REVISTA DE DERECHO PENAL Y CRIMINOLOGÍA, 3.ª Época, n.º 33 (julio 2025), págs. 435-446

UNA RAZONABLE MODERADA ENTRE LAS FILAS DEL FRENTE DE LIBERACIÓN ROBÓTICO

RECENSIÓN A LA MONOGRAFÍA DE ALICE GIANNINI "CRIMINAL BEHAVIOR AND ACCOUNTABILITY OF ARTIFICIAL INTELLIGENCE SYSTEMS". MAASTRICHT, ELEVEN PUBLISHERS, 2023

Andrea Bravo Bolado

Universidad Autónoma de Madrid

I. Introducción

La obra que aquí se comenta se corresponde con el fruto de la tesis doctoral de la profesora Alice Giannini, publicada en lengua inglesa por la editorial Eleven dentro de la colección Maastricht Law Series en 2023. La autora demuestra que, pese a su corta trayectoria académica, la investigación que le valió el doble título de doctora por la Universidad de Maastricht y por la Universidad de Florencia merece la pena ser leída con atención en unos tiempos en los que parece que es de obligada referencia una mención a la potencialidad de las tecnologías para cambiar todo aquello que habíamos dado por sentado, inmersos como estamos en una sociedad donde la IA lo copa todo.

Precisamente por la actualidad, complejidad y ubicuidad de la discusión sobre la llamada cuarta revolución industrial y su afectación al derecho penal, es esperanzador encontrar obras como esta, que le dedican el tiempo y la seriedad necesaria a un tema que ha sido abordado, por el momento, de forma fragmentaria, principalmente a base de *papers* que entremezclan conceptos de tradiciones jurídicas bien diferenciadas. Este sería, por ello, el primero de los méritos a destacar de la obra. La autora, que se retrata como buena conocedora de los sistemas jurídicos

más influyentes en la materia, realiza una revisión de la literatura en tres lenguas —inglés, italiano y alemán— y lo hace poniendo a conversar a autores que, cada uno en su contexto, habían venido manteniendo discusiones alejadas y paralelas. Permite Giannini al lector una inmersión en la discusión dogmática más actual, poniendo al servicio de la investigación un diálogo que venía siendo necesario, con una exposición magistralmente clara y pedagógica. La actualidad del problema queda sintéticamente retratada cuando apenas en la primera página la autora pone a disposición del lector una base de datos donde se recopilan los incidentes causados por sistemas de IA, que se cuentan por cientos, afectando a sectores muy diversos en todas las partes del globo.

Para estructurar la presente recensión se articulará la misma en tres secciones. En primer lugar, se introducirán los problemas centrales que la obra trata, haciendo hincapié en aquello que a nuestro juicio puede aportar más valor, para más adelante desarrollar las ideas importantes en mayor detalle, y finalmente acabar comentando las conclusiones y los puntos débiles de los que, a nuestro parecer, adolece.

II. Las tesis centrales

La pregunta que vertebra la investigación se concreta ya en la p. 4 (*¿to what extent is a theoretical framework of criminal law for liability of non-human agents needed and feasible*?). La autora se propone explorar un campo sin duda polémico, como es el de la posible adscripción de responsabilidad penal a un sistema de IA en aquellos casos en que el delito se comete, al menos funcionalmente, por una máquina, sin que sea sencillo imputar la responsabilidad a un humano. El tópico ha sido tratado por diferentes autores, pero pocos han realizado un trabajo dogmático y comparativo tan completo. En el libro se desarrollan tres grandes subtemas alrededor de esta pregunta inicial a lo largo de ocho capítulos. El capítulo 1, a modo introductorio, explica cuáles son los grandes problemas de fondo y la estructura de la investigación. Estos grandes temas son: la cuestión de la personalidad electrónica de los sistemas y la cuestión relativa a la adscripción del *actus reus* y el *mens rea* a la propia máquina. Además de este núcleo, se tratan otros problemas adyacentes y necesarios. Así; comenzando por la siempre polémica definición de IA (capítulo 2), pasa a realizar un repaso sobre el debate actual entre la doctrina (capítulo 3), para continuar, tras una breve contextualización (capítulo 4), con el debate sobre la adscripción de la capacidad de ser sujeto para el derecho penal (capítulo 5), entrando ya en el siguiente capítulo (6) en el núcleo de la obra, donde se pregunta por la posibilidad de imputar el *actus reus* y el *mens rea* a la propia máquina. La investigación se cierra con un capítulo dedicado a examinar algunos de los sistemas legales existentes (capítulo 7), para concluir con un capítulo en el que resume las conclusiones alcanzadas, dando respuesta a la pregunta que

guía la investigación, proponiendo una serie de retos a los investigadores del futuro (capítulo 8). La metodología adoptada, que utiliza como base el esquema *actus reus-mens rea*, se explica por la autora no como deferencia al sistema anglosajón, sino porque presenta una base común entre ambas culturas, constituyendo una base fértil para un análisis adecuado en un problema de alcance transfronterizo.

III. Desarrollo de las ideas fuerza de la obra

La cuestión con la que se inicia el trabajo no es, por obvia, menos esencial. La definición del fenómeno sobre el que trata el libro ha venido comportando problemas en cada una de sus parcelas de estudio. El derecho penal, señala Giannini, requiere de una seguridad jurídica de la que no goza cuando se trata de aproximarse al fenómeno de la IA, donde cada disciplina tiene algo que decir. Aunque la autora no trata de solucionar por completo una cuestión tan compleja, sí que explica de forma clara cuáles son los mecanismos que hoy en día definen la tecnología más innovadora, acudiendo a símiles y metáforas que hacen entender hasta al más profano una materia de enorme complejidad. Aboga finalmente por acoger la definición que el Grupo de Expertos de Alto Nivel propuso en 2018, definición que, a nuestro parecer, es la más completa y adecuada. Se pone el énfasis, así, no en la similitud de estas tecnologías con la inteligencia humana, desterrando una concepción erradamente antropomórfica de estos sistemas (herencia, probablemente, de los primeros textos de Turing y Asimov), sino una definición más a la Russell y Norvig, donde lo importante es la racionalidad o la funcionalidad de la máquina. Se destaca que la definición del Grupo de Expertos no solo reúne los elementos más importantes según la mayoría de la doctrina, sino que lo hace evitando ambigüedades que hacen referencia a la inteligencia humana como punto de referencia. Concluye, así, que cuando se hable de IA en el trabajo se hará refiriéndose únicamente a los programas en los que se da un comportamiento adaptativo (es decir, aquellos que han podido desarrollar las capacidades derivadas del *machine learning*), cuyas acciones no sean *ex ante* predecibles.

El siguiente capítulo constituye uno de los activos más valiosos para todo aquel que tenga un interés en conocer el estado de la discusión dentro de la dogmática jurídica continental y anglosajona, sistematizando la autora el pensamiento de los autores contemporáneos en tres grandes grupos: los expansionistas, los moderados y los escépticos.

Dentro del bautizado como el expansionista "Frente de Liberación Robótica", la referencia es obligatoria a los ya clásicos textos de Hallevy, expansionista por excelencia, en el que la autora reconoce toda una serie de carencias dogmáticas. Más crédito le otorga, sin embargo, a Ying Hu, que hace un esfuerzo por argumentar por qué cabría un derecho penal

que aplicara solo a los robots más avanzados (capaces de acción moral y autonomía en un sentido colectivo). Con Mulligang y Quark nos presenta concepciones que aluden a la finalidad del castigo directo al robot como forma de satisfacción psicológica, así como las primeras referencias a la posibilidad de responsabilidad colectiva. El segundo grupo —los moderados—, son clasificados en esta categoría porque pretenden modificar solo algunos aspectos de la ley sin que esto llegue a significar un cambio disruptivo para el derecho penal. Se resume aquí el pensamiento de autores que, basándose en la tradición anglosajona, proponen algunas teorías imaginativas o basadas en modelos anteriores, que van desde la aplicación de delitos de peligro, modelos de *welfare offences* o enfoques basados en la responsabilidad objetiva empresarial. Cabe destacar de entre este grupo a las alemanas Simmler y Markwalder, únicas autoras estrictamente penalistas que, con una concepción marcadamente *jakobsiana*, conciben la personalidad como constructo social, poniendo énfasis en la capacidad de cuestionar expectativas normativas como el núcleo del derecho penal; solo en el momento en que los robots tengan reconocida esta capacidad socialmente podrá considerarse su entrada en el mapa del derecho penal. Finalmente, ya en la categoría de los escépticos, Giannini nos presenta a académicos italianos y alemanes, donde, claramente, la tradición jurídica continental apegada a la tradición humanista y a la máxima *nulla poena sine culpa* ve su reflejo. Aunque la discusión en italiano ha estado caracterizada por la fragmentación de los problemas, hace la autora un esfuerzo por combinar todos ellos, reflexionando sobre tópicos comunes como qué utilidad puede tener el castigo para un ente que tiene un "cuerpo que patear pero ningún alma que dañar" (aludiendo al ya clásico texto de Asaro), el papel del humano y su posible culpa o negligencia en el proceso de programación, así como la necesidad de delimitar el riesgo permitido en los casos en que el humano pierde el control. Del lado alemán, cuyo debate tiende a estar cerrado sobre sus propias fronteras, trae a colación trabajos que ponen el foco en la necesaria atención al delito imprudente cometido por el humano de detrás, realizando una interesante reflexión sobre cómo la falta de predictibilidad de algunos sistemas puede ser problemática, destacando que la verdadera tarea pendiente en este campo es la determinación del riesgo permitido, creando, quizá, deberes específicos de vigilancia ("*sleeping obligations*" o "*schlafenden Ingerrenz*"), destacando que es importante reflexionar sobre el nuevo concepto de imprudencia que pueda surgir al albur de la sociedad del riesgo. Si algo aporta la lectura del capítulo es la posibilidad de entender qué problemas preocupan y comparar cómo se abordan desde diferentes perspectivas jurídicas, pudiendo entreverse fácilmente cómo las posibilidades son marcadamente diferentes en los sistemas de *common law* y de *civil law*, lo que sirve para reafirmar algunas tendencias y tradiciones que nuestros sistemas se niegan a dejar escapar.

De toda la exposición razonada y ordenada de sus diferentes frentes, Giannini deriva diez preguntas y siete lagunas, de las cuales merece la

pena destacar la primera de ellas, por constituir, entendemos, eje esencial de sus posteriores pesquisas. Critica la autora, seguramente con razón, que la doctrina mayoritaria ha descartado desde un inicio todos los argumentos de Hallevy sin apenas esfuerzo, tan solo apelando de una forma tajante a que la responsabilidad penal solo le corresponde a la persona, pero lo cierto es que se echa de menos precisamente una profundización seria en los argumentos por los cuales esto no debe ser así. A esta reflexión, muy necesaria si es que se quiere tomar como punto de partida el del equipo de los escépticos, dedica los siguientes capítulos.

Entramos así, de lleno, en el núcleo de su investigación, donde realmente se pregunta qué implica ser sujeto para el derecho penal. El valioso trabajo del capítulo 5, titulado "*criminal capacity*", es que la autora se cuestiona si los sistemas de IA pueden desplegar, bajo alguna interpretación posible, las capacidades que normalmente se entienden como suficientes para adscribir responsabilidad penal. La pregunta esencial es, pues, ¿cuáles son estas capacidades? Se ahonda esta vez en la dogmática netamente penal para tratar de entender lo que significa la capacidad de estar sujeto al derecho penal, haciendo alusión a los elementos que en los diferentes ordenamientos configuran esta capacidad de imputación. La conclusión a la que llega es que en todos los sistemas que analiza (que incluyen los códigos penales alemán e italiano, el *Model Penal Code* americano y jurisprudencia del *common law*), existen al menos dos elementos requeridos: las capacidades cognitivas o epistémicas y las capacidades de control o, si se quiere, las capacidades de conocer y querer los actos propios. El problema que señala la autora es que no sabemos bien si este entendimiento de la ilicitud debe otorgar más peso a la ilicitud en sentido legal o en sentido moral. Se desarrollan, a continuación, algunas ideas sobre cómo las máquinas han sido puestas a prueba, programándolas para entender ciertas normas (ilicitud legal), como es el caso del código de circulación en el caso de programación de vehículos autónomos, o para entender ciertos valores (ilicitud moral), con el paradigma de los denominados *Artificial Moral Agents* a los que se somete a programación para solventar diferentes dilemas morales. Señala la autora los inconvenientes que se han hecho patentes con sendos intentos, pues la codificación de la licitud (ya sea moral o legal) en términos matemáticos adolece de complejidades no solo técnicas sino de tipo ético (afirma la autora, en p. 137, que no tenemos un "sistema óptimo de ética"). Solo una vez superado este primer escalón epistemológico (cuyas dificultades no se ocultan), podrá examinarse el siguiente escalón, relativo al control o voluntariedad sobre las acciones. En este punto la autora menciona, sin profundizar demasiado, la polémica relativa al libre albedrío, para exponer que lo importante no es tanto probar si existe un fenómeno como la libertad de voluntad, pues el énfasis no debe ponerse en la dimensión interna (¿cómo se configura realmente esta voluntad?), sino en la externa; es decir, cómo se percibe desde fuera, tanto por el sistema como por la sociedad, esta capacidad de actuar moral. En definitiva, el

sistema no percibe a la IA como agente dotado de capacidad de acción moral, y tampoco podemos concebir que esté dotado de entendimiento epistemológico. Sin esta capacidad de entender la ilicitud de las acciones o actuar conforme a ella, difícilmente tendremos a un sujeto pleno para el derecho penal. Se echa en falta, empero, en esta sección, una mayor profundización en los argumentos relativos al libre albedrío, tópico harto discutido en esta sede y apenas tratado por la autora.

Concluye el capítulo haciendo una última referencia: incluso aunque pudiera afirmarse que los sistemas son capaces de conocer su entorno, las normas y sus acciones, y actuar conforme a un entendimiento de la legalidad y la moralidad que ha sido en ellos programada, no solo faltaría el reconocimiento externo de que la conducta que realizan la realizan por su propia voluntad, sino que, además, la cuestión de la personalidad para el derecho penal no puede escindirse de la pregunta acerca de la finalidad del castigo.

Pasa entonces la autora a desarrollar las posibilidades de adscribir el *actus reus* y *mens rea* al sistema. Comenzando por los problemas relativos al *mens rea*, diferencia dos posibles modelos; el modelo de culpabilidad de la máquina y el de la culpabilidad del humano de detrás de la máquina. En cuanto a la primera opción, reconoce que los autores que abogan por aceptar una teoría que moldee los elementos del *mens rea* para poder afirmarlos en la máquina son pocos, pero no deja de mencionar sus razonamientos, aun reconociendo las dificultades que entrañan, quizá otorgándoles una importancia que, a nuestro juicio, no debieran tener. En esencia, las construcciones que se presentan parecen confundir la causalidad con la intencionalidad («*machine intent is 'presumably perfectly observable (assuming some access to the algorithm)'*»), p. 157), pero tiene lógica si entendemos que las teorías de las que se parte pretenden una formulación "formal" de la intención, intentando desentrañar un concepto valorativo desde una perspectiva eminentemente técnica.

Pasando a analizar el modelo de responsabilidad del humano, la autora se centra tan solo en el escenario de los delitos imprudentes, pues considera, con acierto, que los delitos dolosos no suponen ningún reto dogmático. Así, centra la discusión en torno al concepto de imprudencia (*negligence*), señalando la dificultad de que en cada sistema jurídico tome una forma. En todo caso, analiza una vez más su mínimo común denominador, condensándolo en dos elementos: la previsibilidad (*foreseeability*) y el incumplimiento de un deber de diligencia razonable (*breach of a duty of reasonable care*), p. 160, señalando la doble vertiente (objetiva y subjetiva) de la misma. Es reseñable la reflexión que realiza sobre el modelo que parece estar adoptando el Parlamento Europeo en lo relativo a la responsabilidad del humano de detrás de la máquina, pues pone de manifiesto algunas preocupaciones que suscribe la autora de estas páginas. Si el modelo correcto para afrontar los riesgos derivados de posibles accidentes es el del humano "*in the loop*", que desempeña un "control

humano significativo", debemos estar muy atentos a que este pretendido control pueda ejercerse en la realidad, y no solo en el papel, pues, con las mayores capacidades de la máquina, será cada vez más difícil que el ser humano pueda controlar *de facto* los errores de un sistema que fue creado para superarlo. El riesgo de que el hombre de detrás se convierta en chivo expiatorio me parece uno de los problemas más actuales en la discusión europea sobre el manido control humano, pieza clave de la ética algorítmica actual. La pregunta que se hace Giannini, en p. 162, es acertada: *"How will a human, in practice, be able to supervise a system that was created to overcome it and make up for her shortcomings?"* Yerra, sin embargo, a mi parecer, cuando concluye con rotundidad que no existe, por el momento, un conocimiento técnico-científico necesario para aplicar reglas de conducta en los nuevos escenarios, pues a mi parecer, el trabajo realizado por las instituciones europeas a lo largo de todo el proceso de legislación que ha dado lugar al Reglamento de IA sirve perfectamente como un primer instrumento aproximativo a las nuevas normas de conducta y cuidado.

Lo que queda claro es que, como constata la autora, la definición de previsibilidad y los deberes de cuidado están cambiando con el objetivo de regular el riesgo. Estar atentos a esta evolución es el camino, considero, más adecuado, para abordar la problemática de los delitos imprudentes. Es de alabar que la autora no solo apunte los problemas, sino que aporte, también, una tímida solución, que comparte con Gless, Silverman o Weigend; solo deberá permitirse un riesgo enmarcado en actividades de IA que reporten un valor social, y esto deberá determinarse caso por caso. Los ejemplos dicotómicos que aporta la autora no dejan lugar a dudas (sí a los coches autónomos que salvan vidas, no a los juguetes sin valor social). Tememos, sin embargo, que el espectro de usos de la tecnología da lugar a aplicaciones en zonas mucho más grises, y se echa en falta la mención a usos con un significado ético no tan unánime.

Aunque señala, asimismo, la escasa o nula atención que se ha prestado a la tarea de individualizar qué humano debe responder de qué tarea o deber, encasillando a cada uno de los posibles roles en el más genérico "programador", tampoco realiza la autora ningún esfuerzo por concretar estos deberes, más allá de señalar la complejidad de la cadena productiva y la escasa atención que se presta a la gobernanza de datos (afirmación que, sin embargo, no podemos compartir del todo, habida cuenta de la atención que el reciente Reglamento de la UE ha prestado a la gobernanza de los datos como fuente de enorme responsabilidad administrativa, sirviendo como base para la conformación de deberes de cuidado que afecten al campo penal).

En cuanto a la asignación del *actus reus,* se aportación resulta, quizá, algo parca. Vuelve a resaltar la idea de que los autores más expansivos se basan en la idea de acción como mero movimiento externo, para contraponerlos con las teorías mayoritarias que exigen algún tipo de volunta-

riedad. Poca atención más dedica a uno de los temas más discutidos (si bien es cierto que el capítulo sobre la adscripción de "capacidad criminal" ahonda mucho más en la cuestión, aunque de forma indirecta). Sí ahonda un poco más en lo que llama, usando la terminología de Pagallo, "*failures of causation*", pues reivindica que los problemas sobre la causalidad no han sido demasiado estudiados entre la doctrina. La idea fuerza aquí es que, debido a diferentes fenómenos propios de la tecnología de IA, el esquema clásico de la causalidad se rompe, pues hay una multitud de factores que interactúan, haciendo cada vez más difícil identificar la causa legalmente relevante. No propone la autora solución alguna, por lo que debe entenderse que todas estas cuestiones hacen difícil asumir que una acción sea imputable a una IA. En el caso de las omisiones, que Giannini trata muy discretamente, se afirma que no debe abrirse la puerta a la comisión por omisión en estos escenarios, echándose de menos, una vez más, algo de profundización en el argumento.

Pasa, a continuación, la autora, a realizar un estudio sobre la posibilidad de imputar responsabilidad penal aplicando el esquema de responsabilidad de las personas jurídicas, para lo cual realiza un repaso por los distintos modelos existentes. Aunque reconoce que la teoría más completa es la desarrollada por Diamantis, su idea de la "*corporate mind*" (que implica reconocer al algoritmo como un trabajador más capaz de vincular a la empresa), que puede ser coherente en un sistema como el americano, resulta chocante en sistemas continentales donde la sola posibilidad de que la empresa responda penalmente es observada con recelo. Se refiere Giannini a los modelos italiano y alemán, pero cabe resaltar que algo parecido ocurre en España y Portugal donde, a pesar de tener ya bien asentados los modelos de responsabilidad penal corporativa, no faltan los ejemplos de voces en la doctrina que no acaban de ver convincentes estas asimilaciones para resolver los problemas de la IA (así, en España, Del Rosal, y en Portugal Aires de Sousa[1]).

En las conclusiones al capítulo sexto la autora deja entrever, por fin, su propia concepción, calificándose de moderada en cuanto a la posibilidad de explorar opciones teóricas para aplicar el sistema penal, pero confesándose escéptica en lo relativo a los fines del castigo, pues no se ve capaz de renunciar a la intrínseca finalidad retributiva (si bien no única) de la pena. Su conclusión es clara; el derecho penal no puede aplicarse de forma directa a los sistemas de IA, y no puede aplicarse por una cuestión relativa a la finalidad que persigue el castigo. Así, apoyándose en

[1] B. Del Rosal Blasco, "El modelo de la responsabilidad penal de las personas jurídicas para los daños punibles derivados del uso de la inteligencia artificial", *Revista electrónica de responsabilidad penal de personas jurídicas y compliance*, *2,* 2023. S. Aires de Sousa, "Não fui eu, foi a máquina": teoria do crime, responsabilidade e inteligência artificial", en A. Miranda Rodrigues, (Coord.), A *inteligência artificial no direito penal,* Coimbra, 2020, pp. 59-94.

un concepto de responsabilidad relacional que bebe de Duff y que acoge Danaher, afirma que ser responsable implica responder por algo ante alguien; los humanos buscan asignar una culpa de carácter retributivo a un agente que reconocen como capaz de soportar las consecuencias de la pena. Si en este campo de relaciones no encontramos un agente al que poder responsabilizar debidamente, debemos buscar esa responsabilidad en otro lugar (si ese otro lugar debe ser el orden civil o el administrativo, cuestión desde luego no baladí, no es algo que la autora llegue a concretar en su obra).

El penúltimo capítulo del libro trata de hacer un repaso por algunas de las propuestas normativas recientes que pueden afectar, si bien de modo tangencial, al objeto de la investigación. Se analizan hasta cinco instrumentos legales de diferente carácter y alcance; algunos todavía en forma de proposición, otros con fuerza vinculante; unos de carácter nacional, otros supranacionales. Algunos regulan cuestiones de carácter general, mientras otros se ajustan a una determinada realidad aplicativa (principalmente la conducción autónoma). Resultan de especial interés las propuestas de reforma en Singapur, que tratan de modificar diferentes aspectos de su Código Penal, con un alcance que va más allá de un solo campo de aplicación, optando por la tipificación de delitos de peligro de contenido muy amplio y con contornos, aún, poco definidos. Si algo se destaca del análisis que realiza la autora de la normativa seleccionada es, precisamente, esa falta de seguridad jurídica que se desprende de las redacciones legales. La inseguridad es patente en ambos grupos de textos; en el caso de la regulación más genérica (más allá de enunciar principios de corte general, como es el caso de las resoluciones del Consejo de Europa), el trabajo del gobierno de Singapur, tratando de evitar la impunidad en el caso de que se produzcan daños imprudentes de cualquier tipo y origen, fija una responsabilidad poco definida sin que existan, aún, unos deberes de cuidado predefinidos. En el caso de las regulaciones nacionales seleccionadas, aplicadas en el campo de la conducción de vehículos autónomos, la conclusión es la misma; en la búsqueda de una posible inmunidad para el conductor, los contornos sobre el papel que debe tener el usuario en calidad de supervisor de la máquina en caso de que esta falle no permiten afirmar, aún, que exista seguridad jurídica.

IV. Conclusiones y valoración

Concluye la autora su monografía con un capítulo final que le sirve para sintetizar todo lo anterior y lanzar algunas preguntas al lector. Giannini nos ofrece, finalmente, la respuesta a su pregunta inicial, dejando claro que su concepción moderada tiene que ver con que la respuesta a si los sistemas de IA pueden ser sujetos del derecho penal no se deriva tanto de su posibilidad (*feasibility*), como de su necesidad (*unnecesity*) –p. 224–. La respuesta a su pregunta de investigación es, pues, un no rotundo. Los

sistemas de IA no pueden ser considerados agentes morales, señalando —p. 230— que solo un agente que tiene la experiencia de autoría (*autorship*), es decir, que puede experimentar una "*constelación de sentimientos*" tales como percibirse a sí mismo como la causa de sus propias acciones y el vínculo entre sus movimientos y sus efectos, puede ser considerado responsable. La agencia intencional en el sentido explicado por Duff resulta esencial e irrenunciable para el derecho penal de la actualidad, por lo que sujetos al derecho penal estarán solo aquellos agentes que sean capaces de "responder" a las demandas o requerimientos del sistema, es decir, cuestionando las normas en un sentido profundo. Redondea este no rotundo aludiendo, una vez más, al fin del castigo penal, confirmando que, más allá de teorías preventivas o comunicativas, la retribución tiene, necesariamente, un papel. Tampoco le convencen a la autora los planteamientos relativos a la agencia colectiva que surgen al realizar paralelismos entre la persona jurídica y la IA, y no le convencen por un motivo ensalzable, a nuestro parecer, pues se fija Giannini en cuál es el propósito esencial de ambas. Si la empresa tiene ínsito en su origen la persecución de un fin económico que le es privado cuando es castigada, la IA no persigue, como "razón fundamental de existir", un propósito económico, así que difícilmente podrá ser dañada o prevenida. Y, en definitiva, si la idea es causar algún tipo de reacción en los humanos que vean y entiendan el castigo, ¿por qué no castigarlos directamente a ellos? La solución pasa, entonces, necesariamente, por castigar al humano. Aunque la aportación que realiza Giannini llegados a este punto es modesta (pues repite hasta la saciedad que no es este el verdadero objeto de su investigación), no es por ello menos valiosa. La idea acertada que destaca es que los humanos, en el papel de supervisores que estamos llamados a desempeñar, corremos el riesgo de desentendernos moralmente de nuestras acciones ("*moral disengagement*"), por lo que nuestra sensación de agencia corre el riesgo de debilitarse, y ello puede tener un impacto innegable en la configuración de los deberes que rodeen al renovado concepto de "imprudencia tecnológica".

Finaliza la obra la autora, por lo tanto, justificando el objeto de su investigación, pues alguien tenía que asumir la incómoda tarea de tomarse en serio planteamientos extravagantes, como siempre lo son las ideas que se formulan por primera vez. Si hace algunos años las personas que defendían la responsabilidad penal de las personas jurídicas podían reunirse subiéndose todas en un taxi[2], quizá hoy ese mismo taxi serviría para llevar a todos los pertenecientes al Frente de Liberación robótica —taxi al que, confieso, hoy no me subiría, razón por la cual no me es difícil compartir las conclusiones que Giannini alcanza en el núcleo duro de su

[2] La ya célebre cita fue pronunciada por el profesor J.M. Zugaldía Espinar en el acto de defensa de la tesis doctoral de la profesora Silvina Bacigalupo, en diciembre de 1997. Zugaldía ironizaba con el, por entonces, escasísimo apoyo que las teorías sobre responsabilidad penal corporativa tenían en España.

obra, si bien con algunos matices—. Y no por ello me parece menos útil una investigación que se tome en serio sus postulados, explorando la realidad del complejo fenómeno que esconden. Desentrañar los problemas comunes a la capacidad de imputación penal, poner a prueba nuestro entendimiento sobre nuestra capacidad de razonamiento moral o reflexionar sobre cómo la sociedad concibe al prudente *"homo tecnologicus"*, si se me permite el artificioso constructo, es una tarea loable. Se echan en falta, por supuesto, algunas reflexiones o referencias. Por nombrar solo algunas, si la autora quería asegurarse de obtener una percepción profunda sobre la dogmática penal, quizá las referencias bibliográficas referentes a los problemas de agencia, imputación, acción y culpabilidad podrían haber estado más inclinadas a dar voz a las grandes voces del derecho penal, sin centrarse tanto en los textos que han abordado los problemas desde el punto de vista tecnológico. Pero esta carencia está justificada si se tiene en cuenta que la autora ha revisado una temática donde la cantidad de información permite difícilmente profundizar en cada tema, por lo inabarcable de la bibliografía actual. Acotar es, en este tópico, necesario.

Se echa en falta, también, una referencia más detallada al impacto que una norma tan relevante como el Reglamento de IA (aunque en el momento no aún aprobado en su versión definitiva, pero sí ampliamente conocido y sustancialmente igual al finalmente aprobado) puede tener en la temática que, de forma tangencial pero presente en toda la obra, aborda. La determinación de los deberes de cuidado y la creación de nuevas normas se está produciendo, *de facto*, no solo con la aprobación del texto legal, sino con todos los trabajos previos que sobre la materia han ido publicando las instituciones y organizaciones internacionales. Que ninguna referencia se haga a los principios éticos universalmente reconocidos y pacíficamente aceptados nos parece cuestionable. Sobre todo, por el papel que los mismos pueden desempeñar a la hora de fijar directrices en las nuevas formas de concebir los deberes del humano de detrás. En todo caso, no era el objeto principal de su investigación hacerse cargo de la responsabilidad del humano, sino dejar apuntados futuros caminos de investigación. Debemos dar gracias, desde luego, los investigadores y, me atrevería a decir, cualquier persona interesada en la temidísima "revolución de los robots", pues la obra de Giannini viene a aportar claridad y coherencia a un tema que debería trascender las fronteras de la ciencia ficción para hacerse un hueco en el de la reflexión calmada y fundada en derecho.