands*, cuando la máquina actúa de manera autónoma y el caso cae bajo el paraguas de la responsabilidad directa de la máquina y la cadena causal con la empresa/programador se interrumpe.

No hay duda de que, en el ejemplo del algoritmo escrito por quince programadores autónomos en un repositorio compartido de Git-Hub la cadena causal que conecta la IA con la empresa está interrumpida. Además, castigar a la máquina puede ser insatisfactorio y sin sentido. Es por eso que la responsabilidad corporativa indirecta con el mecanismo de *respondeat superior* llena los vacíos de responsabilidad. Si la responsabilidad corporativa es una ficción donde se crea subjetividad para abordar la responsabilidad, la relación entre el responsable (la corporación) y el actor (el empleado) puede aplicarse a la relación entre el usuario/supervisor y el algoritmo. Como ya hemos señalado, esta relación es más sólida que la ficción de subjetividad corporativa: la relación usuario-algoritmo se asemeja a la laboral, lo que la hace especialmente apta para albergar el modelo de *respondeat superior* como modelo residual de responsabilidad.

Para cumplir con el requisito de personalidad de la responsabilidad, la base de la relación de *respondeat* superior debe encontrarse dentro de la relación de facto *superior - inferior*. Por ejemplo: en el caso del robot de jardinería, el responsable sería el sujeto responsable de la seguridad de la actividad, que sería el encargado del jardín o el instalador/proveedor de los robots.

B) EL MODELO INDIVIDUALISTA

Un concepto novedoso de responsabilidad penal corporativa, basado en una comprensión realista de las corporaciones, sostiene que las corporaciones poseen personalidades e intenciones individuales independientes de las acciones de sus agentes. Esta perspectiva sobre la responsabilidad penal corporativa podría ofrecer apoyo normativo para los esfuerzos de reforma legal. En la literatura jurídica estadounidense, durante la última década, se ha propuesto un nuevo modelo de responsabilidad penal para entidades legales, basado en la

[77] OSMANI, N. op. cit., p. 71.

doctrina de auto-identidad y reflejando las características de las corporaciones modernas. Según este enfoque, la responsabilidad corporativa tiene prioridad y se basa en una evaluación de varios factores que examinan directamente el comportamiento de la corporación. Específicamente, evalúa lo que la corporación debería haber sabido y hecho para prevenir el daño[78].

Así, como observa una excelente doctrina, considerando el poder sin precedentes que poseen las corporaciones, un enfoque hacia la culpabilidad organizativa podría ser una herramienta poderosa para asignar responsabilidad por los riesgos asociados con las máquinas inteligentes[79].
Esta necesidad se siente cada vez más urgente ante el rápido desarrollo de la tecnología y la IA (y su necesidad de ser sometidas a un marco legal estricto[80]).

Cuando la responsabilidad relacionada con la IA carece de una orientación legal clara, se vuelve imperativo explorar soluciones legales alternativas.

Por ejemplo, en los casos residuales de responsabilidad del producto, y en los casos residuales de responsabilidad directa que caen en el espectro de problemas de "no hands" y "demasiadas manos", la responsabilidad de la empresa es la herramienta necesaria que, por un lado, llena el vacío de responsabilidad y, por otro lado, cumple con la protección necesaria del orden público. Esta doctrina, de esta manera, puede cumplir un doble objetivo: primero, puede mantener el orden social durante épocas de nuevas amenazas para la sociedad, al igual que la revolución digital hoy en día. En segundo lugar, prescinde de la necesidad de establecer la intención o la culpabilidad, centrándose, en cambio, en la asunción del riesgo que un actor lleva a cabo al participar en una actividad particular[81]. En este contexto, la responsabilidad se convierte en una carga para una entidad que, «*aunque por lo demás inocente, se encuentra en una relación responsable con un peligro público*[82]».

Esta doctrina, al igual que el modelo de responsabilidad corporativa indirecta, requiere leyes que destaquen la situación de facto y señalen quién es el responsable legal. En este caso, la ley de responsabilidad muestra la diligencia debida, reconduciendo esta disciplina a los principios constitucionales modernos (personalidad de la responsabilidad y reserva de ley según los artículos 25 y 27 de la Constitución Italiana, por ejemplo).

C) EVALUACIÓN DE LA RESPONSABILIDAD CORPORATIVA: SEIS PUNTOS

Para evaluar la viabilidad del modelo de responsabilidad corporativa, se deben evaluar seis puntos clave: «*(1) (poder) identificar qué corporación es responsable, (2) evitar oportunidades para maniobras, (3) proporcionar incentivos eficientes, (4) generar resultados justos, (5) ser fácil de implementar y (6) promover valores de programación*[83]».

El primer criterio, como se mencionó anteriormente, es poder «*identificar qué corporación es responsable*». El desarrollo algorítmico a menudo implica múltiples corporaciones, por lo que es crucial señalar qué corporación debería ser responsable. Esto incluye a aquellos que diseñaron, ensamblaron, probaron, comercializaron, poseyeron, licenciaron y operaron el

[78] LEDERMAN, E. *Models for Imposing Corporate Criminal Liability,* cit.p. 678.

[79] OSMANI, N.op. cit., pp. 72-73

[80] LEDERMAN, E. op. cit., p. 644.

[81] Esta solución fue principalmente contemplada por CARPENTER, C. L. *On Statutory Rape, Strict Liability, and the Public Welfare Offense Model*, in American University Law review, 53 (2), 2003, pp. 313–391, y N. Osmani, op. cit., pp. 72 e ss.

[82] *Morissette v. United States,* (1952) 342 U.S. 246.

[83] DIAMANTIS, M. E. *Employed algorithms*, cit., p. 829.

algoritmo. El modelo para responsabilizar a las corporaciones debe ofrecer un mecanismo claro para determinar la responsabilidad. El modelo laboral proporciona un método sistemático para identificar a las corporaciones responsables, centrándose en aquellas que emplean el algoritmo en cuestión. Esto implica un análisis intrincado de beneficios y control, alineándose con la realidad económica de las relaciones corporativas. Es importante destacar que el modelo laboral ofrece responsabilidad conjunta para todas las corporaciones involucradas, garantizando la responsabilidad tanto en contextos penales como civiles. También existen casos en los que recurrir a la responsabilidad corporativa sería innecesario o contraproducente, como en casos de malware desarrollado por individuos o sindicatos[84].

El segundo criterio es ser *«lo suficientemente robusto como para evitar maniobras»*. Para evitar maniobras estratégicas, el mecanismo para identificar a las corporaciones responsables debe ser resistente a la manipulación. Si existen lagunas en el marco de responsabilidad, por ejemplo, si la propiedad es el criterio, las corporaciones podrían transferir la propiedad estratégicamente para minimizar la responsabilidad.

Como destaca la mejor doctrina, *«la principal ventaja del Modelo Laboral es que evita los criterios formalistas a favor de pruebas funcionales que siguen la "realidad económica" en lugar de las apariencias superficiales[85]».*

En contraste con los modelos de Responsabilidad Estricta y Causa Natural-Probable, el Modelo Laboral evita abrir vías para que las corporaciones eviten la responsabilidad a través de estrategias creativas. Al emplear pruebas funcionales fundamentadas en la realidad económica, el Modelo Laboral aborda eficazmente la naturaleza dinámica y flexible de las corporaciones. Su enfoque en la sustancia sobre las apariencias superficiales minimiza el riesgo de maniobras, diferenciándolo de otros modelos[86].

El tercer criterio es proporcionar incentivos eficientes a todas las empresas involucradas. El equilibrio es crucial: mientras que imponer una responsabilidad inadecuada no incentiva el desarrollo algorítmico responsable, una responsabilidad excesiva podría frenar la innovación al hacer que los algoritmos sean demasiado costosos. El modelo debe encontrar un equilibrio que fomente el comportamiento corporativo responsable sin obstaculizar el progreso tecnológico[87].

El Modelo Laboral aprovecha los marcos de responsabilidad corporativa existentes, promoviendo la eficiencia al equilibrar las responsabilidades entre las víctimas y los infractores. Al incentivar a las corporaciones a invertir en cumplimiento para los sistemas algorítmicos, el modelo refleja el enfoque adoptado con los empleados humanos. Esto incluye fomentar un mejor cumplimiento a través de equipos de ingeniería diversos, programación cuidadosa, pruebas extensas y actualizaciones continuas, alineando los intereses corporativos con el objetivo de minimizar los daños algorítmicos.

El cuarto criterio es producir resultados justos. La justicia requiere una asignación justa de responsabilidad entre las corporaciones y las víctimas. Dejar que las víctimas soporten los costos de los daños algorítmicos es inequitativo, pero imponer una responsabilidad excesiva a las corporaciones también podría ser injusto. El modelo debe garantizar la equidad tanto

[84] DARYL J. L., *Collective Sanctions*, 56 Stanford Law Review, 345, 393 (2003) ("Firms may externalize liability costs by spinning off risky operations into undercapitalized subsidiaries, as when owners of taxi enterprises incorporate each cab separately").
[85] DIAMANTIS, M. E., *ibid.;* Martin v. Sprint United Mgmt. Co., 273 F. Supp. 3d 404, 422 (S.D.N.Y. 2017) ("En cuanto a la prueba de control funcional, el Segundo Circuito ha identificado una serie de factores pertinentes para determinar si una persona o entidad, incluso si carece de control formal, ejerció un 'control funcional' sobre un empleado").
[86] DIAMANTIS, M. E., *Employed algorithms*, cit., p. 851.
[87] POSNER R. A. y. LANDES, W.M The Positive Economic Theory of Tort Law, 15 1980, GA. L. REV. 851, 868 y ss.

para las víctimas como para las corporaciones, considerando los beneficios y costos sociales más amplios[88].

Alineándose con la ley existente de responsabilidad corporativa, el modelo laboral garantiza la equidad al tratar los daños algorítmicos de manera similar a los causados por empleados humanos. La doctrina observa que: *"Al tratar los daños algorítmicos como los cometidos por empleados, el Modelo Laboral actualiza la ley actual para dar a los demandantes y fiscales el mismo camino razonable hacia la satisfacción por la mala conducta algorítmica que tienen actualmente para los daños causados por empleados. Al mismo tiempo, dado que el Modelo Laboral solo empareja las responsabilidades corporativas anticipadas con los beneficios corporativos anticipados, es justo para los interesados corporativos. El riesgo de pérdida que enfrentan los interesados corporativos por la responsabilidad de los daños algorítmicos se asemeja al riesgo comercial genérico que acompaña a cualquier empresa lucrativa[89]"*.

El quinto criterio es tener barreras bajas para la implementación. La viabilidad de la implementación es esencial, y el modelo debe evitar reformas disruptivas. Los cambios incrementales obtienen un apoyo más amplio y tienen más probabilidades de tener éxito[90]. La simplicidad y la compatibilidad con los marcos legales existentes mejoran las posibilidades de adopción del modelo..

Las virtudes pragmáticas del modelo laboral se extienden a su facilidad de implementación. Basándose en principios legales existentes, el modelo podría integrarse sin problemas en el panorama actual de responsabilidad corporativa sin requerir cambios legislativos radicales. Su potencial para menores barreras políticas en comparación con otros modelos alternativos mejora su atractivo como una solución viable[91].

El sexto criterio es promover valores de programación que se alineen con el desarrollo algorítmico responsable. La transparencia, el respeto por la autonomía humana y la protección de la privacidad son valores esenciales para guiar a los programadores. La responsabilidad corporativa debería servir como una herramienta para fomentar estos valores en el desarrollo algorítmico[92].

Reconociendo la importancia de los valores de programación, especialmente la transparencia, el Modelo Laboral incentiva a las corporaciones a encontrar un equilibrio entre transparencia y precisión. A diferencia de los enfoques de talla única, el modelo permite a las corporaciones evaluar la deseabilidad de la transparencia caso por caso. Al alentar a las corporaciones a alinear las evaluaciones impulsadas por el lucro con resultados socialmente deseables, el modelo laboral promueve eficazmente los valores de programación[93].

CONCLUSIONES

En conclusión, a medida que avanza la tecnología de IA, la cuestión de la responsabilidad penal relacionada con la IA sigue siendo compleja y multifacética.

---

[88] . ALSCHULER, A. W *Two Ways To Think About the Punishment of Corporations,* American Criminal Law Review, 46 2009, pp. 1359, 1366–67.

[89] DIAMANTIS, M. E. *Employed algorithms*, cit., p. 854.

[90] CALABRESI, G. *A Common Law for the Age of Statutes*, 1982.

[91] DIAMANTIS, M. E. *Employed algorithms*, cit., p. 855.

[92] European Commission: High-Level Expert Group on Artificial Intelligence publishes Ethics Guidelines for Trustworthy AI 2–3.

[93] DIAMANTIS, M. E. *Employed algorithms*, cit., p. 855 e ss.

En primer lugar, los tres modelos de responsabilidad pueden utilizarse para construir un primer marco de responsabilidad en torno a la máquina.

En segundo lugar, la idea de responsabilidad penal corporativa, basada en la identidad independiente de una organización, presenta una vía convincente para abordar las brechas de responsabilidad derivadas de la aplicación solo del marco proporcionado por los tres modelos.

Encontrar el equilibrio adecuado entre el progreso tecnológico y la gestión de los riesgos potenciales será un desafío importante para los responsables políticos y la sociedad en su conjunto. El debate en curso subraya la necesidad de continuar explorando y desarrollando marcos legales que puedan adaptarse al panorama en evolución de la IA.