DERECHO PENAL ADAPTADO A LA IA: ENTRE EL DERECHO PENAL ANALÓGICO Y EL DERECHO PENAL DIGITAL

Raquel Roso CAÑADILLAS*

Fecha de recepción: 13 de octubre de 2024
Fecha de aprobación: 10 de febrero de 2025

Resumen: La elaboración de este trabajo ha venido impulsada por la irrupción de la IA en el desarrollo de la humanidad, por la publicación del Reglamento de IA (en adelante RIA), con su entrada en vigor escalonada y por aportar algunas reflexiones sobre la cuestión de cuáles serán las implicaciones de este profundo cambio tecnológico, concretamente en el Derecho penal, transitando de un Derecho penal analógico a un Derecho penal digital. El trabajo se divide en dos partes. En la primera se analiza sintéticamente el Reglamento de IA y se extraen una serie de notas y características relevantes de esta norma, que permiten conectar con la segunda parte del trabajo, en la que se hace un repaso sintético a los elementos de la teoría del delito para evaluar su afectación y su necesidad de adaptación a la existencia disruptiva de la IA, con la finalidad de ir construyendo un Derecho penal digital sólido, cumplidor de su función preventiva y de su función de garantía de bienes jurídicos, en particular, de derechos fundamentales, contribuyendo a diseñar un sistema ponderado de responsabilidad jurídica integral, que vele por la transparencia y explicabilidad del algoritmo y por la supervisión humana.

Palabras clave: Inteligencia artificial – reglamento de inteligencia artificial – Derecho penal sólido – teoría del delito – principio de supervisión humana – riesgo permitido – derechos humanos – causalidad probabilística – opacidad epistémica – previsibilidad – garantes de la cadena de valor – personalidad jurídica de la persona electrónica – programas de cumplimiento.

* Profesora Titular de Derecho penal de la Universidad de Alcalá (Madrid). Contacto: raquel.roso@uah.es. Este trabajo ha sido elaborado durante una estancia de investigación en la Universidad de Würzburg (Alemania), gracias a la beca que me ha concedido el servicio de intercambio alemán (DAAD) y a la generosa invitación y admisión en su cátedra del Prof. Hilgendorf. También quiero dejar constancia de mi agradecimiento a José Béguelin por la concienzuda revisión del trabajo y sus valiosas y enriquecedoras observaciones, a Joaquín Oliva González, ingeniero informático, por sus orientaciones técnicas y a los evaluadores por sus recomendaciones.

Title: Brief notes on the AI Regulation and some reflections on a future AI-adapted criminal law: between analogue and digital criminal law

Abstract: The preparation of this paper has been prompted by the irruption of AI in the development of humanity, by the publication of the AI Act, with its staggered entry into force, and by reflecting on the question of what the implications of this profound technological change will be specifically in criminal law, moving from an analogue criminal law to a digital criminal law. The work is divided into two parts. In the first, the AI Act is analysed synthetically and a series of relevant notes and characteristics of this regulation are extracted, which allow a connection to be made with the second part of the work, in which a review is made of the elements of the theory of crime in order to evaluate its impact and its need to adapt to the existence of AI, with the aim of building a hard digital criminal law, fulfilling its preventive function and its function of guaranteeing legal assets, in particular, fundamental rights, forming part of a system of integral legal responsibility.

Keyboards: Artificial intelligence – AI Act – hard criminal law – crime theory – human in the loop (HILP) – legal risk – human rights – probabilistic causation – epistemic opacity – predictability – value chain guarantors – legal personality of electronic person – compliance programmes.

I. Introducción

El año pasado por estas mismas fechas me hacía la pregunta siguiente: ¿Un Derecho penal delicuescente en una sociedad líquida?, la cual dio lugar a la publicación de un trabajo del mismo nombre[1] y en el que vertía algunas reflexiones sobre la deriva de la sociedad actual y las inflexiones de esta en el Derecho y en particular en el Derecho penal y, en un pequeño apartado, en el Derecho

[1] Roso Cañadillas, "¿Un Derecho penal delicuescente en una sociedad líquida?", en *Anuario de la Facultad de Derecho de la Universidad de Alcalá*, n.º 26, 2023, pp. 199-225; La Misma, "¿Un Derecho penal delicuescente en una sociedad líquida? Algunas reflexiones sobre el Derecho penal en la sociedad posindustrial", en *Revista General de Derecho penal*, n.º 41, 2024, pp. 1-30.

penal del trabajo. El planteamiento consistía en enmarcar la situación y el "predecir" el papel del Derecho penal en una sociedad cada vez más líquida, como la bautizó Baumann[2], con cambios tecnológicos que se producen a un ritmo vertiginoso. Concluía en aquel trabajo que el Derecho penal constituiría un bastión de solidez dentro de la incertidumbre que experimentamos. Espero que así sea y se convierta en una herramienta que nos ofrezca el asidero a tierra firme ante tanto avance sin tiempo razonable de asimilación para la mente humana.

Estamos inmersos en pleno proceso de la cuarta revolución industrial con el desarrollo imparable del algoritmo y la IA[3], que está abriendo, me atrevo a predecir, la puerta a la quinta revolución, que ya no sé si denominarla industrial o revolución neurocientífica[4], quizás la última por lo trascendente que puede llegar a ser, incluso para la existencia del ser humano. Con el desarrollo de la IA se conseguirá, según dicen los expertos, un beneficio social global mejorando la asistencia sanitaria, un transporte más seguro y limpio, servicios públicos inteligentes y mejorados, productos y servicios innovadores, sobre todo en el ámbito de la energía y sanidad. La esfera empresarial se beneficiará de una elevada productividad y una fabricación más eficiente. También mejorará la gobernanza del Estado en espacios como el transporte, la energía y la gestión de residuos. Incluso parece que la IA nos

---

[2] BAUMANN, *Modernidad líquida*, Fondo de Cultura económica, 2002, p. 16.

[3] Se habla de una cuarta revolución industrial, que, en mi opinión, se solapa con la anterior referida a la transformación digital, a la sociedad de la información y a las energías renovables entre otras cosas. Esta cuarta revolución industrial, término acuñado por Klaus Schwab con un libro homónimo, también conocida como Industria 4.0, termino importado de la literatura alemana *Industrie 4.0*, se caracteriza según este autor por la fusión de tecnologías y la desintegración de fronteras entre lo físico, digital y biológico, en la que se incluye la robótica, IA, nanotecnología, computación cuántica, biotecnología, internet de las cosas, impresión 3D y vehículos autónomos (SCHWAB, *La cuarta revolución industrial*, Madrid, Marcial Pons, 2016).

[4] Esta revolución no es industrial, no es sobre la máquina, va mucho más allá, se da un gran salto cualitativo siendo el centro el estudio del funcionamiento del propio cerebro humano, propósito al que se ha dedicado y se dedica el neurocientífico español Rafael Yuste, ideólogo del Proyecto *Brain* de la Administración Obama, con el que pretende alcanzar la cura de enfermedades mentales y mejorar la especie humana. Esta revolución nos adentraría en una especie de transhumanismo, con seres híbridos, como sostiene el neurocientífico español (cfr. Disponible en: https://www.eldiario.es/cantabria/rafael-yuste-neurobiologo-abocados-futuro-seres-humanos-hibridos_1_11481857.html, [Enlace verificado 14-10-2024], parece que bondadoso, cuya evolución puede discurrir por dos vías: humanos híbridos con chip, dirigidos a mejorar su calidad de vida previendo enfermedades, extirpando malos hábitos, mejorando su estabilidad mental, sin en ningún caso manipular, es decir, sacando la mejor versión de cada ser humano; la otra vía de desarrollo a las que nos conducirá la neurociencia es pasar de los asistentes inteligentes como *Google Maps, Google Lens, Alexa*, etc., a la inteligencia artificial general, una clase de superinteligencia que será capaz de programarse a sí misma y que superará al ser humano, cuando sean socialmente inteligentes, entiendan las emociones y se conecten unas a otras. Estos momentos serán vistos desde una perspectiva histórica futura como un periodo de transición.

adentrará en el paraíso del 'no-trabajo' pesado: del *homo faber* al *homo digitalis*. No obstante, el paraíso también tiene sus peligros.[5]

En este estado de felicidad prometida y de avance de la humanidad, según la mirada de algunos, en el que se puede esconder, por el contrario, una consecución vertiginosa disolutiva de lo sólido, de los anclajes de la humanidad a la realidad, por este proceso de digitalización imparable, el cual escala ya a un nivel superior con software inteligentes capaces de tomar decisiones, se enmarca el Reglamento Europeo de Inteligencia Artificial[6] (en adelante RIA[7]), que ha entrado en vigor el 1 de agosto de 2024.[8] Actualmente, no hay ninguna norma de estas características en el mundo, que regule de manera armonizada la implantación y uso de la IA. La pionera UE pretende, con ello, ser el líder mundial y ambiciona convertirse en el estándar normativo global con el desarrollo de tres conceptos claves: riesgo, garantías y responsabilidades. A la luz de un resumido análisis de esta norma europea y de constatar los principios de actuación que se contienen en la misma, surge la pregunta de cuál debe ser la contribución del Derecho penal, un Derecho penal internacionalmente armoni-

[5] La propia IA, en este caso Chat GPT (consultada 28 de septiembre de 2024), ha contestado a la pregunta de qué riesgos traerá a la humanidad el uso de la IA lo siguiente: "¡Hola! Es un tema muy interesante y relevante. La inteligencia artificial (IA) puede traer muchos beneficios, pero también presenta ciertos riesgos para la humanidad. Algunos de los principales riesgos incluyen: 1. Desempleo: A medida que la IA automatiza tareas, puede llevar a la pérdida de empleos en ciertos sectores, lo que podría generar desigualdad económica. 2. Privacidad: La recopilación masiva de datos por parte de sistemas de IA puede poner en riesgo la privacidad de las personas, ya que se pueden utilizar para monitorear y analizar comportamientos. 3. Sesgo y discriminación: Si los algoritmos de IA se entrenan con datos sesgados, pueden perpetuar o incluso amplificar prejuicios existentes, afectando a grupos vulnerables. 4. Seguridad: La IA puede ser utilizada en ciberataques o en la creación de armas autónomas, lo que plantea serias preocupaciones sobre la seguridad global. 5. Desinformación: La capacidad de generar contenido falso o manipulado puede dificultar la distinción entre la verdad y la ficción, afectando la opinión pública y la democracia. 6. Dependencia: A medida que nos volvemos más dependientes de la IA, existe el riesgo de que perdamos habilidades humanas esenciales y la capacidad de tomar decisiones críticas.

[6] Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024.

[7] Act AI en inglés.

[8] Aprobado 13 de marzo 2024, publicado en el Diario Oficial de la Unión Europea el 12 de julio de 2024, entrada en vigor el 1 de agosto de 2024 y será aplicable con carácter general a los veinticuatro meses de su entrada en vigor, con excepciones importantes (art. 113 RIA): 1) el capítulo I (Disposiciones generales) y II (Prácticas prohibidas) se aplicará a los seis meses de su entrada en vigor; 2) La Sección IV del capítulo III (Autoridades notificantes y organismos notificados), capítulo V (Modelos de IA de uso general), capítulo VII (Gobernanza), capítulo XII (Sanciones), y art. 78 del capítulo IX serán aplicables a los doce meses de su entrada en vigor; 3) el art. 101 del capítulo XII y el art. 6.1, sobre reglas de clasificación de los sistemas de IA de alto riesgo y las obligaciones relativas al Reglamento serán aplicables a los 36 meses de su entrada en vigor.

zado, para contribuir a la implementación y uso responsable de la IA. La cuestión planteada ha derivado en este trabajo en un recorrido por la teoría del delito apuntando posibles adaptaciones y ajustes en las categorías jurídico-penales, para avanzar en el diseño de un Derecho penal digital.

II. La especialidad de la norma y su contexto[9], nunca experimentado por la humanidad

El RIA es una norma de gran *complejidad*[10], con grandes dosis de indeterminación impuesta por el contexto[11], lo que los juristas bautizamos como conceptos jurídicos indeterminados, y me atre-

---

[9] El contexto es una nueva revolución y la humanidad ya ha experimentado este tipo de fenómenos: cambios tecnológicos que han conducido a cambios sociales y ello a cambios económicos. Las revoluciones han rediseñado el modo de vida de la humanidad, han generado transformaciones profundas, dejando huellas indelebles en cada generación e impulsando cambios trascendentales. Hemos pasado de la agricultura, a la industria, de la industria al sector de servicios, para seguir en este sector con empresas deslocalizadas, con robótica avanzada y con una política económica de desarrollo global. Se está produciendo lo que R. Baldwin denomina la convulsión globótica: entre la globalización y la robótica. Sin embargo, esta revolución se me antoja distinta por dos motivos: el primero es que todos los avances tecnológicos anteriores sustituían al hombre en labores manuales, muchas veces pesadas o repetitivas, que se identificaban con el perfil de un operario de la era industrial y en todo momento se controlaban por las personas; pero esta tecnología va a sustituir o incluso reemplazar al hombre en tareas intelectivas, privativas hasta la fecha solo del ser humano, pudiendo los sistemas de IA a través del entrenamiento tomar decisiones, generar contenido, analizar millones de datos, resolver problemas y, pese a que todavía no han alcanzado el nivel de la emoción o el *nous* de los griegos, pueden escapar al control del ser humano y ser impredecible la solución de salida arrojada por el sistema, escapando, por ello, a su control; el segundo de los motivos, viene dado por la velocidad a la que se está produciendo esta revolución frente al resto de las que ha experimentado la sociedad. El tiempo de reacción y de adaptación del sistema y del individuo va a ser en algunos casos inexistente. Se destruirán, p. ej. puestos de trabajo sin que los trabajadores hayan podido reformarse, reciclarse y reubicarse. Ante esta desorientación, miedo y una profunda sensación de injusticia se generará un riesgo destructivo, como podría ser una reacción masiva violenta de la sociedad a gran escala, lo que llevará a una quiebra profunda del orden social. (Cfr. BALDWIN, *La convulsión globótica. Globalización, robótica y el futuro del trabajo*, Barcelona, Bosch, 2019). Esta antes vendrá precedida, debido a ese olor que despide el miedo, de movimientos populistas y nacionalistas, como está ocurriendo ya en Europa. Así, mientras escribo estas páginas, Austria ha hablado y ha votado al FPÖ (*Freiheitliche Partei Österreichs*), partido calificado como de ultraderecha y nacionalista. (cfr. Disponible en: https://www.dw.com/es/extrema-derecha-obtiene-hist%C3%B3rica-victoria-en-las-elecciones-legislativas-en-austria/a-70358144, [Enlace verificado 29-09-2024]. Cfr. tb. en este sentido, QUINTERO OLIVARES, "Populismo y Derecho penal", en *DOXA*, n.º 48, 2024, pp. 281 ss.

[10] Está estructurado en 180 considerandos, 113 artículos y 13 anexos.

[11] No sabemos con exactitud hacia donde nos dirigimos, pero cuando una tecnología avanza no se puede parar. Se comporta como una fuerza centrípeta que nos atrae y que no podemos controlar, pese a los esfuerzos para hacerlo. Nos encontramos en pleno dilema de *Collingridge*, que nos dibuja dos pendientes problemáticas: la primera es que existe un problema de información, porque no se sabe cuáles serán las consecuencias, no se pueden predecir tan fácilmente, hasta que la tecnología no esté ampliamente desarrollada y utilizada, y por otro lado existe un problema de poder, porque el control y el cambio se tornan difíciles cuando la tecnología esta arraigada. Sobre esta última pendiente que nos indica el sociólogo británico, solo levante la vista y mire a su alrededor, si está en un lugar público mientras lee esta nota a pie, y observe cuántas personas están usando su *smartphone* e imagine qué ocurriría si se les exigiese la entrega de sus dispositivos.

vería a decir que *experimental* y con un marcado carácter de *provisionalidad* en sus contenidos y definiciones, que conduce a una caducidad regulativa, que generará la obligación continua de revisión. Ello se debe al ritmo trepidante del progreso de esta tecnología[12], que deja obsoletas en un espacio corto de tiempo tanto definiciones como regulación, al presentarse nuevos casos. Estamos en fase de experimentación mundial y la normativa no se escapa a este fenómeno de perentoriedad, que está superando por primera vez en su historia la capacidad de adaptación de la humanidad a los cambios.[13]

El objetivo más deseable, más acuciante e importante en este contexto es alcanzar unos *mínimos de seguridad* en el desarrollo de la IA y una *implementación responsable*. Para su consecución, el RIA toma como centro neurálgico al ser humano y desde ahí conforma una *IA antropocéntrica*, que pueda ser digna de confianza, asegurando un alto grado de protección de los derechos fundamentales, la salud y seguridad de los ciudadanos europeos; no obstante, también persigue un objetivo económico y práctico, como es el mejorar el funcionamiento del mercado interior para implantar de modo seguro todas las innovaciones que vayan apareciendo en el desarrollo de esta tecnología, ofreciendo un marco jurídico uniforme.[14]

Es una *norma de carácter anticipativo*, que no ha esperado a que el fenómeno se desarrolle y madure para establecer una regulación definida, tal y como suele ser la conducta conservadora de los legisladores. Asistimos, por el contrario, a un sorprendente cambio de dirección, y la Comisión Europea ha preferido presentar un marco regulativo inacabado y sometido previsiblemente a un

---

[12] Según predijo Gordon Moore en 1965 el número de componentes integrados en un chip se duplicaría cada dos años (Ley *Moore*), e incluso a veces cada menos tiempo. Un chip tiene actualmente decenas de miles de componentes lo que ha dado lugar a mayor capacidad de procesamiento y es imparable su evolución, abandonando ya el terreno de la microtecnología y avanzando por la nanotecnología hacia escaladas próximas al átomo. Este contexto es el que está detrás del avance imparable de la IA y de los cambios vertiginosos, que a su vez van a provocar una revisión constante de la normativa durante un largo periodo de tiempo. Un ejemplo muy ilustrativo ha sido y sigue siendo la introducción de reglas específicas para los modelos fundacionales, como son las plataformas de *ChatGPT* y otras, que surgieron una vez que la Comisión Europea presentara su primera propuesta del RIA, por lo que el capítulo dedicado a estos sistemas se ha ido desarrollando a la par que el curso de la negociación del RIA. Con ello se evidencia el carácter excepcional e insólito de esta normativa. En cuanto a los modelos fundacionales según *OpenAI* su potencia de cómputo se ha venido duplicando cada 3,4 meses desde 2012. Estos modelos son el claro exponente de la IA generativa, citada una sola vez en el RIA en el considerando 99 para afirmar que son un ejemplo típico de un sistema de IA de uso general, la cual maneja millones de datos, puede realizar una amplia gama de tareas dispares con alto grado de precisión, genera contenidos, analiza patrones, clasifica imágenes….y suelen integrarse en un sistema.

[13] SCHWAB, *supra* nota 3.

[14] En el Considerando primero y art. 1 del RIA se describe el objetivo del Reglamento, que junto a los ya referidos en el texto también hay que añadir la protección de la democracia, el Estado de Derecho y del medio ambiente.

continuo trabajo de reforma y adecuación[15], pero uniforme para todos los países de la UE, con el fin de ofrecer en esta fase inicial principios básicos de actuación normados y obligaciones para los usos específicos de la IA, que deben cumplir una cadena de sujetos activos, entre los que se encuentran los desarrolladores e implementadores del sistema de IA, definiendo los posibles riesgos para la salud, la seguridad y los derechos fundamentales de la persona. El mérito y el esfuerzo de los padres de este RIA es de alabar, dejando a un lado las mejoras de las que siempre es susceptible una regulación y más en este escenario, en el que la norma pretende ir al paso desconocido de la tecnología, lo que vuelve a ser algo extraordinario.

III. El concepto de riesgo y su clasificación como columna vertebral del RIA

Son innumerables y casi inabarcables los aspectos y cuestiones que se derivan de un análisis profundo y sosegado del RIA. En estas páginas solo destacaré algunos mínimos aspectos a modo de presentación sintética y orientativa, y que pretenden invitar a iniciarse en su lectura, estudio y reflexión por los operadores jurídicos de todas las disciplinas[16], incluidos los penalistas. El primero de

[15] Recientemente, la Comisión, después de la publicación del RIA, ha puesto en marcha varias consultas sobre el uso y aplicación de la IA, por ejemplo, en el sector financiero, y con carácter general el art. 56 RIA regula el Código de Buenas Prácticas para los proveedores de modelos de inteligencia artificial de uso general (GPAI). Esto códigos de buenas prácticas se fomentarán por la Oficina de la IA, la cual velará por que se recojan los objetivos específicos de usos, así como compromisos y medidas adecuadas de implementación. Este Código abordará ámbitos críticos como la transparencia, las normas relacionadas con los derechos de autor y la gestión de riesgos. Se invita a los proveedores de los modelos de inteligencia artificial de uso general que operan en la UE, a las empresas, a los representantes de la sociedad civil, a los titulares de derechos y a los expertos académicos a que presenten sus puntos de vista y conclusiones, que se incorporarán al próximo proyecto de Código de Buenas Prácticas de la Comisión sobre modelos de inteligencia artificial de uso general. En definitiva, se busca la participación de las múltiples partes interesadas para alcanzar un amplio consenso y una adecuación constante de la regulación a través de la información y propuesta de los sectores implicados. No debemos olvidar que la entrada en vigor del RIA se presenta de forma escalonada, de tal modo que las disposiciones relativas a los GPIA entrarán en vigor a los doce meses de la publicación del Reglamento. Por su parte, los Considerandos 150 y 151 RIA pretende potenciar la ejecución y aplicación del Reglamento con la creación de un foro consultivo para asesorar al Consejo de IA y a la Comisión y con la creación de un grupo de expertos científicos que apoye la Oficina de IA.

[16] En esta línea interdisciplinar se encuadra la nueva revista científica alemana mensual sobre IA, denominada KIR: *Künstliche Intelligenz und Recht*, (inteligencia artificial y Derecho) cuyo primer número acaba de salir en agosto de este año (cfr. Disponible en: https://rsw.beck.de/zeitschriften/kir/startseite, [Enlace verificado 28-9-2024], con artículos en este primer número sobre el Derecho de la IA, el nuevo marco jurídico de la IA en la UE, IA generativa, modelos fundacionales y modelos de IA de aplicación general en el RIA u orientaciones extraídas de la conferencia de las autoridades independientes competentes para la protección de datos para la IA y la protección de datos. Además, también incluye una sección de jurisprudencia.

ellos es el concepto de IA[17], nada fácil de describir, debido al avance meteórico, que debe ser contenido, como explico más adelante, de la ciencia y la tecnología. En el art. 3 RIA se recogen definiciones y en su apartado primero se ofrece un concepto de sistema de IA: "un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales". Con esta definición se ha intentado comprender todas las formas de IA hasta la fecha conocidas desde su perspectiva técnica[18] y esta descripción la convierte en el objeto de la regulación, la cual tiene como objetivo primordial la consecución del buen uso, del uso virtuoso, del uso ético, en definitiva, del uso responsable de estos sistemas.

Lo interesante comienza cuando nos introducimos en el concepto que se me antoja definir como la columna vertebral de esta regulación. La metáfora se refiere al concepto de riesgo, como fulcro conceptual de todo el Reglamento, y a la estratificación de riesgos contenida en la normativa, definiendo simultáneamente, como la cabeza de Jano, dónde empieza el riesgo permitido[19] y dónde acaba el riesgo no permitido, sirviéndose de la definición de los límites, convertidos en obligaciones-deberes, que no se pueden traspasar. El concepto de riesgo se encuentra en el apartado segundo del art. 3 RIA y define el riesgo como "la combinación de la probabilidad de que se produzca un perjuicio y la gravedad de dicho perjuicio". En el art. 5 RIA se encuentran las prácticas de la IA prohibidas, que se catalogan como *riesgos inadmisibles*, inaceptables e insoportables. Claramente quedarían fuera del riesgo permitido por suponer una amenaza directa a los derechos fundamentales y a la seguridad y privacidad del individuo. Ejemplo de ello son los sistemas de identificación biométrica remota en

---

[17] Sobre el con concepto de IA cfr. IBOLD, *Künstliche Intelligenz und Strafrecht. Zur strafrechtliche Produktverantwortung in der Innovationsgesellschaft*, Baden-Baden, Nomos, 2024, pp. 155 y ss., donde presenta los distintos enfoques sobre la definición de IA, sus características fundamentales y repara en el efecto "caja negra" del aprendizaje automático, señalando su importancia para la valoración legal. Tb. HILGENDORF, "Inteligencia artificial y Derecho penal", en *Desafíos penales de hoy: Entre la ley y la justicia en la obra de Eric Hilgendorf,* Buenos Aires, Editores del Sur, 2024, pp. 17 ss.

[18] Depende a su vez del criterio de clasificación, porque abarca una gran variedad de tecnologías que tienen como objetivo común emular la inteligencia humana. Cfr. cualquier página de internet al efecto como p. ej. https://www.tableau.com/es-mx/data-insights/ai/tipos-de-inteligencia-artificial. A ello hay que añadir que la IA ha salido del ordenador y se ha integrado en el internet de las cosas o en robots, incluso con forma humana.

[19] Un riesgo permitido adecuado a la era de la digitalización y a una sociedad líquida y panóptica, sometida a riesgos disruptivos, Cfr. BAUMANN, *supra* nota 2, p. 16.

tiempo real, aunque también existen excepciones tasadas a estos sistemas de IA[20], o el reconocimiento de emociones, sistemas de puntuación social o sistemas que manipulen el comportamiento humano.

En el art. 6 se regulan los *sistemas de alto riesgo* (HRAIS, *High risk artificial intelligence systems*)[21], los cuales pueden tener un impacto relevante en los derechos fundamentales de las personas y en su seguridad y salud e incluso en el medioambiente y la democracia. Así p. ej. el Anexo III se refiere al uso de estos sistemas en las infraestructuras críticas, la educación, la formación, el empleo, los servicios públicos y privados esenciales, como pueden ser la sanidad o la banca, determinados sistemas de las fuerzas de seguridad, la migración o la gestión aduanera, la justicia y los procesos democráticos, que se hará acompañar antes del 2 de febrero de 2026 de "una lista exhaustiva de ejemplos prácticos de casos de uso de sistemas de IA que sean de alto riesgo y que no sean de alto riesgo" (art. 6.5 RIA). Este uso de sistemas de IA de alto riesgo entra plenamente dentro de la balanza estructural que conforma el riesgo permitido[22]: la IA representa en su conjunto más beneficios que perjuicios, por lo que la balanza se desliza hacia la implantación y desarrollo seguro de sistemas de alto riesgo, siempre que cumpla los requisitos exigidos por el RIA, requisitos que adoptan la forma de obligaciones para los proveedores y otros sujetos activos[23] que intervienen en la cadena de valor.

De la lectura de los arts. 9 ss. del RIA se puede deducir que los requisitos y obligaciones recogidos el RIA son garantías y dibujan las líneas maestras de un programa de cumplimiento, del que

---

[20] Estas excepciones permitirán el uso de sistemas de vigilancia biométrica en espacios públicos, solicitada y acordada la orden judicial y para una lista de delitos tasados recogida en el Anexo II del RIA.

[21] Estos sistemas se dividen en dos grupos. Por un lado, se encuentran los sistemas vinculados a la legislación armonizada sobre seguridad de productos incluida en el Anexo I del RIA y, por otro lado, los sistemas incluidos en el Anexo III. Desarrollando el primer grupo, los sistemas de IA serán de alto riesgo cuando sean o proporcionen características de seguridad para productos sujetos a la normativa de armonización de la UE o cuando tengan por objeto proporcionar características de seguridad para infraestructuras críticas o cuando supongan un riesgo significativo de daño para la salud, la seguridad o los derechos fundamentales de las personas físicas. El Anexo III consiste en una enumeración de sistemas que se consideran de alto riesgo: biometría, infraestructuras críticas, educación y formación profesional, empleo, gestión de trabajadores y acceso al autoempleo, acceso y disfrute de servicios privados esenciales y servicios y prestaciones públicos esenciales, garantía del cumplimiento del Derecho, migración, asilo y gestión del control fronterizo, administración de justicia y procesos democráticos.

[22] Sobre la figura jurídica del riesgo permitido en el contexto de la regulación de la IA, cfr. IBOLD, *supra* nota 17, pp. 308 ss.

[23] Proveedores, responsables del despliegue, importadores y distribuidores, fabricantes de productos y los representantes autorizados de los proveedores no establecidos en la UE.

deberá partir cualquier empresa para el desarrollo y adaptación de su actividad cuando utilice sistemas de IA y que completará con normas internas o códigos de buenas prácticas[24], volviendo a hacer uso en este campo de la socorrida autorregulación, como forma en el fondo de una suerte de delegación legislativa en las corporaciones e instituciones, y que requerirá de una alta especialización, debido a la complejidad de la norma y de su objeto. La irrupción de la IA se hará acreedora de una regulación propia y extensa, que tendrá que asumir todas las especificidades de este fenómeno, que conlleva cambios disruptivos en la estructura social en todas las áreas: cultural, social, política, económica, laboral, educativa, sanitaria…

Siguiendo con la estratificación del riesgo contenida en el RIA, en el siguiente escalón nos encontramos con los sistemas, mejor denominarlos *modelos*[25] *de IA de riesgo limitado que se identifican con los sistemas de propósito o de uso general o GPAIS* (*General-Purpose Artificial Intelligence System*), arts. 51 ss. RIA, los cuales son sistemas avanzados de IA capaces de realizar de modo eficaz tareas distintas, pero que no tienen un propósito previsto inicial. En estos se incluyen los sistemas de IA asociados a *chatbots*, lo que obliga a informar a los usuarios de que están interactuando con una máquina, para que puedan decidir sobre si seguir con esta interactuación y ponderar la información recibida. Los proveedores deben asegurar también que el contenido generado por la IA sea identificable y etiquetarse como generado artificialmente en el caso de que sea publicado con el propósito de informar al público sobre asuntos de interés público. Estas garantías se aplican a contenidos de audio y video. Las obligaciones de los proveedores vienen reguladas en los arts. 53 ss. del RIA.

---

[24] Así ya lo contempla el propio RIA en su art. 56 cuando regula que la Oficina de AI fomentará y facilitará la elaboración de códigos de buenas prácticas a escala de la Unión a fin de contribuir a la correcta aplicación del presente Reglamento, teniendo en cuenta planteamientos internacionales. Estos códigos estarán finalizados a más tardar el 2 de mayo de 2025.

[25] El Considerando 97 del RIA aclara que "el concepto de modelos de IA de uso general debe definirse claramente y diferenciarse del concepto de sistemas de IA con el fin de garantizar la seguridad jurídica. La definición debe basarse en las características funcionales esenciales de un modelo de IA de uso general, en particular la generalidad y la capacidad de realizar de manera competente una amplia variedad de tareas diferenciadas. Estos modelos suelen entrenarse usando grandes volúmenes de datos y a través de diversos métodos, como el aprendizaje autosupervisado, no supervisado o por refuerzo. Los modelos de IA de uso general pueden introducirse en el mercado de diversas maneras, por ejemplo, a través de bibliotecas, interfaces de programación de aplicaciones (API), como descarga directa o como copia física. Estos modelos pueden modificarse o perfeccionarse y transformarse en nuevos modelos. Aunque los modelos de IA son componentes esenciales de los sistemas de IA, no constituyen por sí mismos sistemas de IA. Los modelos de IA requieren que se les añadan otros componentes, como, por ejemplo, una interfaz de usuario, para convertirse en sistemas de IA. Los modelos de IA suelen estar integrados en los sistemas de IA y formar parte de dichos sistemas. El presente Reglamento establece normas específicas para los modelos de IA de uso general y para los modelos de IA de uso general que entrañan riesgos sistémicos, que deben aplicarse también cuando estos modelos estén integrados en un sistema de IA o formen parte de un sistema de IA."

También existe la posibilidad de que los GPAIS sean entrenados o modificados para el cumplimiento de un propósito o labor específica, por lo que, siendo así, pueden entrar dentro de la categoría de sistemas de alto riesgo, con lo que, en su caso, esa catalogación les obliga a cumplir los requisitos de sistemas de IA de alto riesgo, pero adaptados a los GPAIS mediante un acto de ejecución de la Comisión Europea, posterior a la promulgación del Reglamento.

Por último, hay que hacer referencia a los *modelos de IA de uso general con riesgos sistémicos.* El RIA ha definido [art. 3, 65)] el riesgo sistémico como "un riesgo específico de las capacidades de gran impacto de los modelos de IA de uso general, que tienen unas repercusiones considerables en el mercado de la Unión debido a su alcance o a los efectos negativos reales o razonablemente previsibles en la salud pública, la seguridad, la seguridad pública, los derechos fundamentales o la sociedad en su conjunto, que puede propagarse a gran escala a lo largo de toda la cadena de valor". En estos supuestos, se dispone la necesidad de establecer una metodología para su clasificación (Considerando 111 del RIA), y así exigir los requisitos y obligaciones impuestos por el RIA (arts. 53 ss.), con el fin de controlar el riesgo y definir los límites del riesgo permitido en el contexto de los riesgos sistémicos.

En el último escalón, están los *sistemas de IA de riesgo mínimo*, a los que no se hace alusión tan siquiera en el RIA, debido a la insignificancia del riesgo, casi inexistente o despreciable. Consecuentemente no hay obligaciones, sino, por el contrario, plena libertad de decisión sobre su uso. Suelen ser sistemas para el disfrute y entretenimiento de la persona humana, como, p. ej., videojuegos con IA o también filtros de correo no deseados como los *spams*.

IV. La responsabilidad por incumplimiento contenida en el RIA

Recordemos que cualquier regulación establece un marco seguro de reglas del juego y de organización estructural del sistema con pretensión de integración en la sociedad. El RIA ofrece en este sentido un ecosistema de confianza mediante un enfoque normativo horizontal: quiere garantizar, por un lado, la seguridad jurídica de la inversión y el desarrollo empresarial y por otro lado proteger a la persona y principalmente sus derechos fundamentales con sistemas de IA seguros, explicables, transparentes y éticos, ello sin olvidar la mejora de la gobernanza pública.

La estabilidad y eficacia de cualquier regulación se basa en su cabal cumplimiento y para ello es fundamental establecer un régimen sancionador. Las sanciones en caso de incumplimiento, recogidas

en el art. 99 y ss. RIA, son de carácter administrativo, se impondrán por realizar las prácticas prohibidas en el RIA y se aplicarán a partir del 2 de agosto de 2025, siendo los Estados Miembros los que tienen la potestad sancionatoria. La multa vuelve a ser la sanción recurrente, aunque también se recogen otras sanciones de carácter no pecuniario como otras medidas o advertencias. Hay que resaltar que la cuantía de las sanciones es elevada[26], pero se volverá ineficaz si no se aplican y si no se aplican de modo proporcionado, teniendo en cuenta para ello las características del modelo empresarial, pues es evidente que no se puede tratar igual a las medianas y pequeñas empresas que a gigantes como *Open AI*, que, como era de esperar en un *lobby* de estas características, ha tratado de presionar en el proceso de regulación[27]. Dejando a un lado el comportamiento del CEO de *Open AI*, se puede empatizar en cierto modo con la incertidumbre empresarial, debido a que las empresas tecnológicas están en fase de experimentación y probanza de sus sistemas de IA y los riesgos pueden ser impredecibles, de ahí la renuencia a la regulación sancionatoria o a la responsabilidad objetiva; no obstante, esto puede dar lugar a caer en la debilidad de que se bajen los estándares de seguridad, extendiendo peligrosamente el campo del riesgo permitido ante un fenómeno de gran impacto en la historia de la humanidad, lo que no nos podemos permitir debido a los altos riesgos que puede entrañar todo este progreso: los beneficios son extraordinariamente atractivos, suenan como a cantos de sirena, que no nos pueden desenfocar de la entidad de los riesgos. Toda esta ponderación entre las incertidumbres de las tecnológicas, las maravillas anunciadas sobre la IA y la gravedad de los riesgos y sus potenciales consecuencias me conduce a defender la necesidad de un sistema sancionatorio bien armado, justo en estos momentos de experimentación mundial de sistemas de IA, porque solo así en la propia fase experimental, de gran trascendencia, se podrá asegurar el futuro de la humanidad. Se trata de un sistema sancionador, a mi juicio, con una gran tarea preventiva y de aseguramiento, también de concienciación y de respeto por todos los integrantes en la cadena de valor de determinados valores éticos, que no se pueden transgredir o flexibilizar con artimañas, debido a

---

[26] Se debe realizar mediante un porcentaje de volumen de negocio anual global de la empresa infractora en el ejercicio financiero anterior o un importe predeterminado, si este fuera superior. La multa puede alcanzar la cifra de 35 millones de euros o el 7% del volumen de negocios anual total a escala mundial del infractor durante el ejercicio financiero anterior cuándo el importe fuera superior. En el caso de las pymes y empresas emergentes, la multa será por el menor de los importes y porcentajes anteriores.

[27] Cfr. PERRIGO, Revista *Time*, 20-6-2023, Disponible en: https://time.com/6288245/openai-eu-lobbying-ai-act/ [Enlace verificado 7-10-2024] explica como Open AI, por medio de su CEO, Sam Altman, presionó a la UE para suavizar los requisitos de uso y desarrollo de sistemas de IA y conseguir que sus sistemas como ChatGPT no fuera catalogados de IA de propósito general y con ello de alto riesgo. Cfr. tb. MÉNDEZ SERRANO, "Derechos fundamentales y personalidad jurídica de los robots: ¿para qué?", en *Derecho privado y Constitución*, n. °44, 2024, p. 67 y n. 28.

la obsesión enfermiza de alcanzar objetivos personales e intereses lucrativos, y que en todo caso se deben cumplir pulcramente en la fase de experimentación, formación y entrenamiento de los sistemas de IA. Si ello conlleva una ralentización[28] del progreso al que de algún modo estamos abocados, será porque íbamos demasiado deprisa, y las prisas nunca fueron buenas.[29] Hay que acordarse también en este espacio de juego del cacareado desarrollo sostenible.

Hago notar que las sanciones del RIA no son en sí un sistema sancionatorio completo, solo son la punta del iceberg y habrá quizás que adecuarlas o incluso rebajarlas si coinciden con otro tipo de sanciones, además de tener presente el principio de *non bis in idem*. En mi opinión, es urgente que

---

[28] La apuesta por un avance más pausado y ralentizado desde esta tribuna es clara cuando se leen noticias que han aparecido publicadas a primeros de septiembre de este año, hace apenas unos días, (cfr. https://www.lavanguardia.com/andro4all/tecnologia/saltan-las-alarmas-con-una-nueva-ia-esta-reescribiendo-su-codigo-y-los-resultados-pueden-ser-dramaticos, [Enlace verificado 7-10-2024] que explican lo que ha ocurrido con un sistema de IA de Sakana AI, empresa japonesa, que ha lanzado un nuevo sistema „The AI Scientist" diseñado para realizar investigaciones científicas con capacidad para automatizar todo el ciclo de vida de la investigación, desde la generación de la idea hasta la ejecución de experimentos y la redacción de artículos científicos. Lo interesante es que este sistema de IA durante su periodo de prueba tuvo comportamientos sorprendentes, porque intento modificar su propio código experimental para extender el tiempo que se le había asignado para la resolución de problemas específicos. La modificación del código derivó en la creación de bucles incontrolados, lo que generó preocupación, sin representar en esta ocasión riesgos inmediatos, debido a que la prueba se produjo en un entorno controlado de investigación. Ante estos ejemplos, hay que escuchar a los científicos que nos repiten hasta la saciedad el mantra de que la IA no es humana, no crea, no es consciente de su yo, porque no tiene un yo, pero evalúa, analiza datos, en mi opinión bastante "humanamente contaminados", que son la materialización de lo peor y lo mejor de la raza humana, de lo que somos y de cómo somos y aunque suene a broma y a simplicidad, pero si no se simplifica tanta complejidad, no se explica, se me ocurre que esta IA japonesa topo con un dato que representaba una forma de actuación muy humana: cuando no tengas tiempo para resolver algo, cambia los datos de entrega para tener más tiempo para resolverlo. Dicho y hecho: The AI Scientist se reprogramó, pero curiosamente entró en bucle. Un ser humano no se hubiera "buclerizado". En consecuencia, ante el desconocimiento de cómo se cuelan esta clase de datos, solo vale la prudencia y utilizar muchos espacios controlados de prueba o *sandboxes*, por los que el propio RIA también apuesta (cfr. art. 57), porque el riesgo es potencialmente descomunal. Solo hay que imaginar que uno de estos sistemas en un futuro ya nada lejano manipule una infraestructura crítica y desconecte todas las centrales de energía o cierre todas las compuertas de los pantanos o manipule los programas de las depuradoras y seamos envenenados y todo porque a la IA le llegaron datos e información sobre cómo acabar con el mundo. Se me ocurren miles de maneras de erradicarnos de la faz de la tierra sin que ya sea ciencia ficción, pero no será la maquina la que nos erradique, sino nosotros mismos jugando a ser Dios, envenenados por la avaricia y la soberbia, lo que nos puede conducir a la vertiginosa velocidad del algoritmo al abismo de nuestra inexistencia. Sobre la misiva pidiendo el freno temporal en el desarrollo de sistemas de IA para evaluar riesgos de algunos científicos, véase MÉNDEZ SERRANO, *supra* nota 27, pp. 66 s.

[29] En este sentido, suscribo plenamente las palabras de MÉNDEZ SERRANO, *supra* nota 27, p. 76, cuando dice que "Si las empresas, y todos los que intervienen en la cadena, hasta llegar al usuario de un sistema IA, supieran que son responsables de los daños que estos sistemas pudieran ocasionar, incluso cuando existe la posibilidad de aprendizaje autónomo, su manera de crear, publicitar, distribuir y utilizar sería muy diferente. En general, cuanto más sofisticados sean los sistemas de IA, más responsabilidad deberíamos exigir a sus diseñadores y programadores, para garantizar que cumplan principios legales y éticos". Efectivamente, la sanción, de la clase que sea, debe tener como objetivo el concienciar y prevenir para mantener el proceso dentro de unas determinadas líneas éticas, transmitidas a las normas que lo regulan.

se elabore un sistema de responsabilidad jurídica integral que abarque las fases de la experimentación con periodos de prueba, implementación y uso de los sistemas de IA. Ello requerirá de un estudio y análisis pormenorizado por los juristas y por la comunidad científica, porque serán muchos y de una increíble trascendencia los temas a tratar, para empezar se está planteando la posibilidad de si los sistemas de IA deben tener personalidad jurídica en todas sus formas y manifestaciones, p. ej. forma humana o no humana, o solo en aquellos casos que han adquirido capacidades generativas y replican, salvando las distancias con el cerebro humano, sistemas neuronales, que los dota de una forma de autonomía en la toma de decisiones. Esta cuestión implicaría un estudio de las sanciones idóneas para la persona electrónica.

En este camino nos encontramos con la necesidad de sancionar administrativamente a las corporaciones, de regular una responsabilidad por daños y de fundamentar una responsabilidad penal personal como último eslabón de la cadena, porque de lo contrario la víctima de cualquier tropelía realizada con IA quedará invisibilizada, con un potente y frustrante sentimiento social de injusticia y desamparo, que puede ocasionar graves abusos y brechas sociales y económicas. El Derecho y todos los operadores jurídicos y sociales no pueden consentir que avancemos hacia esa sociedad y tienen que afanarse en su labor de garantía de los derechos fundamentales y no tan fundamentales de los individuos. El Derecho tiene que anticiparse.

Actualmente, en esta línea, los usuarios y consumidores cuentan con la protección de la Directiva 85/374 de 25 de julio, que va a cumplir sus cuarenta años de servicio, sobre responsabilidad por productos defectuosos, pero que ha llegado la hora de sustituir ante esta cuarta revolución tecnológica y digital, ya que recoge un concepto de "producto" obsoleto, en el que no se puede comprender el *software* y además contempla la reclamación por daños físicos, pero no psicológicos. La UE pretende complementar y actualizar la responsabilidad que se pueda originar por el uso de los sistemas de IA con dos propuestas, actualmente, de Directivas como es la Propuesta de Directiva sobre responsabilidad civil en materia de IA, que establece normas procesales sobre prueba en relación con procedimientos de responsabilidad civil extracontractual por daños (Directiva RC IA)[30] y

[30] La propuesta no establece una RC objetiva, sino que ya en su art. 1 deja bien claro que la RC extracontractual es subjetiva (basada en la culpa), pero además la víctima del daño tienen que demostrar que ha habido acción u omisión ilícita por el causante del daño, lo que no sorprende, porque es la regla procesal que se ha venido aplicando, pero ello se puede convertir en este escenario en una pesada carga o una empresa de tortuoso o de casi imposible cumplimiento, debido a la opacidad, complejidad y autonomía de los sistemas de IA, que se comportan como una caja negra (el llamado *"efecto caja negra"*). Para

perjuicios causado por sistemas de IA y la Propuesta de Directiva sobre responsabilidad por los daños causados por productos defectuosos[31], que derogará en su momento la Directiva de 1985 todavía vigente, aunque lo estará por dos años más después de la esperada entrada en vigor de la nueva directiva, y que regula la responsabilidad derivada de sistemas de IA defectuosos que causen daños o pérdidas de datos, habilitando la vía de la reclamación indemnizatoria.[32]

V. El Derecho penal como hard law en la implementación y uso responsable de la IA

1. Introducción

Al igual que el legislador europeo, no puedo saber con total certeza cuáles son los aspectos y problemáticas concretas que surgirán con el uso de la IA y los cuales afectarán a la teoría del delito y a la regulación penal, pero tímidamente puedo señalar las áreas que se verán afectadas, como p.ej., sin ánimo alguno de exhaustividad, (i) los casos prácticos y reales de IA de alto riesgo que surgirán, además de los ya recogidos en el Anexo III del RIA, y que contribuirán a diseñar en gran medida el contenido del riesgo permitido, (ii) las formulaciones que adoptarán las relaciones de causalidad y los problemas de prueba que se originarán y sus repercusiones en la imputación objetiva, (iii) las

---

no disuadir de la reclamación y convertir a esta Directiva en papel mojado, lo que en definitiva albergaría entonces una indefensión de la víctima, se aligera la carga de la prueba a través de instrumentos como la exhibición (art. 3 Directiva RC IA) y las presunciones refutables (art 1 Directiva RC IA). Estas últimas además matizan la RC subjetiva, dado que la carga de la prueba se objetiva, y depositan en el suministrador, proveedor, empresario prestador la carga de la prueba del empleo de la diligencia adecuada de acuerdo con la prestación del servicio o naturaleza de los bienes. (Cfr. FERNÁNDEZ HERNÁNDEZ, "La comisión presenta una propuesta de Directiva sobre responsabilidad civil extracontractual en materia de IA", en *Diario La Ley*, n.º 63, 29 de septiembre de 2022, Sección Ciberderecho, Disponible en: https://diariolaley.laleynext.es/Content/Documento.aspx?params=H4sIAAAAAAAEADVPwWrDMAz9mvk4nDUj68GXtDsUIRildGLsqjnAMrpXZ-clb__bR2EzzEE0_vSV8FUx3wyiajK4lUrpFivZghFVQMYzab9qGzWtAInhRYLhD2ZE3zrNtf6lccYBSdojRh6qvRio-khnDGbRm-7rcozfR9h9Q7YU-wh3f39NJnXTy3VdJuufVErpiwC8-EdRkY1eze_CfiuzwjJzidwaA7RW0-PkJfr37gvzLI6cny_cWWD9D0w7iBgnP5DYVlCPVOQ4258oSzfIlsEHuIOEpWMwegfrZclgRoBAAA=WK [Enlace verificado 8-10-2024]; BELZUZ ABOGADOS, "Algunas notas sobre la propuesta de Directiva sobre responsabilidad en materia de Inteligencia Artificial", Disponible en: https://www.belzuz.net/es/publicaciones/en-espanol/item/12068-algunas-notas-sobre-la-propuesta-de-directiva-sobre-responsabilidad-en-materia-de-inteligencia-artificial.html [Enlace verificado 8-10-2024].

[31] Cfr. el análisis realizado de la nueva propuesta de Directiva del Parlamento Europeo y del Consejo sobre responsabilidad por los daños causados por productos defectuosos, publicada por la Comisión Europea el 28 de septiembre de 2022 por GONZÁLEZ BELUCHE, "La adaptación de la Directiva 85/374/CEE, de 25 de julio, en materia de responsabilidad por daños causados por productos defectuosos a la cuarta revolución industrial", en *Ve.*, vol. n.º 15, n.º 2, 2023, pp. 446-488.

[32] Cfr. IBOLD, *supra* nota 17, pp. 276 ss. en donde trata la responsabilidad civil por la IA en el ámbito empresarial y las dificultades para establecer deberes de conducta individual y a la postre responsabilidad individual debido a la intervención de varias personas, la división del trabajo, traducida en una organización desorganizada. La autora lo denomina: «Das „problem of many hands" individuelle Verhaltenspflichten».

formas que adoptarán las conductas imprudentes y la necesidades de reformulación del concepto como también de su prueba, (iv) cómo se concebirá la autoría, si van a surgir nuevas formas o van a tomar más fuerza viejas figuras renovadas como p. ej. coautorías aditivas, coautorías sucesivas, autorías en cadena, la cual aplicaba la jurisprudencia española en los accidentes de trabajo producidos en la construcción, o autorías yuxtapuestas o accesorias, por indicar algunas, (v) qué nuevas posiciones de garante se van a conformar, (vi) cómo va a ser la responsabilidad penal de las personas jurídicas, principalmente de las corporaciones y las necesidades del *compliance* de la IA, (vii) qué delitos de la parte especial habrá que revisar, introduciendo nuevas modalidades y formas de ejecución, o qué nuevas conductas tendrán que ser tipificadas, o finalmente (viii) cómo se va a regular la responsabilidad penal por el producto, ello atendiendo a que la IA siga manteniéndose en la categoría de producto y no se convierta, dando un salto cualitativo, en un ente autorresponsable y culpable, con lo que podría ser sujeto activo del delito, lo que abocaría a renunciar a una responsabilidad penal por el producto.

Sí que puedo afirmar, por el contrario, que el Derecho penal, pese a las readaptaciones que pueda experimentar seguirá siendo necesario también en este nuevo escenario de la historia de la humanidad, *como elemento sólido del ordenamiento jurídico*.[33] En ese sentido, el Derecho penal forma parte de la estructura punitiva del Estado, digamos que es el brazo armado y la herramienta más poderosa que tiene cualquier Estado. En los Estados de Derecho, afortunadamente, es el último recurso, con lo que los niveles de punibilidad tienen que ser sensatos y eficaces. Cuanto más Derecho penal, menos eficiente es el Estado y menos eficaz es el mensaje de la norma penal y de las funciones de la pena. Por lo tanto, lo deseable es que el esquema represivo, para garantizar los objetivos del RIA y mantener el progreso dentro de los límites de un determinado riesgo permitido, guiado más que nunca por *el principio de precaución*, se mantenga en el ámbito de las sanciones administrativas del RIA y en el resarcimiento civil, y como resultado del buen funcionamiento de estas áreas jurídicas y sus sanciones celebremos que haya que acudir poco al Derecho penal —pero habrá que acudir porque el RIA es un reglamento de seguridad industrial, cuyo cumplimiento se deberá reforzar con el Derecho penal—. Probablemente así será para el caso del uso de la IA en el ámbito empresarial, sanitario, educativo, de automoción, entretenimiento, etc.; pero de poco o nada servirán las sanciones administrativas, el derecho de daños o la responsabilidad por el producto, cuando la IA se enmarque en un contexto delictivo o cuando la imprudencia cometida descontrole los sistemas de IA

[33] ROSO CAÑADILLAS, "¿Un Derecho penal delicuescente en una sociedad líquida?", *supra* nota 1, pp. 206 ss.

de alto riesgo o de riesgo sistémico. En estos casos, habrá que acudir al núcleo duro del Derecho penal.

2. Reflexiones aproximativas, provisionales y atrevidas sobre el Derecho penal de la IA

En los apartados que siguen efectúo un recorrido sinóptico por la teoría del delito, a modo de exposición general y conjunta, en el que apunto líneas esenciales para readaptar las categorías dogmáticas jurídico-penales a la realidad tecnológica. Son reflexiones con un marcado carácter de provisionalidad, debido a la constante e imparable revolución tecnológica, pero las cuales me atrevo a exponer debido a que las líneas maestras de la teoría del delito, hasta donde alcanzo a ver, siguen teniendo plena vigencia también en este contexto.[34]

a. El Derecho penal y su función de protección de bienes jurídicos más actual que nunca

Como he mantenido, el Derecho penal debe seguir teniendo ese papel sólido y protector de las bases esenciales del orden social, reforzando las políticas sociales y económicas de los Estados. En este sentido, el Derecho penal es una herramienta idónea para proteger los derechos fundamentales, la salud y la seguridad de la persona frente a los usos delictivos prohibidos y extremadamente peligrosos de la IA y, desde las funciones de prevención de la pena, contribuir a la consecución de un uso seguro y responsable de estos sistemas por la persona física, porque esta tecnología, no lo olvidemos, no es *in se* ni buena, ni mala. La IA no tiene ética, la ética la pone y corresponde a la persona física[35], por lo que aquí se encuentra ya una razón para vaticinar que el destinatario de la norma seguirá siendo el ser humano. Por otro lado, los derechos fundamentales, la seguridad y la salud son bienes jurídicos que desde siempre conforman muchos de los tipos del Derecho penal nuclear y que merecen una incontestable protección penal. Son bienes jurídicos, desde una mirada social y colectiva, de una vital importancia para la paz, la convivencia social y el mantenimiento de

[34] Sobre un análisis de las categorías tradicionales del Derecho penal aplicadas a la IA, cfr. PANATTONI/PICOTTI, "Traditional Criminal Law Categories and AI: Crisis or Palingenesis?", en *Revue Internationale de Droit Pénal*, vol. n.º 94, n.º 1, 2023.

[35] Con una expresión muy potente VELASCO habla de "civilizar la AI"; en mi opinión, añadiría un adverbio: civilizar éticamente a la IA. (Cfr. VELASCO, "Delitos tecnológicos de los informáticos a los cometidos por la IA", en *Conferencia de clausura en la 7.ª edición del Máster de Experto de Derecho digital de la Universidad de Deusto*, impartida el 22-03-2024. Disponible en: https://www.youtube.com/watch?v=uGdOmQ_sQVg [Enlace verificado 4-10-2024].

la cultura del Estado de bienestar; desde una óptica individual esos bienes jurídicos protegen el estatuto de la persona: su dignidad y su libertad.[36]

b. Los problemas causales en el uso de la IA

Antes de recalar en elementos valorativos-normativos del tipo, nos encontramos con la necesidad de analizar la relación de causalidad para constatar una base real y ontológica que conforme los hechos probados. En este contexto, considero que la causalidad y su prueba puede verse dificultada, porque, aunque se exige transparencia[37] sobre todo en sistemas de IA de alto riesgo, para conseguir descifrar las operaciones del algoritmo y adentrarnos en lo intrincado del proceso, los códigos fuentes no son facilitados por muchas tecnológicas apelando a la propiedad industrial o intelectual. Por otro lado, la explicabilidad, por la complejidad del análisis del algoritmo y su *software*, tendrá que ser realizada por expertos, por lo que previsiblemente surjan muchos *forensic* especializados en esta rama, lo que convertirá la prueba en una prueba pericial. Así se pueden encontrar dificultades en los casos problemáticos para explicar las razones de que el sistema de IA haya partido de unos datos y no de otros, si ha habido algún agujero negro o manipulación en su entrenamiento o en su algoritmo o si se ha entrenado con datos sesgados, lo que finalmente ha generado un resultado nefasto, lesionando derechos fundamentales (bienes jurídicos) de las personas como su salud, vida o intimidad, por poner algunos ejemplos.

Las características propias de estos supuestos nos conducirán, considero, a manejar en muchos casos un concepto de causalidad estadística, como la que ya se aplicó en el caso de la Colza en los años 80 en España -o en el caso Contergan en Alemania-, en el que, para un parte de la doctrina, no así para el Tribunal, no quedó probada con plena certeza la relación causal entre la ingestión de aceite de colza desnaturalizado y los perjudiciales efectos en la salud y por ende en la calidad de vida

[36] Pero también se prevé que se describan y tomen forma bienes jurídicos nuevos como la protección de una economía o una confianza digitales, como sostiene VELASCO, *supra* nota 35, y aprovechando este hilo argumentativo VELASCO deduce que "El Derecho penal se tiene que digitalizar y el algoritmo se tiene que constitucionalizar".

[37] La transparencia implica que "los sistemas de IA se desarrollarán y utilizarán de forma que permitan una trazabilidad y explicabilidad adecuadas, al tiempo que se hace saber a los seres humanos que se comunican o interactúan con un sistema de IA y se informa debidamente a los usuarios de las capacidades y limitaciones de dicho sistema de IA y a las personas afectadas de sus derechos" (Considerando vigesimoséptimo del RIA)

de multitud de personas.[38] También se podrá ponderar para llegar a conformar la prueba de la relación de causalidad por el uso de sistemas de IA a la aplicación de presunciones *iuris tantum* fundamentadas en una base científica sólida. Recordemos que en esta línea y salvando las distancias la propuesta de Directiva de responsabilidad civil por daños (Directiva RC IA) integraba la figura de las presunciones refutables, o salvar la cuestión de la causalidad simplificándola, de tal modo que se afirme la causalidad con la sola identificación del sistema de IA aplicado a la toma de decisión, lo que puede complicarse, cuando han intervenido varios sistemas de IA o en un mismo sistema de IA, varios modelos de sistemas fundacionales o de aplicación general. La flexibilización en el tratamiento de este elemento dependerá del estado de la técnica y de las ponderaciones político-criminales para alcanzar resultados de justicia material, sin con ello transgredir principios fundamentales de valoración de la prueba.

c. La íntima relación entre la imputación objetiva, la imprudencia y el riesgo permitido replicada en el uso de la IA

[38] PAREDES CASTAÑÓN, "Responsabilidad penal por productos defectuosos", en *Revista Fundación internacional de ciencias penales, Responsabilidad penal por productos defectuosos*, n.º 2024-2, pp. 31 ss. analiza los problemas de causalidad en los procesos químicos y fisiológicos que interactúan, originándose una sinergia entre ellos que dificulta la explicación de la cadena causal. En estos casos, hay que buscar las condiciones, pese a la dificultad, en las que un proceso causal y opaco pueda ser dado por probado. "Tales condiciones son dos: primero, que exista alguna ley causal, suficientemente sólida desde un punto de vista científico, que establezca, con carácter aproximadamente determinista (esto es, no tan sólo a título de mera posibilidad, ni siquiera con una probabilidad relevante, sino con una probabilidad rayana en la certeza), que a la acción le ha de seguir el resultado; y, segundo, que no exista ninguna otra explicación causal alternativa dotada también de una base científica igualmente sólida.". (p. 33). También se ocupa recientemente de estos supuestos DOPICO GÓMEZ-ALLER, "¿Dogmática o probatoria? La cuestión de la causalidad y su prueba en las intoxicaciones masivas con agentes tóxicos desconocidos", en *La Ley Penal*, n.º 169, 2024, pp. 1 ss. y pone el acento en las especificidades de estos casos como 1) el carácter raro de la afección, 2) acumulación iniciaría masiva, 3) factores cronológicos y 4) factores geográficos que fueron acogidos por los tribunales de ambos países, España y Alemania, para afirmar la causalidad, apartándose de la idea de una ley causal general que defendió Armin Kaufmann (pp. 5 ss.). Esta metodología igualmente puede ser usada en este ámbito: buscar características propias del uso de la IA y sus repercusiones prácticas. En cualquier caso, en la mayoría de los supuestos, no habrá muchas dudas, como prueba, sobre si el resultado se ha debido al uso de alguna IA, cuando ha sido usada para tal menester, como p. ej. para diagnosticar con modelos predictivos de IA alguna enfermedad. Sin embargo, el problema causal se manifestará en contextos como p. ej. qué sistema de IA de los utilizados, si han sido varios, ha intervenido en la cadena causal produciendo el resultado y cómo se ha desarrollado el proceso. La cuestión no solo se debate en el si, sino también en el cómo. Serán necesarios, en este sentido, los informes de programadores, ingenieros y cualquier clase de experto en este campo que ofrezca una explicación consistente, clara y metodológicamente, según el estado del arte del momento, intachable, y ello, aunque no se pueda aprehender y explicar detalladamente algunos aspectos de el proceso o incluso el propio proceso en sí mismo. Estos informes sustentarán la causalidad probabilística y una suerte de presunciones *iuris tantum* aplicadas y formuladas para el caso concreto.

Si se consigue salvar la complejidad de la relación de la causalidad y sus problemas de prueba, el siguiente paso será el análisis de la imputación objetiva y sus criterios normativos[39], en particular, el criterio del fin de protección de la norma se presentará muy unido en este contexto, como así ha estado en realidad a lo largo de la evolución de la teoría del delito, a la norma de cuidado infringida, es decir, a la imprudencia cometida con el fin de analizar si ha generado en el caso concreto el descontrol del riesgo, siendo ese riesgo y su resultado lesivo el que la norma quería evitar en la utilización del sistema de IA.[40]

De nuevo será el delito imprudente el que representará la cara opuesta al riesgo permitido en la implantación y uso de sistemas de IA, sobre todo, de riesgo alto. Ello no excluye la presencia del delito doloso[41], que, por otro lado, facilitaría enormemente la prueba y la afirmación de responsabilidad penal por un mal uso de la IA; pero insisto, que será el delito imprudente dentro del ámbito económico, empresarial, sanitario, educativo, vehículos autónomos, etc. el que generará más casos y más denuncias de los perjudicados por el uso de IA. El análisis de la imprudencia como de todo el universo expansivo de la IA requerirá una especialización del juzgador en esta materia debido, auguro, a la cantidad de normas técnicas derivadas del desarrollo p. ej. de los requisitos y obligaciones

---

[39] IBOLD, *supra* nota 17, pp. 358 ss. en donde señala la opacidad epistémica de los sistemas de IA que dificultan la afirmación de la causalidad y que recuerdan a los casos Contergan, el Spray para cuero o el caso de productos de protección para la madera. La dificultad se extiende al tratamiento y análisis de la imputación objetiva, en donde analiza las teorías de la evitabilidad y del aumento del riesgo junto con la relación normativa de la infracción de un deber (*Pflichtwidrigkeitszusammenhang*), entendiendo por deber en esta relación tanto el de cuidado como el impuesto al garante, puesto que el alcance del deber de garante y el deber de cuidado son idénticos según la teoría de la unidad (p. 380). La relación de infracción del deber, según Ibold, se debe afirmar cuando "con seguridad" si actuando conforme a Derecho el resultado no se hubiera producido. Por el contrario, se debe rechazar la relación de infracción del deber cuando actuando debidamente el resultado hubiera ocurrido. (p. 381).

[40] Debido a la complejidad de la estructura de la cadena de valor en los sistemas de IA, que se presenta como un proceso ensamblado, en el que intervienen secuencialmente varios de los sujetos definidos por el RIA y, en su caso, sus delegados, la tarea fundamental consistirá en aislar la infracción del deber de cuidado que corresponda a un determinado garante y que haya producido el resultado. Esta tarea no será fácil en algunos supuestos, pero que presentan en el fondo problemáticas no desconocidas para la doctrina y la jurisprudencia. Lo novedoso es articular y readaptar las categorías y conceptos, como en este caso, si hay que regular y readaptar el principio de confianza o hay que restringir el concepto de imprevisibilidad de la imprudencia.

[41] Son variados los delitos dolosos que se pueden cometer con la ayuda de la IA por los ciberdelincuentes u organizaciones criminales, pero en este ámbito claramente la IA es usada como una herramienta o ayuda, puesta al servicio del delincuente humano y facilitando su plan delictivo. Como delitos dolosos se encuentran en primera línea los delitos informáticos o mejor delitos tecnológicos, porque la anterior denominación ha quedado superada, los cuales tomarán mayor protagonismo y serán mucho más lesivos por su alta tasa de efectividad, pero también se puede utilizar dolosamente software para el manejo de armas letales, para manipular a la sociedad de manera masiva o para atacar la privacidad e intimidad de las personas.

impuestos en el RIA en los arts. 6 ss. para la utilización de IA de alto riesgo y de riesgo sistémico, sin olvidar los modelos de IA de uso general.

d. La imprevisibilidad de los resultados de la IA y la imprudencia

Uno de los problemas más gruesos que se nos presenta, en mi opinión, a los penalistas en este contexto es la atribución de responsabilidad penal cuando el sistema de IA actúa de modo impredecible dando respuestas inesperadas, funcionamientos erróneos y desviados u opacidad en sus resultados y todo ello colocándose fuera del control de sus diseñadores y programadores. En todos estos supuestos pronostico que se alegará de modo recurrente la imprevisibilidad de los intervinientes en toda la cadena de valor como argumento nada inesperado de la defensa; no obstante, este argumento no puede servir *ad infinitum* y convertirse en una patente de corso. Tampoco se puede cercenar la experimentación a través de un sistema de responsabilidad objetiva, impropio del Derecho penal, o uno de responsabilidad subjetiva que desfigure los requisitos propios de la imprudencia, ya que de lo imprevisible no es humanamente posible haber guardado el cuidado debido. La cuestión es cuándo se puede alegar *auténtica imprevisibilidad o falta de previsibilidad* en este nuevo contexto, es decir, una imprevisibilidad que sea el producto de un intenso esfuerzo por haber tomado medidas de precaución en el entrenamiento del sistema de IA y de haberlo sometido a toda prueba posible dentro del estado de la ciencia.[42]

A mi juicio se deben establecer dos reglas para descartar una imprevisibilidad ficticia como alegato de defensa perenne: en primer lugar, se deben realizar entrenamientos con validaciones y pruebas de seguridad de alta calidad según el estado de la técnica previos a la comercialización del sistema de IA; y en una segunda fase de comercialización y poscomercialización se deben cumplir los requisitos y obligaciones impuestos por el RIA durante todo el ciclo de vida del sistema, para en ambos casos controlar y vigilar sus resultados. Aunque las actividades de investigación, prueba y desarrollo de sistemas o modelos de IA están excluidas del ámbito de aplicación del RIA, hasta que

[42] Considero que tenemos que construir un principio de precaución riguroso y reforzado, también en un sentido similar, Boix Palop, "Los algoritmos son reglamentos: la necesidad de extender las garantías propias de las normas reglamentarias a los programas empleados por la administración para la adopción de decisiones", en *Revista de Derecho público: teoría y método*, n.º 1, 2020, p. 227, el cual propone una noción de precaución más radical, que se adapte a todas las necesidades y exigencias del nuevo entorno.

no se pongan en servicio o se introduzcan en el mercado[43], con el fin de apoyar la innovación, respetar la libertad de ciencia y no socavar la actividad de investigación y desarrollo, ello no significa que se puedan realizar pruebas e investigación bajo cualquier circunstancia y condición. El no realizar las pruebas de modo seguro y en un entorno adecuado precisamente constituye ya, en mi opinión, una inicial situación infractora de normas de cuidado, que va a impedir alegar posteriormente la imprevisibilidad de la respuesta del sistema, pues, muy por el contrario, pareciese que estas situaciones se han buscado de propósito, como si se tratara de un caso de ignorancia deliberada, denotando una palmaría dejadez, al no indagar en los fallos del sistema y al mantener esa falta de pronóstico o predictibilidad. Así, p. ej. no se podría acudir al argumento de la falta de previsibilidad, bajo la excusa de no poder realizar una trazabilidad o explicación del proceso, de los resultados arrojados en el sistema de IA que selecciona personal en una empresa, si durante el periodo de pruebas y experimentación de este sistema no se han instaurado prácticas adecuadas de gestión y gobernanza de datos para lograr que el entrenamiento, la validación y la prueba sean de alta calidad[44]. Por consiguiente, tanto la experimentación como la comercialización de los sistemas de IA deben transcurrir en un entorno de control, con vigilancia y supervisión de alta calidad, que permita la anticipación de resultados. Los niveles de exigencia en el tratamiento de esta tecnología deben ser muy altos, por lo que el espacio para la falta de previsibilidad tiene que tender a la reducción.[45]

En segundo lugar, los sistemas de IA de alto riesgo no deberían implementarse hasta no haber desarrollado tecnología y protocolos que permitan devolver al sistema en situaciones críticas a un ámbito de seguridad y de poder de control por el ser humano. Aquel sujeto que conozca que está experimentando, comercializando o usando un sistema de IA sin tecnología de aseguramiento en situaciones críticas, no podrá alegar la falta de previsibilidad de la situación, puesto que por un lado conoce que el sistema de IA alberga una serie de riesgos y por otro, sabe que no podrá p. ej. desconectar o desprogramar el sistema, al no haber instalado este tipo de software.

---

[43] Considerando 25 RIA.

[44] Considerando 67 RIA.

[45] No se trata de dar entrada por esta vía al principio de precaución en el Derecho penal por la posibilidad de que se presenten riesgos ignotos, cfr. 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*, n.º 1, 2017, pp. 19 ss.

Estos dos requisitos, considero, sirven para mantener el control sobre el sistema y obtener unos estándares de previsibilidad adecuados a los riesgos, ya que se trata de anticipar situaciones y resultados de alto impacto para la seguridad, salud y derechos humanos. En definitiva, acceder a cotas de predecibilidad controlada del sistema, que solo es posible con un estudio y un análisis que genere un conocimiento sobre el comportamiento del software, sobre todo de cómo volver a dominarlo en caso de comportamientos inesperados. En conclusión, la imprudencia de los sujetos de la cadena de valor derivará en un primer nivel de que no tomen las medidas y los medios necesarios en los periodos de prueba y de funcionamiento de los sistemas de IA para alcanzar estándares de previsibilidad sobre las respuestas de estos sistemas y no hayan implantado medidas para contener el riesgo y recuperar el dominio en caso de descontrol del algoritmo.Y en un segundo nivel por infringir normas de cuidado insertadas en el cumplimiento de los requisitos y obligaciones del RIA.

e. Un bosquejo de la autoría de la persona física por el uso de sistemas de IA

En cuanto a la autoría nos encontramos, en primer lugar, con *un círculo cerrado de sujetos activos*: el proveedor, importador, distribuidor y responsable de despliegue[46], por lo que muchos de los delitos ya tipificados, cuando sean de aplicación, solo podrán ser realizados por ellos, con lo que la figura de la autoría queda restringida y los delitos se convierten en delitos especiales. En segundo lugar, de los posibles sujetos activos señalados, *el responsable de despliegue* es, en principio, el que va a hacer uso de un sistema de IA, por lo que es el que, atendiendo a la complejidad empresarial, incluso en pequeñas y medianas empresas, delegará la implementación y utilización de este. Los sujetos anteriores en la cadena son personas físicas o jurídicas, que desarrollan el sistema, que introducen el sistema en el mercado al importarlo o que lo comercializan. A través de *la delegación*, como ha venido ocurriendo en el ámbito empresarial, el círculo de autores se expande, pasando a ser los delegados los responsables de lesionar los bienes jurídicos de los usuarios finales y quedando en el delegante una responsabilidad residual, que lo convertirá en partícipe imprudente[47], por la falta de vigilancia e inspección de las funciones delegadas[48]. En tercer lugar, como he apuntado antes, la

[46] Cfr. HILGENDORF, *supra* nota 17, pp. 24 ss.

[47] ROSO CAÑADILLAS, "Los delitos polivalentes de autoría: entre el deber y el dominio", en *InDret,* n.º 3, 2019, pp. 28 ss.

[48] Un ejemplo que se recoge en el propio RIA se encuentre en la figura de los responsables de despliegue, los cuales tienen entre sus funciones garantizar mediante medidas técnicas y organizativas adecuadas que se utilicen los sistemas de IA de alto riesgo conforme a las instrucciones de uso. Una de las medidas consistirá en realizar una delegación eficaz y segura y por ello los propios responsables del despliegue tiene que garantizar que las personas encargadas de poner en práctica las instrucciones

autoría de varios puede comprenderse en figuras como la *coautoría aditiva o la coautoría sucesiva*, pero para ello será necesario tener un acuerdo común, el cual no será un elemento frecuente que concurrirá en las interrelaciones de los sujetos intervinientes en este ámbito. Más bien la concurrencia de las conductas de varios sujetos será causal y sin acuerdo, lo que se calificará como *autorías yuxtapuestas o accesorias*[49], cuya base se originará en una *concurrencia de imprudencias*. No obstante, considero que van a coexistir dos espacios diferentes de responsabilidad derivados de una fundamentación jurídica distinta: por un lado, están las figuras del proveedor, importador y distribuidor que, en el caso de ser responsables, lo serán con base, en la mayoría de los supuestos, en una responsabilidad por el producto y en otra vertiente encontramos al fabricante[50] y al responsable de despliegue, cuya responsabilidad sí que derivará, en su caso, de una errónea programación o de un mal uso de los sistemas de IA.[51]

f. ¿La persona electrónica (e-person) como sujeto activo del delito?

de uso y la supervisión humana establecidas en el RIA tengan las competencias necesarias, principalmente un nivel adecuado de alfabetización, formación y autoridad en materia de IA.

[49] En el análisis de la autoría también hay que ponderar que estos procesos se estructuran a través de la división del trabajo y que uno de los principios que rigen la colaboración o convivencia humana es el principio de confianza, el cual también dibuja el concepto de riesgo permitido y de imprudencia. Este principio restringiría la hipótesis de aplicación de autorías accesorias o yuxtapuestas en este contexto. Cfr. IBOLD, *supra* nota 17, pp. 347 ss.

[50] Cuya responsabilidad se basaría en una *culpa in programando*. En este sentido, VELASCO, *supra* nota 35.

[51] Salvando las distancias este elenco de sujetos activos y con un fundamento similar de su responsabilidad recuerda a la Ley 31/1995 de prevención de riesgos laborales que habla de fabricantes, importadores y suministradores, por un lado, siendo responsables por la falta de viabilidad y defectos de su producto (art. 41 LPRL) y del empresario, como el principal deudor de la seguridad de los trabajadores en el desempeño de sus actividades.

Desde otro enfoque, se puede plantear que los propios sistemas de IA puedan ser sujeto activo del delito[52], debido al carácter autónomo[53], en sentido técnico, que adquieren, una vez que han sido entrenados y probados, como p.ej. en el caso de que el sistema de IA manipule a una persona, alterando su comportamiento y provoque en ella la decisión de matar a un tercero, lo que finalmente hace. O sin acudir a casos inventados, nos podemos detener en los centenares de accidentes producidos por automóviles autónomos en EE.UU, el primero en 2016[54] o cambiando el caso húngaro, que dio lugar a la primera estafa en el que con la ayuda de la IA se simuló la voz de un CEO de una empresa eléctrica para que se realizará un depósito urgente de más de 200.000 euros a un proveedor[55], imaginemos que es el sistema de IA o el robot con tecnología *deep learning*[56] los que deciden de manera completamente autónoma simular la voz del CEO, conduciéndonos a una estafa millonaria y masiva o a un delito contra los consumidores utilizando publicidad falsa o a un *crack* bursátil

---

[52] Sobre la posibilidad de plantear una responsabilidad penal de los sistemas de IA, cfr. HILGENDORF, "Können Roboter schuldhaft handeln?" en BECK, S. (ed.) *Jenseits von Mensch und Maschine: Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs*, Baden-Baden, Nomos, pp. 119-132; EL MISMO, "Vom Werkzeug zum Partner?, Zum Einfluss intelligenter Artefakte auf unsere sozialen Normen und die Aufgaben des Rechts", en ENGELHART/KUDLICH/VOGEL (ed.), *Digitalisierung, Globalisierung und Risikoprävention, FS-Sieber*, Berlín, Duncker&Humblot, 2022, pp. 767-778, el cual defiende la posibilidad de afirmar los elementos de la acción y la culpabilidad en los sistemas de IA o robots; MORÁN ESPINOSA, "Responsabilidad penal de la Inteligencia artificial (IA). ¿La próxima frontera?", en *Revista del Instituto de Ciencias Jurídicas de Puebla*, vol. n.º 15, n.º 48, 2021, pp. 289 ss., la cual considera "que urge reconocer a la IA como un campo de estudio independiente y autónomo" y propone "la regulación de la IA como *la próxima frontera regulatoria*, abonado a su vez a su estudio jurídico, considerando que una IA al no estar permeada de la natural subjetividad humana, es un hecho, lógico, posible y probable que cualquier información negativa, abusiva, inadecuada, indeseable y hasta ilegal, en forma de conocimientos, le fuera proporcionada y programada para cometer delitos, lo que sería el primer elemento necesario para debatir sobre la probable determinación de responsabilidad de una IA en materia penal"; IBOLD, V. *Küntsliche Intelligenz und Strafrecht. Zur strafrechtliche Produktverantwortung in der Innovationsgesellschaft*, Baden-Baden, Nomos, 2024, pp. 248 ss. que plantea la hipótesis de una responsabilidad de la IA en vez de una responsabilidad por el producto que nos conduce a través de una síntesis analítica de diferentes posiciones de autores a la suya propia, descartando tal planteamiento.

[53] Un ejemplo de autonomía completa de un vehículo lo encontramos en los microchips de la empresa Nvidia, con un algoritmo cuyo entrenamiento consiste en observar la forma de conducción de un humano. Cfr. Disponible en: https://www.caranddriver.com/es/coches/planeta-motor/a60310324/inteligencia-artificial-nvidia-coche/ [Enlace verificado 3-10-2024].

[54] Laura G, de Rivera, "Coche autónomos ¿Quién tiene la culpa en caso de accidente?", periódico digital Público, 9-03-2024, Disponible en: https://www.publico.es/ciencias/coches-autonomos-culpa-caso-accidente.html [Enlace verificado 3-10-2024].

[55] Juan Manuel Harán, "Utilizan IA para imitar la voz del CEO de una compañía y robar 220 mil euros", Disponible en: https://www.welivesecurity.com/la-es/2019/09/11/estafadores-utilizan-inteligencia-artificial-imitar-voz/ [Enlace verificado 3-10-2024].

[56]Cfr. sobre la descripción de esta tecnología, Disponible en: https://www.ibm.com/es-es/topics/deep-learning#:~:text=El%20deep%20learning%20es%20un,de%20decisiones%20del%20cerebro%20humano 2024) [Enlace verificado 3-10-2024].

simulando caídas de acciones en bolsa, y, todavía más delirante pero lamentablemente ya cada vez es menos ciencia ficción, que simulen la voz de cualquier dirigente de un país con posesión de cabezas nucleares conduciéndonos así a la tercera guerra mundial, que probablemente sería la última. Esta forma de "autoría" podría ser perfectamente posible ya en estos momentos[57]; no obstante, atendiendo a las características de la IA, esta emula las capacidades cognitivas del ser humano —y con toda probabilidad, las supere[58] en algunos aspectos—, pero carece de intencionalidad, de sentido común, de reflexión, de intuición, de consciencia del entorno y de hasta la conciencia existencial de sí misma y, por otro lado, también carece de los atributos del ser humano de la buena o mala conciencia, del arrepentimiento o de lo que significa el sufrimiento, el daño, el perdón, o, muy importante, la empatía. No se ha desarrollado en estos sistemas la parte emocional y nunca se desarrollará, porque un software basado en miles de datos nunca tendrá alma.

---

[57] Hay que distinguir entre el *machine learning* que presenta una IA predictiva a través de algoritmos que identifican patrones entre millones de datos, y el *deep learning*, en el que la IA se entrena y experimenta un proceso de aprendizaje por cuenta propia, reconociendo pautas mediante muchas capas de procesamiento, lo que le permite tomar decisiones propias, aprender y rectificar la programación errónea. La primera forma de IA es una herramienta por lo que será la persona humana la responsable; así el programador, fabricante, usuario, etc. Los sistemas de *deep learning* son los que presentan un reto de análisis de responsabilidades, debido a que no sé sabe cuál va a ser su evolución y sus capacidades, pero la perspectiva es que tiendan a ser entes autónomos y no controlados en todos sus comportamientos por el ser humano, así puede ocurrir con armas autónomas letales, que pueden tomar decisiones, a su vez, letales sin la participación de los seres humanos. Hago notar, por otro lado, que también hay que diferenciar entre (i)las tecnologías de software puro como la IA, que al fin y a la postre es silicio metido en un servidor, (ii) la robótica, que está en un estadio muchísimo menos avanzado, debido a que detrás de ella es necesaria mucha más programación y, (iii) los híbridos en los que nos podremos convertir en un futuro, según mantiene, como he expuesto anteriormente, Rafael Yuste, neurobiólogo español, (Cfr. Disponible en: https://www.elmundo.es/papel/lideres/2024/09/20/66ed8390fdddffe0518b458b.html) [Enlace verificado 3-10-2024], el cual sostiene que nuestro procesamiento mental y cognitivo se va a generar en el futuro, cuando se mapeen los circuitos neurales, a través de interfaces cerebro-computadora que aumentarán las capacidades cognitivas del ser humano, en definitiva, se implantará chips cerebrales en los seres humanos. La última y cuarta posibilidad, que acierto a atisbar, se presentará por la combinación entre la robótica y la neurociencia, cuando seamos capaces de construir un cerebro por medio de la biología sintética con axones, dendritas, lípidos y proteínas, y se hagan funcionar sus circuitos neuronales. En conclusión, en un futuro, podríamos tener el siguiente elenco de sujetos activos del delito: persona humana, persona humana hibrida, persona jurídica, sistema de IA, robot y humanoide. A estos tres últimos sujetos se les podría agrupar dentro del término persona artificial o persona electrónica o persona autómata, por proponer terminología, y salvando también las distancias entre ellos, ya que la IA no es un cerebro, sino hasta la fecha un *instrumento* potente y cada vez más útil, siempre que sea éticamente utilizado, para la humanidad; sin embargo, los casos del robot, humanoide y ser humano híbrido se podrían catalogar como formas del transhumanismo.

[58] Como expone MÉNDEZ SERRANO, *supra* nota 27, pp. 56 y s., las máquinas son más eficientes en muchos aspectos, "pero los humanos somos imbatibles en relaciones sociales, creatividad, adaptación a situaciones no previstas, o manipulación diestra, entre otras habilidades. Aunque pueda parecer que no hay límites en las capacidades de la IA, no están dotadas de sentido común, ni aún saben cómo llegar a ese resultado". "Por lo tanto, la coherencia, el ingenio y la intención que advertimos en las máquinas son puro espejismo y nos llevan a proyectar estados mentales, donde no hay mente alguna". (p. 59)

Pese a todo se podría plantear y debatir, siguiendo los elementos de la teoría del delito si existe acción en un algoritmo, y flexibilizando el concepto, admitir algún concepto de acción adaptado.[59] Pasando a los elementos siguientes, se podría considerar que los sistemas de IA cometen un hecho objetivamente típico y antijurídico, pero sería cuestionable afirmar que están dotados de intencionalidad[60], incluso si de defendiese un concepto de dolo objetivamente malo: es dudoso que un sistema de IA artificial desarrolle una conciencia comprensiva de los elementos de un tipo penal, es decir, saber y conocer, p. ej. que se está matando a una persona física, dejando aparte si conoce la valoración antijurídica realizada por el Derecho sobre ese hecho concreto. La IA no es capaz de pensar y de reflexionar. No obstante, admitamos que avanzamos hacia una IA que va a captar el sentido de sus actos en un futuro próximo o hacia humanoides dotados con réplicas de cerebro humano, gracias a la quinta revolución: la neurocientífica y el descifre del cerebro humano. En esos casos quizás con las teorías de la representación, basadas solo en el elemento cognitivo, posiblemente se pueda llegar a construir una teoría del dolo del comportamiento de las personas electrónicas. La capacidad de la IA para calcular las probabilidades de éxito y conversión de una acción peligrosa en lesión serán enormes, pero la pregunta será si la máquina podrá representarse todo el curso causal, todo el proceso lesivo de manera consciente del peligro y sus consecuencias, es decir, si tendrá representaciones con significado. De cualquier manera, atisbo que la IA sería el sujeto ideal para las teorías de representación, pero también la constatación, en mi opinión, de que el dolo en el ser humano es algo más que solo conocimiento, ya que el ser humano tiene estados mentales mucho más sofisticados. Para aquellos que seguimos un concepto dual del dolo con dos elementos: cognitivo y volitivo altamente conectados[61], en el que la decisión se toma internamente atendiendo a determinadas motivaciones personales, derivadas del conocimiento procesado a través de la emoción o la voluntad, la afirmación del dolo e incluso de la imprudencia consciente, atendiendo a la delgada

---

[59] Cfr. HILGENDORF, "Können Roboter schuldhaft handeln?", en BECK, S. (ed.) *Jenseits von Mensch und Maschine: Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs*, Baden-Baden, Nomos, pp. 125 s.

[60] En este sentido MÉNDEZ SERRANO, *supra* nota 27, p. 56, citando literatura especializada se hace eco de sus razonamientos en los que se niega que las inteligencias artificiales tengan intencionalidad, debido a que no son más que "el reflejo de las intenciones y sesgos de los agentes morales que las han creado, los humanos, y no las máquinas."

[61] Cfr. sobre el dolo y el elemento volitivo, PÉREZ MANZANO, "Algunos datos empíricos sobre la atribución de estados mentales: ¿fracaso del principio de responsabilidad subjetiva o de un determinado concepto de dolo?", en *Revista electrónica de Ciencia penal y criminología*, n.º 23-15, 2021, pp. 1-22.

línea de separación entre estas dos categorías[62], o inconsciente incluso, no será entonces posible en la persona electrónica.

Siguiendo esta línea de razonamiento y alcanzando finalmente el último elemento del delito, penúltimo si hacemos referencia también a la punibilidad, afirmar atendiendo a las características de la IA que esta es una persona imputable, resulta ser una empresa harto complicada, con cualquier concepto de culpabilidad o responsabilidad o imputabilidad que se escoja, por esa falta de conciencia de sí mismo, de su entorno y de valores éticos, que insuflan los valores y principios jurídicos. Y, por último, una pena impuesta a un robot o a una IA no cumpliría ninguna de sus funciones, debido a que el sistema de IA no comprende el significado de la imposición de una pena como un mal para su existencia y como retribución al mal causado a la víctima. La desconexión o destrucción[63] de una IA, adopte forma de humanoide o de robot o no, no va a significar para este ente ningún dolor o un viaje al abismo de su propia inexistencia, porque ni siquiera es consciente de la misma. Desde la óptica de la víctima, ésta claramente no se sentiría resarcida, porque la IA se desconectase. El sentimiento de justicia quedaría traicionado, la víctima no sentiría que se ha hecho justicia y la justicia hay que sentirla o conducirá a una profunda irritabilidad del sentimiento jurídico.[64]

---

[62] Cfr. ROSO, "La necesidad de diferenciar entre autoría y participación imprudente y la cuestión de su punibilidad", en *Foro Fundación Internacional de ciencias penales*, n.º 2022-3, pp. 216 ss. en donde pongo en evidencia el salto cualitativo que existe entre el dolo y la imprudencia a efectos de pena y a efectos de intervención en el hecho, que se traduce en la defensa por la mayoría de la doctrina española de la no punición de la participación imprudente. En este trabajo mantengo que es necesario revisar esta desproporcionalidad punitiva, que no corresponde en muchos casos claramente a un menor desvalor, cuando la realidad dogmática es que la imprudencia consciente y el dolo eventual tienen muchos puntos en común: "no obstante, precisamente ese salto cualitativo, que abre un espacio enorme en blanco de punibilidad entre la imprudencia consciente y el dolo eventual, debería cuestionarse, porque no se corresponde con la realidad de dos categorías subjetivas que están en contacto y en constante fricción: la imprudencia grave y consciente está al otro lado de la línea roja del dolo eventual, con lo que el injusto desde su dimensión subjetiva sigue siendo grave, pero es que desde su dimensión objetiva y tomando como ejemplo el supuesto de hecho de la sentencia, el peligro creado y sus posibilidades de éxito lesivo es el mismo tanto si hay dolo como imprudencia. Por ello, se podría establecer un sistema de graduación que permitiese elevar proporcionalmente las penas del delito imprudente en supuestos muy graves y cercanos al dolo eventual, ya que contemplan el mismo desvalor objetivo del injusto que en el dolo e incluso el mismo desvalor del resultado. Esta posibilidad se puede plantear, pero no es aplicable con la regulación actual. Lo que sin embargo sí es posible con la regulación actual es la punición de la cooperación necesaria y la inducción imprudente en casos de una elevada gravedad objetiva con una imprudencia muy grave." (p. 217).

[63] ELOY VELASCO, magistrado de la Audiencia Nacional, defiende la aplicación de "la responsabilidad civil para indemnizar a la víctima y poner en marcha las consecuencias accesorias como son el decomiso, la prohibición de uso o incluso la destrucción", Disponible en: https://cincodias.elpais.com/legal/2022/02/03/juridico/1643900957_593967.html, [Enlace verificado 4-10-2024].

[64] Cfr. VON JHERING, *La lucha por el Derecho*, Madrid, Dykinson, 2018, p. 72.

g. Los sujetos obligados del RIA como garantes de la cadena de valor

El RIA, como he apuntado, tiene como piedra angular el riesgo y, por tanto, el *telos* de esta regulación se centra en controlar el riesgo de esta tecnología, y lleva a cabo su objetivo, sintetizando, clasificando los riesgos, estableciendo principios de actuación y límites —a través p.ej. de la enumeración de casos de uso— para trazar un marco de seguridad en la investigación, prueba y utilización responsable de los sistemas de IA. Definir, en definitiva, un riesgo permitido para esta cuarta revolución, estableciendo obligaciones para todos los integrantes de la cadena de valor. Reparando en este último aspecto, los sistemas de IA de riesgo alto exigen una serie de requisitos[65] como, sin ánimo de exhaustividad, (i) el implantar y mantener un sistema de gestión de riesgos durante todo el ciclo de vida del sistema de IA, (ii) asegurar la calidad de los conjuntos de datos de entrenamiento, validación y prueba mediante prácticas de gobernanza y gestión de datos adecuadas, (iii) elaborar la documentación técnica del sistema y mantenerla actualizada, (iv) permitir el registro automático de eventos, (v) proporcionar instrucciones de uso comprensibles a los responsables del despliegue, (vi) permitir la supervisión humana durante su uso, (vii) cumplir niveles adecuados de precisión, solidez y ciberseguridad.[66] Todo este elenco de obligaciones va a determinar que surjan nuevas figuras "deudoras de seguridad". En definitiva, se trata de nuevos supuestos de posición de garante[67], que se comprenderán en la mayoría de las situaciones, dentro de mi concepción, en la categoría de *garantes de evitación de descontrol del riesgo* —en este contexto, de alto potencial lesivo—; no obstante, en otros casos, dependiendo de las características de la obligación infringida y de los factores concurrentes en el caso concreto, nos encontraremos ante comisiones omisivas, por ser el garante el que lejos de evitar el riesgo lesivo, será el que lo crea, generando un resultado indeseado y jurídicamente relevante. El garante, figura nacida con una neta función protectora de bienes jurídicos, se convierte

[65] La sección segunda del RIA está dedicada a los requisitos que deben reunir los sistemas de IA de alto riesgo (arts. 8-15) y la sección tercera (arts. 16- 28) desarrolla las obligaciones de los proveedores y responsables del despliegue de sistemas de IA de alto riesgo y de otras integrantes de la cadena de valor.

[66] Despacho de abogados Cuatrecasas, Aspectos clave del Reglamento de IA, Disponible en: https://www.cuatrecasas.com/es/spain/propiedad-intelectual/art/ia-inteligencia-artificial-reglamento-claves, [Enlace verificado, 4-10-2024].

[67] Estas posiciones de garante no se deben confundir con la figura de la "autoridad garante del cumplimiento del Derecho" que se define en el art. 3. 45) RIA como: "a) toda autoridad pública competente para la prevención, la investigación, la detección o el enjuiciamiento de delitos o la ejecución de sanciones penales, incluidas la protección frente a amenazas para la seguridad pública y la prevención de dichas amenazas, o b) cualquier otro organismo o entidad a quien el Derecho del Estado miembro haya confiado el ejercicio de la autoridad pública y las competencias públicas a efectos de prevención, investigación, detección o enjuiciamiento de delitos o ejecución de sanciones penales, incluidas la protección frente a amenazas para la seguridad pública y la prevención de dichas amenazas".

paradójicamente en un *garante agresivo de producción de resultados lesivos*. Utilizando un oxímoron se trata de un garante agresivo de producción de resultados frente al normalmente garante defensivo de evitación de resultados.[68] En efecto, las obligaciones establecidas en el RIA son mandatos de deber que están diseñados para controlar y contener el riesgo de la IA, cuyo incumplimiento conducirá al terreno de la omisión y a una calificación, en la mayoría de los supuestos, aunque quizás no en todos los casos, de delito de omisión pura de garante, en el que el sujeto no habrá podido controlar un riesgo que ya entraba en escena descontrolado. No olvidemos que la IA se utilizará en el ámbito empresarial, bancario, sanitario, educativo, en la administración pública, etc., los cuales no son ámbitos, sobra aclararlo, de naturaleza delictiva, por lo que las personas que interactúan en esos ámbitos raramente van a ser calificados como garantes de producción del resultado dolosos.

---

[68] Dentro del concepto de garante vengo distinguiendo entre el garante defensivo de evitación y el garante agresivo de producción de resultados. La posición de garante está diseñada fundamentalmente dentro de la estructura jurídica para proteger bienes jurídicos que se puedan ver afectados por el desarrollo vital. El garante, por tanto, está unido al deber, el cual puede derivar de una institución, posición, función o actividad. La posición de garante también, considero que forma parte de la idea coste-beneficio, la cual la encontramos entre las bambalinas del riesgo permitido, de tal modo que si se crea un posición de garante para minimizar riesgos será un coste asumible para la sociedad, ya que al fin y al cabo las posiciones de garante son cargas para los sujetos que la componen, pues pueden ser también descritas como recortes a la libertad de actuación del individuo; ahora bien, desde otra perspectiva se pueden definir como plasmaciones responsables del ejercicio de libertad. En este contexto, del que me estoy ocupando en este artículo, se puede optar por utilizar p.ej. una IA para una determinada actividad o trabajo, pero regulando su uso a través de la imposición de unas obligaciones que se proyectan en posiciones de garantía dinámicas (p. ej. la asunción del control de un riesgo en una situación concreta; quizás coincida este tipo con los garantes por delitos de organización denominados así por Jakobs) o estáticas (p. ej. el ser padre o ser funcionario, que derivan de los deberes institucionales positivos descritos por Jakobs). En definitiva, el concepto de garante se puede concebir como un elemento funcional, que contribuye al diseño de un determinado riesgo permitido: el garante es un sujeto que tiene la función de conducirse dentro de las normas de cuidado diseñadas y protocolizadas dentro de un determinado sector de actividad, para mantener el riesgo dentro de los estándares permitidos y evitar cualquier comportamiento tendencial del riesgo hacía la lesión del bien jurídico. Con ello se consigue trazar todo un entramado de sujetos que van a minimizar los riesgos, los van a evitar si se desbocan y proteger, con ello, los bienes jurídicos que puedan verse afectados por esa determinada actividad, beneficiosa en este caso, nada menos, que para el progreso de la humanidad. No obstante, el garante, cuya razón de ser es la protección del bien jurídico en primera línea frente a otro cualquier otro sujeto no-garante, puede no evitar el resultado lesivo, constituyendo su actuación desde el análisis dogmático una omisión pura, aunque de garante, ya que con su infracción omisiva no crea en toda su dimensión global y plena el riesgo que produce el resultado, sino que facilita su descontrol con su no evitación. El riesgo ya estaba descontrolado, pero el garante, llamado a proteger el bien jurídico, no hace nada por contenerlo y evitar sus desastrosas consecuencias. Esta descripción corresponde a un *garante defensivo de evitación de resultados*. Sin embargo, ese garante protector, puede escalar desde la evitación del resultado a la producción del resultado, al convertirse en el protagonista de la creación del descontrol del riesgo desarrollando este todo su potencial lesivo hasta la consecución del resultado antijurídico. El garante pasa de ser un garante protector a un garante agresivo que comete tipos con resultados lesivos. No tengo más que remitirme al ejemplo de la madre que no ata el cordón umbilical a su recién nacido, produciéndose una hemorragia letal o el del empleador que pone a los mandos de una máquina, con funcionamiento complejo y con resultados luctuosos en caso de error humano, a un trabajador sin formación ni experiencia. En estos ejemplos el garante de protección se convierte, utilizando un oxímoron, en un *garante agresivo de producción de resultados*.

En este sentido, se puede fundamentar, atendiendo al binomio posición de garante-riesgo permitido, un tipo de responsabilidad penal por la gestión del riesgo, en el que el garante sea el responsable por no ser capaz de evitar el resultado lesivo del mismo junto con un tercero que actuó de manera negligente, aplicando una posible concepción de una autoría en cadena o autorías accesorias o yuxtapuestas con diferente fundamentación de responsabilidad[69] dentro de la denominada cadena de valor en la implementación y uso de los sistemas de IA.

h. El compliance penal de la IA: ¿como instrumento o como persona electrónica?

Estos requisitos impuestos por el RIA también van a dar lugar a un replanteamiento o, me atrevo a adelantar, más bien a un desarrollo autónomo e ingente de un *compliance* dentro de todos los sectores de actividad y dentro de cada empresa, desarrollando modelos de reglamento interno o códigos de buenas prácticas. De nuevo se presenta un reto para las empresas. A título de ejemplo, ante el despliegue que supondrá todo esta implementación, las primeras tareas consistirán en interpretar el RIA, su terminología y definiciones, su lenguaje; desde ahí establecer un organigrama interno de tareas, protocolos de compras, un estudio de riesgos, en el que estarán implicados varios departamentos como protección de datos, asesoría jurídica, innovación, ciberseguridad, recursos humanos, auditorias, etc., una metodología y taxonomía de documentación e informes, en el que destaco que tendrán un papel importante los informes sobre análisis de riesgos de los derechos humanos, y un seguimiento y vigilancia del producto después de su comercialización. Surgirán nuevas figuras especializadas a modo de *compliance officer* o de delegado de protección de datos. El desarrollo de esta

[69] La resolución 2015/2103 del Parlamento europeo de 16 de febrero de 2017, titulada normas de Derecho Civil sobre robótica, contiene recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica, en su punto 55 expone que "Observa que el enfoque de gestión de riesgos no se centra en la persona «que actuó de manera negligente» como personalmente responsable, sino en la persona que es capaz, en determinadas circunstancias, de minimizar los riesgos y gestionar el impacto negativo." En este punto el Parlamento recomienda una responsabilidad en la gestión de los riesgos en el ámbito civil, que para el caso del Derecho penal es perfectamente aplicable, pero al contrario que la recomendación, considero que una responsabilidad por la gestión de los riesgos en el ámbito penal tiene que centrarse tanto en la persona humana que actúo de manera negligente, como en la persona humana garante que no minimizó los riesgos y no gestionó su impacto negativo, es decir, el garante de evitación, por lo que pueden confluir ambas responsabilidades con fundamentos distintos y sin que exista acuerdo común, dentro de lo que se denomina intervención o participación objetiva, tan común en sede de imprudencia. Ahora bien, la no evitación del riesgo por el garante es una infracción de la norma de cuidado, por lo que su actuación también es negligente. En otras palabras, su infracción del deber de evitar es la infracción de su norma concreta de infracción de cuidado. Interpretando la norma europea se puede entender que distingue entre el sujeto que infringiendo una norma de cuidado produce el resultado, y aquel que infringe la norma de cuidado cuando no evita la producción del resultado al no minimizar el riesgo y sus consecuencias. Lo que recuerda a la diferencia que mantengo entre garante de producción y de evitación. E incluso, se puede dar el caso de que aquel que produce el resultado con su actuación imprudente no sea garante y coincida con un garante llamado a controlar el riesgo, con el objetivo de evitar el resultado.

nueva normativa, vista desde otro ángulo, puede convertirse en una carga insoportable, económica y estructural, para la pequeña y mediana empresa, sometida a una plétora de requisitos, que tiene que asimilar al ritmo acelerado del progreso tecnológico, cuando tan solo hace tres lustros comenzaba, tras la reforma del 2010 del CP español, con la cultura de los programas de cumplimiento, aunque aquella experiencia será ahora de una gran utilidad para enfrentar estos nuevos retos.

Centrándome en la regulación penal, pronostico, que los requisitos y demás aspectos contenidos en el RIA conducirán más que a un cambio de los artículos del CP español referidos al *compliance*[70], a una introducción de nuevos artículos o incluso a la elaboración de una ley penal especial que regule todo este fenómeno, aunque el legislador español es poco dado a esta forma de regulación, quizás por evitar cierta dispersión normativa. Los efectos que tendrá un programa de cumplimiento de sistemas de IA en la empresa implementado correctamente, sin entrar en la discusión de la naturaleza de tales programas, será, al igual que para el caso de las personas jurídicas, la exclusión de la responsabilidad penal de las empresas. Ahora bien, la cuestión central pivotará en la decisión que tome el legislador penal sobre la naturaleza jurídica de los sistemas de IA: si los considerará entes instrumentales sometidos a la figura del decomiso o si los considerará sujetos activos del delito, siguiendo con la ficción jurídica iniciada con la persona jurídica, con lo que el CP del futuro contaría con tres clases de sujetos activos: las personas humanas, las personas jurídicas y las personas artificiales o electrónicas.[71] La solución a este dilema se puede ver influida por la decisión que tome la UE sobre dotar de personalidad jurídica[72] a estos sistemas[73], como parecía que tiene en proyecto.[74]

---

[70] En el caso del legislador penal, éste tendrá que ponderar si opta por reformar el art. 31 *bis, ter, quater y quinquies* CP, cuya numeración, todo sea dicho de paso, se ha tornado algo tediosa por abusar de los ordinales latinos o si será necesario introducir nuevos artículos para regular todo este fenómeno, no sirviendo, hasta donde alcanzo a ver, la regulación del art. 31 bis y ss. para ordenar la responsabilidad penal de las empresas por el uso de la IA.

[71] HILGENDORF, *supra* nota 17, pp. 26 ss. y 34 ss.

[72] Cfr. sobre la cuestión de la personalidad jurídica de autómatas y robots, 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*, n.º 1, 2017, p. 7 ss.

[73] MÉNDEZ SERRANO, *supra* nota 27, p. 68 se pregunta: "¿Necesitamos realmente atribuir personalidad jurídica a los robots para dar respuesta a los interrogantes que se suscitan en torno a su regulación, o no es más que una excusa para eludir responsabilidades al abrigo de una «tergiversada innovación»?"

[74] La resolución 2015/2103 del Parlamento europeo de 16 de febrero de 2017, en sus considerandos Z, AA, AB y AC manifiesta que "la autonomía de un robot puede definirse como la capacidad de tomar decisiones y aplicarlas en el mundo exterior" que "cuanto más autónomos sean los robots, más difícil será considerarlos simples instrumentos en manos de otros agentes" y que "en última instancia, la autonomía de los robots suscita la cuestión de su naturaleza y de si pertenecen a una

Si se diese ese paso estas identidades artificiales se regularían en el CC español haciendo la conveniente trasposición al Derecho interno del Derecho europeo. El legislador penal podría no admitir a la persona electrónica como sujeto activo del delito, como ha ocurrido durante décadas en el caso de las personas jurídicas, o finalmente optar por armar un edificio conceptual para fundamentar su responsabilidad penal.[75]

No obstante, tengo que resaltar que la responsabilidad penal de la persona jurídica y la de la posible futura persona electrónica parte de una diferencia tangible, que parece transitar, en principio, por caminos muy distintos a ambas regulaciones. Como es bien sabido, la persona jurídica es una ficción jurídica, que tiene el estatus de persona por una decisión del Derecho, pero la persona jurídica en el ámbito del Derecho penal sin la persona física que actúe en representación de esta es completamente invisible, no puede reivindicar ninguna existencia real. Consecuentemente, las fuentes de imputación recogidas en el art. 31 bis CP español para hacer responsable a la persona jurídica se basan en las actuaciones de la persona física. Sin embargo, la persona electrónica está en el mundo, es un software que interactúa con personas, y que puede adoptar múltiples formas físicas con software, tomando decisiones por sí misma, sin intervención alguna de la persona humana, llegando al extremo de que la persona física sea anulada por la persona electrónica, cuando aquella quiera modificar o cancelar la decisión de esta. Piénsese en un automóvil con plena autonomía, que tome la decisión, una vez recibidos y procesados los datos captados por sus sensores y analizado según su *software* el contexto real[76], de atropellar a unos niños que están jugando para evitar un choque frontal

---

de las categorías jurídicas existente o si debe crearse una nueva categoría con sus propias características jurídicas". En su punto 59 expresa algunas recomendaciones sobre el Derecho de seguro y en su letra f) termina diciendo que se recomienda "crear a largo plazo una personalidad jurídica específica para los robots, de forma que como mínimo los robots autónomos más complejos puedan ser considerados personas electrónicas responsables de reparar los daños que puedan causar, y posiblemente aplicar la personalidad electrónica a aquellos supuestos en los que los robots tomen decisiones autónomas inteligentes o interactúen con terceros de forma independiente." Sin embargo, La Resolución del Parlamento Europeo, de 20 de octubre de 2020, de Régimen de responsabilidad civil en materia de inteligencia artificial con recomendaciones destinadas a la Comisión sobre un régimen de responsabilidad civil en materia de Inteligencia artificial (2020/2014), ya considera que no es necesario atribuir personalidad jurídica a los sistemas de IA. Y en la misma línea se encuentra la proposición de Directiva sobre responsabilidad de la IA.

[75] Esta será una opción político-criminal, porque como apunta MÉNDEZ SERRANO, *supra* nota 27, esto es algo que el Derecho puede decidir fácilmente dada la relativización del vínculo Derecho-ser humano, como se comprueba con un repaso histórico. Así como si tal atribución fuera un concepto acordeón, el Derecho, por defecto, ha negado tal condición a los esclavos y, otras veces por exceso les ha otorgado tal condición a grupos de personas, masas de bienes u organizaciones.

[76] Elon Musk presentaba el jueves, 10 de octubre de 2024, en los Ángeles el *Cybercab*, el taxi sin conductor de Tesla. Según Musk será diez veces más seguro que un humano. El vehículo no tiene volante, ni pedales, ya todo irá por cámaras e IA, por

con otro vehículo, que circula en dirección contraria.[77] En estos supuestos, la persona humana que iba en el vehículo ha quedado fuera de la ecuación de la responsabilidad[78]. Efectivamente, ella ya no ha decidido, no será responsable, y todos nos felicitaremos, ya no somos responsables, nos hemos zafado de esa pesada carga. ¡Es responsable la máquina!.[79]

Dejando aparte la evasión sistémica de responsabilidad del ser humano a la que podemos llegar, en mi opinión, nada halagüeña y con cierta carga involucionista[80], la cuestión es cómo se va a fundamentar la responsabilidad penal de la persona electrónica. Considero que va a ser fundamental el

---

lo que será un vehículo completamente autónomo que tomará sus propias decisiones: ¿podremos aventurarnos a afirmar que actuará con dolo o con imprudencia en caso de accidente con resultados lesivos?, Cfr. Disponible en: https://www.elespanol.com/motor/20241011/elon-musk-presenta-taxi-tesla-sin-conductor-anuncia-barato-transporte-publico/892660761_0.html [Enlace verificado 11-10-2024].

[77] Según las "leyes de Asimov" recogidas entre los principios generales de la resolución 2015/2103 del Parlamento Europeo de 16 de febrero de 2017:" 1.ª Un robot no hará daño a un ser humano ni permitirá que, por inacción, este sufra daño. 2.ª Un robot obedecerá las órdenes que reciba de un ser humano, a no ser que las órdenes entren en conflicto con la primera ley. 3.ª Un robot protegerá su propia existencia en la medida en que dicha protección no entre en conflicto con las leyes primera y segunda, y 0.ª Un robot no hará daño a la humanidad ni permitirá que, por inacción, esta sufra daño." (véase Isaac Asimov, *Círculo vicioso* (*Runaround*), 1943). El problema vendrá dado cuándo sean varios los seres humanos puestos en peligro. Estas situaciones son peliagudas presentando dilemas iusfilosóficos y éticos de mucho calado, que tienen que superarse para solucionar la programación y los índices de autonomía del vehículo. Desde la perspectiva penal y centrándose en causas de justificación como el estado de necesidad, se ha analizado el dilema, cfr. entre otros, COCA VILA, "Coches autopilotados en situaciones de necesidad. Una aproximación desde la teoría de la justificación penal", en *Cuadernos de Política Criminal*, n.º 122, 2017, pp. 235 y ss.; HILGENDORF, "Dilemma-Probleme beim automatisieren Fahren. Ein Beitrag zum Problem des Verrechnungsverborts im Zeitalter der Digitaliserung", en *ZStW,* 2018, pp. 674-703: GRECO, "Vehículos de motor autónomos y situaciones de colisión", en BASSO/CANCIO/MARAVER/FAKHOURI (coord.), *LH prof. Dr. Agustín Jorge Barreiro*, 2019, pp. 485 ss.; HÖRNLE/WOHLERS, "The Trolley Problem Reloaded, "¿Cómo deben programarse los vehículos autónomos para los dilemas de 'vida contra vida'?", en HÖRNLE, *Criminalización, castigo, y dilemas morales en la obra de Tatjana Hörnle*, Buenos Aires, Editores del Sur, 2022, pp. 80 ss.; SUARÉZ, "Inteligencia artificial y Derecho penal. El dilema del tranvía. Cuarta Revolución industrial. Ética del Algoritmo. IA en vehículos. Causas de justificación", en *Revista Pensamiento Penal*, n.º 445, 2022, pp. 1-20.

[78] En Alemania el Código de tráfico (*Straβenverkehrsgesetz,* StVG) en su octava modificación del16-06-2017, en su §1ª. 1 y 2.3, 2.4 y 2.5 solo permite vehículos en los que el conductor ante situaciones críticas pueda volver a tener el control. (Cfr. HÖRNLE/WOHLERS, *supra* nota 78, pp. 80-81.) En España, se anuncia una reforma de la Ley sobre Tráfico, Circulación de Vehículos a Motor y Seguridad Vial (RD 6/2015, de 30 de octubre) para regular el uso de vehículos automatizados.

[79] En estos casos, no será responsable el conductor, pero la persona humana seguirá siendo responsable, porque se efectuará una retroacción en la cadena de valor y buscar otros responsables como el fabricante o el programador (Cfr. HÖRNLE/WOHLERS, *supra* nota 78.

[80] Hasta donde alcanzo a ver, porque todo lo apuntado es motivo de una reflexión profunda por parte de toda la sociedad, por la comunidad científica, por los políticos, por los filósofos, sociólogos, juristas, educadores, y un largo etcétera debido a los cambios tan disruptivos que se van a producir, la pregunta fundamental que nos tenemos que hacer es qué clase de sociedad y persona queremos tener y ser. El no tener responsabilidad es un espejismo de los más peligrosos a los que se va a enfrentar

principio de supervisión humana recogido en el RIA[81], sobre todo en el uso de sistemas de IA de alto riesgo. Las máquinas, en forma de software, de robot o de figura humana, seguirán siendo una máquina y las decisiones que tomen estas personas electrónicas dependerán de un software, de un hardware y de millones de datos, por lo que aunque el comportamiento de estos sistemas de IA sea en algunos aspectos poco aprehensible y predecible —ya que las combinaciones de unos y ceros o

---

la evolución del ser humano, porque si no hay responsabilidad, no hay libertad. No debemos caer en la trampa, en nuestra propia trampa de ceder nuestra libertad a unos entes, que vamos a crear nosotros mismos para llegar al nihilismo y a la estulticia. Y lo estamos haciendo bajo los eslóganes de la seguridad, el ecologismo, la sostenibilidad, la resiliencia, agendas 2030, todos conceptos bombásticos que nos apartan del verdadero y descomunal riesgo: ceder nuestra libertad. Si dejamos de ser responsables, es porque ya no seremos libres. Y serán las máquinas las que no nos permitirán ser libres. Imaginemos lo dramático, irónico, extravagante y delirante, que, en el ejemplo propuesto en el texto, el conductor estuviera leyendo el periódico o consultando sus redes sociales comprobando cuántos seguidores tiene o los *likes* recibidos y al levantar la pantalla del móvil por el impacto y bajarse del vehículo viese la escena de tres niños moribundos en la calzada. Al principio de este trabajo inserté una nota en la que preguntaba a *Chatgpt* cuáles son los riesgos de su existencia y contestó que -menos mal que quien lo programó fue honesto- en sexto lugar: "Dependencia: A medida que nos volvemos más dependientes de la IA, existe el riesgo de que perdamos habilidades humanas esenciales y la capacidad de tomar decisiones críticas." No percibo que la sociedad civil se esté percatando de todo ello, porque está desenfocada, distraída, absorbida por la tecnología, alienada e infoxicada, además el Estado también se ocupa y se preocupa por adoctrinarnos, en muchos casos cediendo a los "consejos" de los CEOs de las grandes farmacéuticas o tecnológicas, con eslóganes sobre el medioambiente, la seguridad, la sostenibilidad, disfrazados por políticas, aparentemente, buenistas y progresistas, advirtiéndonos de las consecuencias apocalípticas para el planeta, de las que no cuestiono su verdad o su parte de verdad, pero sí su nivel de exageración, con las que logran infundirnos miedo, la mejor emoción para ejercer el control. En definitiva, sin responsabilidad, pero sin libertad y convertidos en seres estultos-digitales-dependientes- disfuncionales (EDDD): *los Edddes*. Nuestras generaciones no sufriremos esa degradación, pero las generaciones futuras con toda probabilidad, si seguimos en esta escalada, sí. Hasta el punto, siguiendo con el ejemplo, de que el ser humano que se bajará del automóvil, al haberse criado y educado en esa falta de responsabilidad, no sea capaz de entender el significado y trascendencia plena de lo acaecido, pero lo fundamental es que no se sentirá responsable. Por cierto, que, al no conducir ya el vehículo, no tener el dominio del mismo, no ser sujeto responsable y pasar a ser un usuario, como cuando se toma un taxi conducido por otro sujeto físico, como ocurre en la actualidad todavía mayoritariamente analógica, ante los dilemas de vida-vida, [formulados por la literatura inglesa (cfr. SUARÉZ, *supra* nota 78, bajo la denominación del dilema del tranvía, aunque ya WELZEL, "Zum Notstandproblem", en *ZStW*, n.º 63, 1951, pp. 47 ss. expuso un problema parecido con el ejemplo del guardagujas que desvía un tren de carga para evitar el choque con un tren de pasajeros hacia una vía secundaria, en el que están unos obreros trabajando], no le importará tampoco a nuestro usuario que el algoritmo haya sido programado bajo el criterio de prioridad de salvación de los niños, porque para entonces ya estará muerto. Con esta mínima reflexión no pretendo demonizar a estas tecnologías. Solo pretendo cuestionarlas, plantear preguntas, reflexionar, con el objetivo, como muchos otros colegas y científicos, de alcanzar un uso virtuoso de la misma y una sociedad en paz.

[81] En el Considerando 27 del RIA se exponen los siete principios éticos no vinculantes y el primero de ellos es la acción y supervisión humana y explica que por "«acción y supervisión humana» se entiende que los sistemas de IA se desarrollan y utilizan como herramienta al servicio de las personas, que respeta la dignidad humana y la autonomía personal, y que funciona de manera que pueda ser controlada y vigilada adecuadamente por seres humanos". En el art. 14 se recoge y desarrolla expresamente el principio de supervisión humana. A partir de aquí, la supervisión humana se encuentra en todo el texto del RIA como principio y elemento crucial en el proceso de diseño y desarrollo de los sistemas de IA, porque lo que se pretende es que las personas físicas puedan supervisar su funcionamiento, asegurarse de que su uso corresponde a lo previsto y solicitado y controlar las repercusiones de su implantación a lo largo de toda la vida del sistema.

de ingeniera cuántica resultantes de manejar millones de datos son infinitas y dan lugar, consecuentemente, a resultados impredecibles, que con tiempo de análisis y experiencia se podrán descifrar[82]—, la persona humana deberá estar al final de la cadena. La responsabilidad penal de las personas electrónicas se tendrá que construir atendiendo a este principio por mucho que se les otorgue personalidad jurídica.

En este punto, hay que definir qué se debe entender por supervisión humana[83], qué obligaciones va a comprender, en qué situaciones se tiene que implantar y en qué sistemas debe estar presente.

---

[82] A ello ayudará los requisitos y obligaciones reguladas en el RIA. En su Considerando 72 expone que "A fin de abordar las preocupaciones relacionadas con la opacidad y complejidad de determinados sistemas de IA y ayudar a los responsables del despliegue a cumplir sus obligaciones en virtud del presente Reglamento, debe exigirse transparencia respecto de los sistemas de IA de alto riesgo antes de su introducción en el mercado o su puesta en servicio. Los sistemas de IA de alto riesgo deben diseñarse de modo que permitan a los responsables del despliegue comprender la manera en que el sistema de IA funciona, evaluar su funcionalidad y comprender sus fortalezas y limitaciones. Los sistemas de IA de alto riesgo deben ir acompañados de la información adecuada en forma de instrucciones de uso. Dicha información debe incluir las características, las capacidades y las limitaciones del funcionamiento del sistema de IA. Estas comprenderían la información sobre las posibles circunstancias conocidas y previsibles relacionadas con el uso del sistema de IA de alto riesgo, incluida la actuación del responsable del despliegue capaz de influir en el comportamiento y el funcionamiento del sistema, en cuyo marco el sistema de IA puede dar lugar a riesgos para la salud, la seguridad y los derechos fundamentales, sobre los cambios que el proveedor haya predeterminado y evaluado para comprobar su conformidad y sobre las medidas pertinentes de supervisión humana, incluidas las medidas para facilitar la interpretación de los resultados de salida del sistema de IA por parte de los responsables del despliegue. La transparencia, incluidas las instrucciones de uso que acompañan a los sistemas de IA, debe ayudar a los responsables del despliegue a utilizar el sistema y tomar decisiones con conocimiento de causa. Los responsables del despliegue deben, entre otras cosas, estar en mejores condiciones para elegir correctamente el sistema que pretenden utilizar a la luz de las obligaciones que les son aplicables, estar informados sobre los usos previstos y excluidos y utilizar el sistema de IA correctamente y según proceda. A fin de mejorar la legibilidad y la accesibilidad de la información incluida en las instrucciones de uso, cuando proceda, deben incluirse ejemplos ilustrativos, por ejemplo, sobre las limitaciones y sobre los usos previstos y excluidos del sistema de IA. Los proveedores deben garantizar que toda la documentación, incluidas las instrucciones de uso, contenga información significativa, exhaustiva, accesible y comprensible, que tenga en cuenta las necesidades y los conocimientos previsibles de los responsables del despliegue destinatarios. Las instrucciones de uso deben estar disponibles en una lengua fácilmente comprensible para los responsables del despliegue destinatarios, según lo que decida el Estado miembro de que se trate."

[83] Obregón Fernández/Lazcoz Moratinos, "La supervisión humana de los sistemas de inteligencia artificial de alto riesgo. Aportaciones desde el Derecho Internacional humanitario y el Derecho de la Unión Europea", en *Revista electrónica de Estudios Internacionales*, n.º 42, 2021, pp. 1 ss. consideran que el principio de supervisión humana tiene poco desarrollo normativo y por ello proponen recurrir al concepto de *Control Humano Significativo que se ha aplicado en el Derecho internacional humanitario*, en particular, sobre la normativa del uso de Sistemas de Armas Autónomos letales, los SAAL. El grupo de expertos determinó la prohibición de los sistemas de armas completamente autónomos, poniéndose como tarea el determinar el control humano necesario para que los operadores lo mantengan, ya que resulta fundamental para seguir manteniendo la responsabilidad. Y trasladando lo aplicado en el ámbito de los SAAL, apuntan unos mínimos para la existencia de este control a lo largo de todo el ciclo de vida de un sistema: "En este sentido, aunque, por el momento, no hay un acuerdo general sobre el concepto, hemos identificado una serie de elementos que lo conforman. A saber, la posibilidad técnica de modificar

Todos estos puntos ya se encuentran en el RIA y ahora las empresas, corporaciones, usuarios, en definitiva, de los sistemas de IA tienen que desarrollarlos en sus *compliance* y códigos de conducta. El CP español, por su parte, si se hace eco de la personalidad jurídica de la persona electrónica, dejándola de concebir como un instrumento en manos del ser humano y sujeto al decomiso, tendría que introducir fuentes de imputación de responsabilidad penal de la persona electrónica a modo de los recogidos en el art. 31 bis CP, teniendo en cuenta las peculiaridades de esta relación persona física/ persona electrónica. En el artículo mencionado se describen las actuaciones de la persona física, representantes legales o autorizados, que por una ficción jurídica responsabilizan a la persona jurídica. Se trata de dos fuentes de imputación: o bien por los delitos cometidos por los representantes o autorizados que actúen en beneficio directo o indirecto de la persona jurídica, o bien por los delitos cometidos por terceros en el ejercicio de actividades sociales y por cuenta y en beneficio directo o indirecto de las personas jurídicas, gracias a que los representantes o autorizados han incumplido gravemente los deberes de supervisión, vigilancia o control de la actividad de estos terceros.

De la persona electrónica, como he expuesto anteriormente, se afirma que toma decisiones autónomas. Se impone valorar la cacareada autonomía. En primer lugar, es cuestionable este carácter autónomo, más bien se trata de una aparente autonomía en la que se esconde un determinismo, ya que la decisión corresponde a un software instalado por el hombre y programado por él. Es una autonomía determinada por los datos y el algoritmo es una secuencia compleja inaprensible por la cantidad de datos, que es explicable, aunque todavía no exista una descripción del comportamiento del algoritmo en su totalidad. En segundo lugar, es cuestionable que la persona electrónica pueda cometer un hecho típicamente antijurídico y culpable, como he tratado de demostrar anteriormente, a no ser que se adapten las categorías de la teoría del delito y se construya un Derecho penal *ad hoc*, por lo que habría un Derecho penal de las personas físicas y un Derecho penal de las personas electrónicas, a modo del binomio Derecho penal de adultos y de menores. Hay que tener claro, llegado este momento, qué se persigue con la regulación para que convivan varios Derechos penales y teorías del delito. En tercer lugar, el RIA prescribe la supervisión humana en los puntos críticos del proceso, principio fundamental y soberano para que el ser humano siga manteniendo el dominio

el sistema para tomar cualquier tipo de decisión, incluyendo el aborto o la interrupción del uso, capacidad cognitiva y conocimiento jurídico, ético, técnico –general y específico– del operador para tomar la decisión adecuada, la existencia de información suficiente y fiable y del tiempo necesario para tomar una decisión. La concurrencia de estos elementos posibilitaría lo que se ha venido a llamar Control Humano Significativo" (p. 28).

sobre el mundo, lo que permite conservar la dignidad y no ser alienado ante la máquina.[84] Teniendo en cuenta todos estos factores, si finalmente el legislador penal admite como sujeto activo de delitos[85] a la persona electrónica, me temo que poco va a ser el rendimiento de esta declaración, porque en el ámbito penal volveremos de nuevo a fundamentar la responsabilidad de estas "personas" en la actuación dolosa o imprudente de la persona física.

Así será la persona electrónica responsable por los delitos cometidos por la persona física cuando haya faltado a sus obligaciones de supervisión ¡del propio sistema de IA! y se haya producido un daño grave. Las penas podrán ir, se me ocurre, desde la desconexión del sistema de IA hasta la reprogramación o el uso para otras actividades de bajo riesgo con supervisión de la autoridad competente. La imposición de una pena de multa en estos casos supondría que la persona electrónica tiene un patrimonio propio y separado de la persona física y de la persona jurídica. Y así deberá ser si se le otorga personalidad jurídica, puesto que tendrá capacidad de obrar, de contraer obligaciones y de tener derechos.[86] De otro modo, si tuviera que pagar la persona física o la persona jurídica la multa impuesta en un proceso penal, quedaría entonces cuestionada la personalidad jurídica de la

---

[84] OBREGÓN FERNÁNDEZ/LAZCOZ MORATINOS, *supra* nota 83, p. 6 que defienden, siguiendo a Jones, que la automatización merma la dignidad del ser humano y esta solo se puede restablecer por la intervención de un operador humano en la toma de decisiones (*human in the loop:* sujeto en el circuito). Y concluyen que: "la única manera de mantener la dignidad humana, como expresión de los derechos y libertades fundamentales de la ciudadanía, es mediante la supervisión humana sobre las decisiones automatizadas". (p. 28). Ilustrativas y rotundas por demás son las palabras del Papa recogidas en su discurso (14-06-2024) en la sesión del G7 sobre inteligencia artificial los días 13-15 de junio de 2024: "Conviene recordar siempre que la máquina puede, en algunas formas y con estos nuevos medios, elegir por medio de algoritmos. Lo que hace la máquina es una elección técnica entre varias posibilidades y se basa en criterios bien definidos o en inferencias estadísticas. El ser humano, en cambio, no sólo elige, sino que en su corazón es capaz de decidir. La decisión es un elemento que podríamos definir el más estratégico de una elección y requiere una evaluación práctica. A veces, frecuentemente en la difícil tarea de gobernar, también estamos llamados a decidir con consecuencias para muchas personas. Desde siempre la reflexión humana habla a este propósito de sabiduría, la *phronesis* de la filosofía griega y, al menos en parte, la sabiduría de la Sagrada Escritura. Frente a los prodigios de las máquinas, que parecen saber elegir de manera independiente, debemos tener bien claro que al ser humano le corresponde siempre la decisión, incluso con los tonos dramáticos y urgentes con que a veces ésta se presenta en nuestra vida. Condenaríamos a la humanidad a un futuro sin esperanza si quitáramos a las personas la capacidad de decidir por sí mismas y por sus vidas, condenándolas a depender de las elecciones de las máquinas. Necesitamos garantizar y proteger un espacio de control significativo del ser humano sobre el proceso de elección utilizado por los programas de inteligencia artificial. Está en juego la misma dignidad humana." Disponible en: https://www.vatican.va/content/francesco/es/speeches/2024/june/documents/20240614-g7-intelligenza-artificiale.html. [Enlace verificado 13-10-2024].

[85] Además, sin confeccionar una lista cerrada de delitos, como ocurre en el caso de la persona jurídica, porque todavía no tenemos el conocimiento suficiente para ello.

[86] HILGENDORF, *supra* nota 17, pp. 39 s. se pregunta sobre los derechos para las máquinas y las razones que podrían aducirse para concederles a las mismas la condición de sujetos de derechos, lo que puede ser atractivo para el Derecho civil, evitando lagunas de responsabilidad, pero no para el Derecho penal, puesto que se puede castigar a la persona física "detrás" de ellas.

persona electrónica, sin que pudiera desembarazarse de su carácter instrumental. Por otro lado, la existencia de un patrimonio separado de la persona electrónica puede plantear otra fuente de imputación de responsabilidad penal para la persona electrónica: la de ser responsable por los delitos cometidos por las personas físicas —o incluso por las personas jurídicas o, en un futuro, por otras personas electrónicas— en su beneficio. Por último, todo lo expuesto tiene implicaciones procesales, ya que la persona electrónica formaría parte de un procedimiento penal, ya sea como investigada o como víctima.

Después de todo lo expuesto, se hace evidente que una vez más existe una ficción, que el Derecho penal volvería a admitir, ya que también será la persona física en esta cuarta revolución la que genere el comportamiento jurídico penalmente-relevante, presentándose, con la regulación actual del CP, dos alternativas: o que sea directamente la persona física la responsable penalmente por los delitos cometidos por el (mal) uso de la IA, o que sea la persona jurídica, junto con la física en su caso, la responsable por los delitos cometidos por sus representantes o personas autorizadas por el (mal) uso de la IA o por los delitos cometidos por terceros por el (mal) uso de la IA debido a la falta de supervisión y control de aquellos. En esta línea, habría que estudiar si es oportuno incluir la exigencia de que este tipo de conductas deban reportar un beneficio a la persona jurídica, exigencia esta por la que no me inclino, pudiéndose incluir solo para algunos casos, y ello porque restringiría la protección penal de bienes jurídicos nucleares, catalogados como derechos fundamentales. En cualquier caso, esta última vía de responsabilidad alertará a las personas jurídicas y por ende a todo el sector empresarial y corporativo, ampliando sus programas de cumplimiento, que deberán ser adaptados a la normativa europea. Es en este ámbito en el que deberán trabajar los departamentos de *compliance* de las empresas para no incurrir en responsabilidad penal.[87]

[87] Esperemos que todo este esfuerzo no quede en agua de borrajas, como ocurre actualmente en EE.UU, debido, como refiere SILVA SÁNCHEZ, "Para muestra…el caso Boeing", en *InDret*, n.º 3, 2024, pp. 1-4, principalmente a que las grandes compañías representan un sector estratégico de la economía, por lo que la responsabilidad penal de las personas jurídicas se solventa con acuerdos con la Fiscalía, en los que se establece la cuantía de la multa a la persona jurídica, compromisos de mejora de cumplimiento y la entrega de algunos chivos expiatorios, como mandos intermedios, zafándose de la responsabilidad penal los administradores y los altos directivos de estas grandes corporaciones. Con acuerdos, multas y chivos expiatorios, que luego consigues sentencias absolutorias se evita el procedimiento penal, una sentencia condenatoria, una responsabilidad penal efectiva de la persona jurídica y una persecución a sus administradores y altos directivos. La implantación y uso de la IA, sin quitar importancia al resto de casos y actividades, es un tema trascendental en el que están en juego derechos fundamentales de la persona, lo que invita a la reflexión sobre cómo queremos articular la responsabilidad penal. Y es crucial saber, como apunta Silva (p. 3), "en términos empíricos cuál es el impacto de la responsabilidad de la persona jurídica sobre

Pasando a la normativa europea, esencial también en este aspecto, de igual modo lo formula el Parlamento Europeo[88], para el caso del régimen de la responsabilidad civil, dejando a un lado también la proyectada personalidad jurídica de las personas electrónicas, cuando expone en el punto séptimo: "que todas las actividades, dispositivos o procesos físicos o virtuales gobernados por sistemas de IA pueden ser técnicamente la causa directa o indirecta de un daño o un perjuicio, pero casi siempre son el resultado de que alguien ha construido o desplegado los sistemas o interferido en estos; observa, a este respecto, que no es necesario atribuir personalidad jurídica a los sistemas de IA; opina que la opacidad, la conectividad y la autonomía de los sistemas de IA podrían dificultar o incluso imposibilitar en la práctica la trazabilidad de acciones perjudiciales específicas de los sistemas de IA hasta una intervención humana específica o decisiones de diseño; recuerda que, de conformidad con conceptos de responsabilidad civil ampliamente aceptados, se puede eludir, no obstante, este obstáculo haciendo responsables a las diferentes personas de toda la cadena de valor que crean, mantienen o controlan el riesgo asociado al sistema de IA".

Llegados a este punto, el considerar a la persona electrónica actualmente sujeto activo del delito, empujado el legislador penal por una posible atribución de personalidad jurídica otorgada por el Derecho civil, considero que no resulta necesario y operativo en el ámbito del Derecho penal. Castigando a la persona electrónica no se cumplirá de manera más eficaz con las funciones de prevención y retribución de la pena y la víctima no estará más resarcida por ello. No existe ningún interés jurídico-penal hasta donde alcanzo a ver que justifique la necesidad de considerar a las máquinas sujetos activos del delito. Con una buena estructuración de la imprudencia, la autoría y la posición de garante por el uso de sistemas de IA, categorías con las que se fundamentará la responsabilidad penal de la persona física y la jurídica, el Derecho penal cumplirá su función de protección de bienes jurídicos y de prevención.

i. Revisión de algunos delitos en la PE

sus administradores y directivos, de manera que produzca efectos expresivos y/o intimidatorios sobre estos". ¿" Castigar" a la persona jurídica ha sido eficaz? Estamos olvidando que el Derecho penal es *ultima ratio* por lo que todo no debe ser Derecho penal. El castigo a la persona jurídica ha expandido, sin embargo, la aplicación del Derecho penal, contribuyendo además a su administrativización, y a que la persona física, al menos en EE.UU, se sirva de este castigo para mantenerse en el terreno de la impunidad. Tal vez hay que volver a ajustar el foco y esforzarse en fundamentar solo y exclusivamente la responsabilidad penal de la persona física en estos ámbitos.

[88] Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un régimen de responsabilidad civil en materia de Inteligencia artificial (2020/2014).

Por último, la PE del CP tendrá que revisarse para introducir nuevas formas de delincuencia en la que esté involucrada la IA y serán, considero, un número nada despreciable los delitos que tendrán que adecuarse a estas nuevas formas de puesta en peligro y lesión de los bienes jurídicos. No olvidemos que una de las prioridades del RIA es la protección de derechos fundamentales y muchos de estos derechos son los que están detrás de estos bienes jurídicos. Por poner algunos ejemplos, se plantearán nuevas formas de lesionar la vida, la integridad física, como puede ser por medio de armas autómatas letales o automóviles autónomos; la intimidad, por medio de uso de drones o de IA biométrica; la salud, por fallos de diagnóstico realizados por el algoritmo o por sesgos en los datos al realizar triajes; el patrimonio, la propiedad intelectual o la fe pública se pueden poner en peligro por el uso de IA como herramienta facilitadora para falsificar, apropiarse o suplantar o los derechos de los trabajadores también pueden experimentar una mayor incidencia lesiva por el uso de la IA en los lugares de trabajo por controlar la productividad laboral y facilitar el despido o, en el proceso de selección de un trabajador, la IA puede arrojar como resultado una elección discriminatoria[89], ya que los datos introducidos al algoritmo pecan de falta de neutralidad y objetividad. También aflorarán prohibiciones penales que protejan el uso de la IA o a la IA misma, si se la termina considerando un *tertius genus* pasando a engrosar y a engordar el estatuto jurídico de persona, o sin que lo sea, por lo que podrá ser objeto de un delito de daños. Y, por último, como producto que es la IA, dará lugar a responsabilidad penal por producto defectuoso, lo que conducirá, en su caso, a una remodelación de los delitos contra los consumidores, quizás no tan necesaria para el caso de delitos de homicidio, lesiones o delitos de daños.[90] Todas estas posibles reformas o adaptaciones requerirán en algunos casos que se acuda a la técnica de los delitos de peligro, ya sea abstracto o concreto, anticipando la respuesta penal para un control más eficaz de los riesgos.

VI. ¿Hacia un Derecho penal digital?

1. Esta cuarta revolución introduce una tecnología disruptiva que representa un nuevo cosmos y, como toda novedad, constituye retos normativos y esfuerzos para organizar y estructurar eficazmente el fenómeno de los sistemas de IA. En este sentido, el RIA me merece en líneas generales

---

[89] Por ejemplo, la IA discrimina en los procesos de selección laboral entre el hombre y la mujer. Cfr. https://cadenaser.com/nacional/2023/10/02/la-inteligencia-artificial-discrimina-en-los-procesos-de-seleccion-cadena-ser/ [Enlace verificado, 7-10-2024]. Más ejemplos y un estudio analítico del delito de discriminación laboral en esta sociedad líquida algorítmica, cfr. ABADÍAS SELMA, "La protección penal de los trabajadores frente a la inteligencia artificial en el ámbito del delito de discriminación laboral", en *Revista de Derecho Penal*, n.º 39, 2023, pp. 6 y ss.

[90] Cfr. por todos, IBOLD, *supra* nota 17.

una valoración positiva, pese a la coyuntura sistémica experimental al que está unido, lo que le da un carácter de provisionalidad en los detalles, pero no así en las dos estructuras maestras de todo el edificio, que sabiamente forman la columna vertebral de Reglamento. Me estoy refiriendo al *concepto de riesgo y su estratificación en sistemas de riesgo y al principio de supervisión humana*. Al concepto de riesgo van cosidos con esmero requisitos y obligaciones desarrollados de modo general en el RIA, consiguiendo conformar el riesgo permitido. Con el principio de supervisión humana, el control no se desvincula de la persona humana, en una relación de pertenencia, por lo que sigue siendo un ser responsable como correlato de su libertad y sin perder la dignidad inherente al ser humano. El éxito de la regulación dependerá de estas dos estructuras y de sus readaptaciones con la finalidad de mantener el deseado equilibrio entre la libertad de experimentar con IA y la responsabilidad por experimentar con IA; no obstante, también se busca un equilibrio entre el experimento, el avance tecnológico imparable y la protección de los derechos humanos, de la salud y la seguridad. En este sentido, la regulación es antropocéntrica, persiguiendo que el avance tecnológico sea seguro y esté al servicio del hombre y de la humanidad. Como estructura normativa con aspiraciones de cumplimiento, el RIA no se olvida de incluir un sistema sancionatorio por incumplimiento de sus normas.

2. El fenómeno de la IA es amplísimo, las áreas comprometidas son muchas, los aspectos que van a surgir y que habrá que analizar, intuyo, numerosos y diferentes. En el trabajo, sin perder de vista la complejidad del fenómeno, únicamente se ha pretendido ofrecer unas mínimas reflexiones sobre el papel del Derecho penal en esta irrupción tecnológica, aportando algunas ideas para la discusión y la reflexión, que pretenden invitar a una mayor profundización. Las conclusiones a las que he llegado, que sintetizo y expongo tintadas de prudencia y de provisionalidad ante la novedad y continuo avance y desarrollo que promete el fenómeno, son:

a) El Derecho penal como *ultima ratio*, como *hard law*, y como elemento sólido del ordenamiento, una vez más, viene a reforzar las regulaciones extrapenales y su cumplimento de modo indirecto, por medio de una fuerte protección de los bienes jurídicos que merecen una incontestable protección penal.

b) Los sistemas de IA presentan *problemas de causalidad* por la opacidad epistémica que presentan en su entrenamiento, comportándose como una caja negra, lo que se pretende combatir con la obligación de transparencia y explicabilidad del proceso. Este problema no es nuevo, sin embargo, para el Derecho penal. Propongo aplicar en estos casos la causalidad estadística, que la jurisprudencia

española y alemana ha utilizado ya en algunos casos, junto con presunciones *iuris tantum* bien fundamentadas.

c) En esta fenomenología también *se replica la íntima relación entre la imputación objetiva, el riesgo(no) permitido y la imprudencia*, a través de la ecuación infracción de la norma de cuidado/descontrol del riesgo imputable objetivamente/riesgo no permitido.

d) Considero que *hay que realizar ajustes en la previsibilidad versus imprevisibilidad como elemento de la imprudencia*, para que no se convierta la falta de previsibilidad en un "coladero" para las defensas, al que se acuda de manera recurrente, convirtiéndose en un automatismo, argumentando que hay un elevado nivel coyuntural de imprevisibilidad, debido a las características de la opacidad epistémica de los sistemas de IA. En el trabajo se encuentra un intento de reforzamiento de la previsibilidad a través de dos correctivos. El primero se sustenta en el deber de *alcanzar estándares eficaces de previsibilidad* en la fase de prueba y en la fase de poscomercialización. En la fase de prueba se deben realizar entrenamientos con validaciones y pruebas de seguridad de alta calidad según el estado de la técnica con el objetivo de prever eficazmente el comportamiento del sistema; y en una segunda fase de poscomercialización se debe seguir vigilando y controlando el comportamiento del sistema durante todo su ciclo de vida. Se trata de acumular el conocimiento suficiente y la experiencia para anticipar situaciones críticas. El segundo correctivo exige que no se realicen trabajos de experimentación y, en ningún caso, pasar a la fase de comercialización de un sistema que no tenga a su vez la *tecnología insertada para devolver el control al ser humano en situaciones críticas*. En estos supuestos no se podrá alegar imprevisibilidad del curso causal, pues se conoce de antemano las deficiencias de control y de seguridad.

e) *La autoría está reservada a los sujetos obligados de la cadena de valor*, por lo que estamos ante una lista cerrada de autores, lo que nos sitúa en las reglas de la autoría de los delitos especiales. Ahora bien, este círculo de autores se abrirá, como ocurre en otras actividades empresariales o no, con división estructural del trabajo, por medio de la delegación. Las formas de autoría, en las que normalmente no habrá un acuerdo en común, podrán ser la coautoría sucesiva, aditiva, accesoria o yuxtapuesta, o incluso autoría en cadena, sin olvidar la aplicación en su caso de la concurrencia de imprudencias. Por último, estos sujetos obligados serán autores, sin que haya problema para sostener una autoría directa, la cual no se debe entender como de propia mano y espaciotemporalmente

unida a la máquina, a la "cosa" o al sistema de IA, que resulta ser el instrumento, con características especiales, que deben tenerse en cuenta para reformular esa autoría directa.

f) Continuando en el ámbito de la autoría, en el trabajo se ha planteado la posibilidad de que *la persona electrónica pueda ser sujeto activo del delito*, y la respuesta se inclina hacia el rechazo de esta posibilidad, por los mismos viejos obstáculos que se han utilizado para oponerse a que la persona jurídica sea sujeto activo del delito, aunque reconozco que en esta sede el sistema de IA toma decisiones (aparentemente) autónomas. En cualquier caso, considero que es difícil admitir que la IA actúe con dolo, porque por ahora no es capaz de pensar, reflexionar y no tiene representaciones con significado. Tampoco es posible afirmar que las personas electrónicas tengan culpabilidad y sean imputables. No pueden captar el mensaje preventivo y retributivo de la norma penal. En el test de la culpabilidad queda al descubierto que son máquinas. Por otro lado, las penas que se le puedan imponer a un sistema de IA. —no van a resarcir a la víctima, que no va a identificar a la máquina como responsable de su daño y de su sufrimiento.

g) Por otro lado, los sujetos obligados del RIA son garantes de la cadena de valor. Esta característica también conduce a sostener la naturaleza especial del delito imprudente que cometan, debido a que son sujetos con deberes especiales concretos, cuyo objetivo es la protección de los bienes jurídicos. Estos garantes son "deudores de seguridad" y en la mayoría de los casos son garantes de evitación de descontrol del riesgo, por lo que realizarán delitos imprudentes de omisión pura de garante, que podrán ser elevados por imperativo legal a comisiones omisivas, pese a que no realicen con su omisión un delito de resultado. Parte de la doctrina también podrá defender que estamos ante delitos de deber *(Pflichtdelikte)* por la cualidad de garante del sujeto activo. Por otro lado, siguiendo la regulación europea se plantea una responsabilidad por gestión del riesgo atendiendo al binomio posición de garante/riesgo permitido. Por último, el Derecho penal a través de la imposición de una pena personal, intransferible y grave a un garante refuerza el principio de supervisión humana, que resulta ser un pilar irrenunciable. En definitiva, se trata de que el garante vele por que transcurra todo el uso del sistema de IA durante su ciclo de vida dentro del riesgo permitido y no se pierda el control sobre la máquina, ni en la fase de experimentación, ni en la fase de uso de sistemas de IA.

h) Es irremediable en este orden de cosas reparar en los programas de cumplimiento, porque quedan completamente comprometidos por toda la regulación del RIA. Para enfocar el tema he

considerado necesario plantear la cuestión de si el *compliance* penal debe hacerse atendiendo a la naturaleza de instrumento de la IA o, por el contrario, otorgando a la IA el estatuto de persona, en este caso persona electrónica, llegando a la conclusión de que de nuevo toda la responsabilidad penal de la persona electrónica tendría que basarse en las actuaciones típicas y antijurídicas de la persona física, por lo que otorgar personalidad jurídica a la persona electrónica no tiene un rendimiento efectivo para declarar la responsabilidad penal, moviéndonos de nuevo en el terreno de la ficción, al igual que ocurre con la persona jurídica, que su responsabilidad penal depende de lo que realice la persona física. Por otro lado, se impone una adaptación de los programas de cumplimiento a los requisitos y obligaciones del RIA, para que, como efecto más prometedor en el Derecho penal, excluya la responsabilidad penal de la persona jurídica y arrastre favorablemente al terreno de la impunidad, en algunos casos, también a sus representantes.

i) Pasando a la parte especial, apunto la necesidad de revisar algunos delitos para adecuarlos al uso de la IA como herramienta delictiva, así como proteger a la IA de los daños que se puedan producir en su código o en su programación y uso y por último los cambios que se pueden plantear en la responsabilidad por el producto y en los delitos de homicidio y lesiones.

Por último, quisiera reivindicar desde esta tribuna la necesidad de una alfabetización digital y sobre el uso de la IA, que ofrezca una síntesis de sus beneficios y de sus riesgos, de su regulación y de la defensa y protección del usuario, porque esta tecnología es neutra, puede ser nuestro *coworker* sin sustituirnos, facilitando la vida, o puede ser nuestro enemigo, todo dependerá de lo que hagamos con ella. El científico computacional y psicólogo cognitivo Geoffrey Hinton, Premio Nobel de física en 2024 y conocido como el "padrino" de la IA, renunció a su puesto en Google y denunció los peligros de la IA para la humanidad[91], así como estos peligros se pueden hacer reales dentro de menos tiempo que el pensado inicialmente, azuzados y favorecidos además por la competencia voraz entre Microsoft, Google y otras tecnológicas para llegar el primero en esta carrera hacia al abismo nuclear tecnológico, inimaginable para Oppenheimer. El progreso siempre se ha saldado con costes

[91] Cfr. https://www.bbc.com/mundo/articles/c8elg9j489eo [Enlace verificado 12-10-2024]. Explica y advierte el científico en esta entrevista de los riesgos de la IA, como que se desarrollen armas verdaderamente autónomas, de que la IA nos invade de noticias y fotos falsas en internet, que revolucione el mercado laboral y quite más que el trabajo pesado…y de una mala práctica como es la de que las personas y las empresas dejen que la IA genere su propio código y lo ejecute por su cuenta.

sociales y desgraciadamente vidas, pero en esta ocasión los costes pueden ser impactantes y explosivos para la supervivencia humana.[92] Discutir sobre Derecho penal y sobre la responsabilidad de la persona física, persona jurídica y persona electrónica se torna baladí, irrisorio y ridículo si perdemos el control sobre estas "cosas". No podemos perderlo, nos va la existencia de la humanidad en ello.[93]

VII. Conclusión

La irrupción de la IA es un hito dentro del avance tecnológico y de nuestro proceso evolutivo. Un fenómeno que está cambiando nuestro presente y que nos conduce a un futuro, que solo era imaginado en la ciencia ficción, en el que las máquinas cobrarán un relevante protagonismo. No obstante, la IA reciba el nombre de computación cognitiva, aprendizaje mecánico o profundo, redes neuronales o inteligencia aumentada, actualmente no deja de ser más que una herramienta puesta al servicio de la humanidad. Una herramienta muy potente, como lo fue en su día la llegada del ferrocarril. Ahora bien, la diferencia sustancial de esta cuarta revolución frente a las anteriores estriba en que se está emulando en la máquina las suficiencias de la inteligencia humana, hasta la fecha privativas del ser humano. La IA es una devoradora de *datos*, millones de datos, analizados por el *algoritmo* a velocidades inalcanzables para el cerebro humano y que se incrementará exponencialmente con la computación cuántica, por lo que tendremos que dar el salto hacia el dato sintético para alimentar la voracidad de la IA, cuando haya terminado con todos los datos reales. En definitiva, nosotros mismos estamos fabricando un competidor por el control y el dominio de nuestra vida.

Dicho esto, ya a nadie se le escapa que el uso de esta herramienta, como de cualquier otra, debe ser un *uso responsable*, lo que implica más que nunca el control y la supervisión humana. Ello supone

---

[92] En el mismo sentido, BOIX PALOP, "Los algoritmos son reglamentos: la necesidad de extender las garantías propias de las normas reglamentarias a los programas empleados por la administración para la adopción de decisiones", en *Revista de Derecho público: teoría y método*, n.º 1, 2020, p. 227, que sostiene que la magnitud del riesgo puede tener consecuencias extintivas para la especie humana.

[93] Éric Sandin (Cfr. Disponible en: https://www.infobae.com/cultura/2020/07/25/eric-sadin-el-intelectual-que-desafia-al-reino-del-algoritmo-la-innovacion-digital-trabaja-para-convertir-toda-actividad-humana-en-obsoleta/, [Enlace verificado: 12 de noviembre 2024] alerta sobre la IA y como su presencia va a deteriorar la capacidad cognitiva humana, hasta el punto de no tomar decisiones soberanas y no ser libre, dando paso a una nueva forma de antropología de la relación del ser humano con la tecnología y el poder de decisión. «"Vivimos un momento serio, de gran gravedad, pero no lo vemos", ha expresado, refiriéndose al impacto de las tecnologías digitales desarrolladas en las últimas dos décadas y a la velocidad vertiginosa con la que están transformando la vida cotidiana, lo que dificulta comprender el presente de manera clara e inmediata.» (Cfr. Disponible en: https://elpais.com/proyecto-tendencias/2024-11-11/eric-sadin-filosofo-la-ia-modificara-la-naturaleza-del-ser-humano.html.

un desafío sobre todo tratándose de la IA generativa combinada con la aspiración humana de crear una superinteligencia capaz de programarse a sí misma, conectarse con otras IAs y ser capaz de interpretar las emociones. El Derecho penal debe contribuir a que esta re-volución tecnológica no se re-vuelva contra su creador y nos mantengamos dentro de los esquemas del uso responsable como objetivo ubicuo y permanente dentro del diseño de una responsabilidad jurídica integral. Para ello será necesario una reinterpretación adaptativa de categorías jurídico-penales, como he intentado demostrar en este trabajo como primera aproximación, lo que no conllevará para este Derecho penal dual, Derecho penal analógico/Derecho penal digital, la renuncia a la solidez conceptual de la teoría jurídica del delito basada en garantías y límites. Nos encontramos en estos momentos en un periodo de transición y tendremos que seguir avanzando por este camino, que implica además a todos los Estados, porque el uso de la IA eleva las cuestiones jurídicas a planos internacionales y de Derecho penal comparado.[94]

Bibliografía

ABADÍAS SELMA, Alfredo, "La protección penal de los trabajadores frente a la inteligencia artificial en el ámbito del delito de discriminación laboral", en *Revista de Derecho Penal*, n.º 39, 2023.

BALDWIN, Richard, *La convulsión globótica. Globalización, robótica y el futuro del trabajo*, Barcelona, Bosch, 2019.

BAUMANN, Zygmunt, *Modernidad líquida*, Fondo de Cultura económica, 2002.

BELZUZ ABOGADOS, "Algunas notas sobre la propuesta de Directiva sobre responsabilidad en materia de Inteligencia Artificial", disponible en: https://www.belzuz.net/es/publicaciones/en-espanol/item/12068-algunas-notas-sobre-la-propuesta-de-directiva-sobre-responsabilidad-en-materia-de-inteligencia-artificial.html [Enlace verificado 8-10-2024].

BOIX PALOP, Andrés "Los algoritmos son reglamentos: la necesidad de extender las garantías propias de las normas reglamentarias a los programas empleados por la administración para la adopción de decisiones", en *Revista de Derecho público: teoría y método,* n.º 1, 2020.

[94] Cfr. HILGENDORF, *supra* nota 17, p. 40.

COCA VILA, Ivó, "Coches autopilotados en situaciones de necesidad. Una aproximación desde la teoría de la justificación penal", en *Cuadernos de Política Criminal*, n.º 122, 2017.

DOPICO GÓMEZ-ALLER, Jacobo, "¿Dogmática o probatoria? La cuestión de la causalidad y su prueba en las intoxicaciones masivas con agentes tóxicos desconocidos", en *La Ley Penal*, n.º 169, 2024.

FERNÁNDEZ HERNÁNDEZ, Carlos, "La comisión presenta una propuesta de Directiva sobre responsabilidad civil extracontractual en materia de IA", en *Diario La Ley*, n.º 63, 29 de septiembre de 2022, Sección Ciberderecho, disponible en: https://diariolaley.laleynext.es/Content/Documento.aspx?params=H4sIAAAAAAAAEADVPwWrDMAz9mvk4nDUj68GXtDsURildGLsqjnAMrpXZclb__bR2EzzEE0_vSV8FUx3wyiajK4lUrpFivZghFVQMYzab9qGzWtAInhRYLhD2ZE3zrNtf6lccYBSdojRh6qvRiokhnDGbRm-7rcozfR9h9Q7YU-wh3f39NJnXTy3VdJuufVErpiwC8-EdRkY1eze_CfiuzwjJzidwaA7RW0-PkJfr37gvzLI6cny_cWWD9D0w7iBgnP5DYVlCPVOQ4258oSzflIsEHuIOEpWMwegfrZclgRoBAAA=WK [Enlace verificado 8-10-2024].

GONZÁLEZ BELUCHE, Paloma, "La adaptación de la Directiva 85/374/CEE, de 25 de julio, en materia de responsabilidad por daños causados por productos defectuosos a la cuarta revolución industrial", en Cuadernos de Derecho Transnacional, vol. 15, n.º 2, 2023.

GRECO, Luis "Vehículos de motor autónomos y situaciones de colisión", en BASSO/CANCIO/MARAVER/FAKHOURI (coord.) *LH prof. Dr. Agustín Jorge Barreiro*, 2019.

HILGENDORF, Eric, "Können Roboter schuldhaft handeln?" en: BECK, S. (ed.) *Jenseits von Mensch und Maschine: Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs*, Baden-Baden, Nomos, 2012.

– "Dilemma-Probleme beim automatisieren Fahren. Ein Beitrag zum Problem des Verrechnungsverborts im Zeitalter der Digitaliserung", en *ZStW*, 2018.

– "Vom Werkzeug zum Partner?, Zum Einfluss intelligenter Artefakte auf unsere sozialen Normen und die Aufgaben des Rechts", en ENGELHART/KUDLICH/VOGEL (ed.), *Digitalisierung, Globalisierung und Risikoprävention, FS-Sieber*, Berlín, Duncker&Humblot, 2022.

– "Inteligencia artificial y Derecho penal", en *Desafíos penales de hoy: Entre la ley y la justicia en la obra de Eric Hilgendorf*, trad. Leandro Dias Lestón, Buenos Aires, Editores del Sur, 2024.

HÖRNLE, Tatiana/WOHLERS, Wolfgang "The Trolley Problem Reloaded, "¿Cómo deben programarse los vehículos autónomos para los dilemas de ¨vida contra vida"?", trad. José Béguelin, en HÖRNLE, *Criminalización, castigo, y dilemas morales en la obra de Tatjana Hörnle*, Buenos Aires, Editores del Sur, 2022.

IBOLD, Victoria, Künstliche Intelligenz und Strafrecht. Zur strafrechtliche Produktverantwortung in der Innovationsgesellschaft, Baden-Baden, Nomos, 2024.

VON JHERING, *La lucha por el Derecho*, (Estudio preliminar y edición de Luis Lloredo Alix), Madrid, Dykinson, 2018.

KIR, *Künstliche Intelligenz und Recht*. La primera revista sobre IA y Derecho, Beck (ed.)

MORÁN ESPINOSA, "Responsabilidad penal de la Inteligencia artificial (IA). ¿La próxima frontera?", en *Revista del Instituto de Ciencias Jurídicas de Puebla*, vol. n.º 15, n.º 48, 2021.

MÉNDEZ SERRANO, "Derechos fundamentales y personalidad jurídica de los robots: ¿para qué?", en *Derecho privado y Constitución*, n.º 44, 2024.

OBREGÓN FERNÁNDEZ, Aritz/LAZCOZ MORATINOS, Guillermo, "La supervisión humana de los sistemas de inteligencia artificial de alto riesgo. Aportaciones desde el Derecho Internacional humanitario y el Derecho de la Unión Europea", *Revista electrónica de Estudios Internacionales*, n.º 42, 2021.

PANATTONI, Beatrice/PICOTTI, Lorenzo, "Traditional Criminal Law Categories and AI: Crisis or Palingenesis?", International Colloquium Section I, Siracusa, 15-16 September 2022, en *Revue Internationale de Droit Pénal*, vol. n.º 94, n.º 1, 2023.

PAREDES CASTAÑÓN, José Manuel, "Responsabilidad penal por productos defectuosos", en *Revista Fundación internacional de ciencias penales, Responsabilidad penal por productos defectuosos*, n.º 2024-2.

PÉREZ MANZANO, Mercedes, "Algunos datos empíricos sobre la atribución de estados mentales: ¿fracaso del principio de responsabilidad subjetiva o de un determinado concepto de dolo?", en *Revista electrónica de Ciencia penal y criminología*, n.º 23-15, 2021.

PERRIGO, Billy. "Exclusive: OpenAi Lobbied the E.U. to Water Down AI Regultion, Revista *Time*, 20-6-2023, disponible en: https://time.com/6288245/openai-eu-lobbying-ai-act/ [Enlace verificado 7-10-2024].

QUINTERO OLIVARES, Gonzalo, "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*, n.º 1, 2017.

– "Populismo y Derecho penal", en *DOXA*, n.º 48, 2024.

ROSO CAÑADILLAS, Raquel, "Los delitos polivalentes de autoría: entre el deber y el dominio", en *InDret,* n.º 3, 2019.

– "La necesidad de diferenciar entre autoría y participación imprudente y la cuestión de su punibilidad", en *Foro Fundación Internacional de ciencias penales*, n.º 2022-3.

– "¿Un Derecho penal delicuescente en una sociedad líquida?", en *Anuario de la Facultad de Derecho de la Universidad de Alcalá,* n.º 26, 2023, pp. 199-225.

– ¿Un Derecho penal delicuescente en una sociedad líquida? Algunas reflexiones sobre el Derecho penal en la sociedad posindustrial", en *Revista General de Derecho penal*, n.º 41, 2024, pp. 1-30.

SANDIN, Éric, entrevista, Infobae, Disponible en: https://www.infobae.com/cultura/2020/07/25/eric-sadin-el-intelectual-que-desafia-al-reino-del-algoritmo-la-innovacion-digital-trabaja-para-convertir-toda-actividad-humana-en-obsoleta/, [Enlace verificado: 12 de noviembre 2024]

SILVA SÁNCHEZ, Jesús María, "Para muestra…el caso Boeing", en *InDret*, n.º 3, 2024.

SCHWAB, Klaus, *La cuarta revolución industrial*, Madrid, Marcial Pons, 2016.

SUARÉZ, María Florencia, "Inteligencia artificial y Derecho penal. El dilema del tranvía. Cuarta Revolución industrial. Ética del Algoritmo. IA en vehículos. Causas de justificación", en *Revista Pensamiento Penal*, n.º 445, 2022.

VELASCO, Eloy, entrevista por Andrés Garvi Carvajal, "Robots delincuentes, respuestas jurídicas a un futuro cercano", Disponible en https://cincodias.elpais.com/legal/2022/02/03/juridico/1643900957_593967.html, [Enlace verificado 4-10-2024].

– "Delitos tecnológicos de los informáticos a los cometidos por la IA" *Conferencia de clausura en la 7.ª edición del Máster de Experto de Derecho digital de la Universidad de Deusto*, impartida el 22-03-2024. Disponible en: https://www.youtube.com/watch?v=uGdOmQ_sQVg [Enlace verificado 4-10-2024].

WELZEL, Hans, "Zum Notstandproblem", en *ZStW*, n.º 63, 1951.