

SOBRE UNA ADECUACIÓN DEL DERECHO PENAL ANTE LOS CONFLICTOS DERIVADOS DE LAS TECNOLOGÍAS DE INTELIGENCIA ARTIFICIAL

ON THE ADAPTATION OF CRIMINAL LAW TO CONFLICTS ARISING FROM ARTIFICIAL INTELLIGENCE TECHNOLOGIES

ROBERTO NAVARRO-DOLMESTCH* **

RESUMEN

Este artículo sostiene la hipótesis que para que las estructuras y conceptos de los que dispone el derecho penal sean adecuados para responder ante los conflictos jurídico-penales en los que, directa o indirectamente, se involucre una máquina dotada de tecnología de inteligencia artificial (TIA), se necesita modificar el esquema binario sobre el que se construye normativamente el sistema penal que distingue entre ‘sujetos’ y ‘objetos’. En ese sentido, se propone incorporar una tercera categoría estructural, adicional a las dos indicadas, que consiste en la de los ‘actuantes’, reservada para los dispositivos dotados de TIA. La identificación de las máquinas inteligentes como ‘actuantes’, se propone, contribuirá a una mejor operativización de las estructuras penales en la resolución de los conflictos.

Palabras clave: Inteligencia artificial, derecho penal, responsabilidad penal, actuantes, delito-IA, causalidad.

*Doctor en Derecho. Profesor en la Facultad de Derecho de la Universidad de Tarapacá, sede Iquique, Chile. Correo electrónico: rnavarro.dolmestch@gmail.com. ORCID: <https://orcid.org/0000-0003-0907-5714>.

**Este artículo se ha elaborado en el marco del proyecto de investigación Fondecyt-Iniciación (2023) titulado “Aplicación del derecho penal a la inteligencia artificial” N° 11230216, financiado por la Agencia Nacional de Investigación y Desarrollo de Chile y del que el autor es su investigador responsable.

Trabajo recibido el 30 de marzo de 2025 y aceptado para su publicación el 23 de junio de 2025.

ABSTRACT

This paper advances the hypothesis that, in order for the structures and concepts available within criminal law to be adequate for addressing criminal conflicts in which a machine endowed with artificial intelligence technologies (AIT) is directly or indirectly involved, it is necessary to modify the binary framework upon which the criminal justice system is normatively constructed—namely, the distinction between ‘subjects’ and ‘objects’. In this regard, the paper proposes the incorporation of a third structural category, in addition to the aforementioned two: that of ‘actuants’, reserved for devices equipped with AIT. It is intended that the identification of intelligent machines as ‘actuants’ will contribute to a better operationalization of criminal law structures for the resolution of conflicts of this nature.

Keywords: Artificial intelligence, criminal law, criminal liability, actants, AI crime, causality.

INTRODUCCIÓN

El surgimiento de las máquinas inteligentes, esto es, las distintas aplicaciones de las tecnologías de inteligencia artificial (en adelante, TIA), le ha dado a un viejo problema un sentido completamente nuevo. El viejo problema es el de los conflictos que siempre han producido los desarrollos tecnológicos y que plantean para el Derecho constantes desafíos de regulación.¹ El nuevo sentido consiste en que las TIA parecen tener unas ciertas características que producen como consecuencia la aparición de dudas sobre la idoneidad de las estructuras jurídicas existentes para responder adecuadamente a los conflictos que las máquinas inteligentes producen. Este artículo se sitúa en el tema de la gestión de los conflictos jurídico-penales que se producen a propósito del involucramiento de las máquinas inteligentes en conductas que revisten los caracteres de delito.

Sobre la base de una premisa optimista, según la cual “el derecho penal ya cuenta con los recursos para hacer frente a la actividad delictiva relacionada con la TIA, al menos a mediano plazo”,² se sostiene que para lograr un adecuado rendimiento de las estructuras del derecho penal ante las máquinas

¹ MANDEL, Gregory N. (2007). History Lessons for a General Theory of Law and Technology. Minnesota Journal of Law, Science & Technology, vol. 8 n° 2, *passim*.

² DSOUZA, Mark, “Don’t panic. Artificial intelligence and Criminal Law”, en: BAKER, D.J.; ROBINSON, P.H. (Eds.), *Artificial intelligence and the Law. Cybercrime and criminal liability*, Routledge, Oxon, 2020, p. 247.

inteligentes, se necesita asignarles a esos dispositivos relevancia estructural y teórica, reconociéndoles a estos dispositivos una nueva categoría intermedia: los ‘actuantes’. En efecto, el modelo conceptual bipartito, sobre el que se ha construido el moderno derecho penal, solo distingue entre ‘sujetos’ y ‘objetos’³ y que solo está en condiciones de atribuir a los dispositivos basados en TIA el mismo estatus que a las tecnologías convencionales (no-inteligentes), incluyéndolas en la categoría de los ‘objetos’, esto es, como meros instrumentos. Sin embargo, esa estructura dual desconoce las características ontológicas de las máquinas inteligentes y que las distinguen de las tecnologías convencionales. Tales características –entre ellas, sus niveles de autonomía y la capacidad técnica que tienen para involucrarse en hechos que revisten los caracteres de delito– hacen que estos dispositivos desborden la tradicional categoría de los objetos, presentando rasgos propios de la de los sujetos. Y es esta estrechez conceptual la que puede explicar que se estime, *prima facie*, que las estructuras o la teoría del derecho penal sean consideradas inadecuadas para la gestión de casos con involucramiento de dispositivos basados en TIA,⁴ cuyo surgimiento demuestra la necesidad de superar el señalado modelo bipartito.

Para el desarrollo de la hipótesis, el artículo se estructura según pasa a indicarse. En el § I se expone, a modo de premisa, el problema del desarrollo de las TIA y de su capacidad actual para generar conflictos jurídico-penales. El § II se destina a describir la estructura dual sobre la que se ha construido el derecho y las características que presentan las máquinas inteligentes que hacen cuestionar la idoneidad de esa estructura. El § III describe la solución que propone este artículo, y que consiste en describir para los dispositivos basados en TIA la categoría de los ‘actuantes’. El § IV, por su parte, se destina a exponer algunos casos en los que la categoría propuesta de los ‘actuantes’ prestaría utilidad para encontrar soluciones en el derecho penal. Este trabajo debe comprenderse dentro de una línea de investigación actualmente en desarrollo y, por consiguiente, como una aproximación al tema abordado.

³ INATANI, Tatsuhiko, ““Moralizing Technology” and Criminal Law Theory”, en: BORGES, G., SORGE, C. (Eds.), *Law and Technology in a Global Digital Society. Autonomous Systems, Big Data, IT Security and Legal Tech*, Springer, Cham, 2022.

⁴ KING, Thomas C., “Projecting AI-Crime: A Review of Plausible Threats”, en: ÖHMAN, C., WATSON, D. (Eds.), *The 2018 Yearbook of the Digital Ethics Lab*, Springer, Cham, 2019, pp. 67-68.

I. UNA PREMISA: LA PERSPECTIVA TEMPORAL DEL DESARROLLO DE LAS TIA

Los análisis del tratamiento jurídico de los conflictos penales en los que se involucre una máquina inteligente están influidos por una dimensión temporal: si la capacidad de esos dispositivos para involucrarse en actividades delictivas es presentada por las máquinas en la actualidad o si es una capacidad que puede llegar a ser técnicamente desarrollada en un futuro más o menos cercano. El punto puede ilustrarse en la elección del modo gramatical correcto para traducir al castellano el título del libro de HALLEVY, “*When robots kill?*”.⁵ Al respecto, hay tres posibilidades. La primera, traducirlo en el modo indicativo del tiempo verbal presente, como “cuando los robots matan”; y la segunda y tercera, al tiempo verbal presente o al futuro, pero ahora del modo subjuntivo, esto es, “cuando los robots maten” o “mataren”, respectivamente. El indicativo es el “modo de la realidad, de la objetividad, de lo seguro o de lo actual”,⁶ mientras que el subjuntivo expresa incertidumbre o futuridad indefinida.⁷ Gramaticalmente, el modo informa sobre la “actitud del hablante ante la información suministrada y, en particular, sobre el punto de vista que este sostiene en relación con el contenido de lo que se presenta o se describe”.⁸ No hay consenso sobre el estado de desarrollo actual de las TIA y de ello se deriva una dificultad en la validación del modo gramatical adecuado o correcto que debería utilizarse. Tal ausencia de acuerdo puede explicarse por varios motivos. Por un lado, porque no siempre hay información disponible sobre esos desarrollos, por razones de seguridad (como es el caso del ámbito militar y defensa) o por la necesidad de mantener el secreto industrial. Y, por otro, porque las TIA se han desarrollado no solo sobre distintos enfoques técnicos, sino también en diversas áreas, generando aplicaciones con variadas funciones y ámbitos de uso (dominios). Al panorama anterior debe agregarse que la acelerada velocidad de los avances que han registrado las TIA producen un problema adicional: la dificultad del derecho penal para “seguir el ritmo evolutivo de la tecnología informática que, cada vez a mayor velocidad, está materializando lo que hace pocos años parecía solo fruto de la imaginación”.⁹

⁵ HALLEVY, Gabriel, *When Robots kill: Artificial intelligence under criminal law*. Northeastern University Press, Lebanon NH, 2013.

⁶ REAL ACADEMIA ESPAÑOLA, *Gramática descriptiva de la lengua española* (BOSQUE, I. y DEMONTE, V., Eds.), Espasa, Madrid, 1999, p. 3218.

⁷ REAL ACADEMIA ESPAÑOLA, cit. (n. 6).

⁸ REAL ACADEMIA ESPAÑOLA, *Nueva gramática de la lengua española, Morfología y sintaxis I*, Espasa, Madrid, 2009, p. 1866 (§25.1c).

⁹ DE LA CUESTA, Paz, “Inteligencia artificial y responsabilidad penal”, *Revista Penal México*. 2020,

Sin perjuicio de esas dificultades, basado en la evidencia experimental¹⁰ y de

Nº 16-17, p. 60.

¹⁰ Aunque no es posible hacer un análisis detallado en este lugar, se pueden mencionar algunos casos, como el de la posibilidad de que las TIA aprendan que la manipulación del mercado es una estrategia de inversión óptima (en este sentido, MARTINEZ-MIRANDA, E., MCBURNEY, P., HOWARD, M., “Learning unfair trading: A market manipulation analysis from the reinforcement learning perspective”, en AA.VV., *2016 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, Natal, Brazil, 2016; CARROLL, Micah; CHAN, Alan; ASHTON, Henry; KRUEGER, David, “Characterizing Manipulation from AI Systems”, en: AA.VV., *EAAMO '23: Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, ACM, New York, 2023); “aunque el desarrollador no haya tenido intención de manipular el mercado” (MIZUTA, Takanobu, “Can an AI perform market manipulation at its own discretion? – A genetic algorithm learns in an artificial market simulation”, en: AA.VV., *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, ACT, Canberra, 2020, p. 411). O los que informan KING y sus colegas, que son experimentos que demostraron que aplicaciones basadas en TIA pueden ‘cometer delitos’ por medio de la obtención de valiosa información personal de usuarios (KING, Thomas C.; AGGARWAL, Nikita; TADDEO, Mariarosaria; FLORIDI, Luciano; “Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions, *Science and Engineering Ethics*, 2020, vol. 26, nº 1; KING, Thomas C.; AGGARWAL, Nikita; TADDEO, Mariarosaria; FLORIDI, Luciano; “Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions”, en: COWLS, J., y MORLEY, J. (Eds.), *The 2020 Yearbook of the Digital Ethics Lab*, Springer, Cham, 2021). O el de una aplicación de IA usada en el diseño de moléculas para el desarrollo de nuevos fármacos, con la que se hizo la prueba de modificar sus incentivos originalmente construidos para producir moléculas benignas. Bajo los nuevos incentivos, en menos de seis horas la aplicación produjo alrededor de 40 mil moléculas tóxicas, algunas ya conocidas y otras nuevas, muchas de ellas plausibles de realizar, potencialmente más tóxicas que las ya conocidas y que entrañan un riesgo concreto de su uso ilícito para la producción de armas químicas (URBINA, Fabio; LENTZOS, Filippa; INVERNIZZI, Cédric; y EKINS, Sean, “Dual use of artificial-intelligence-powered drug Discovery”, *Nature Machine Intelligence*, 2022, N° 4).

la realidad,¹¹ comparto la posición de una parte de la doctrina¹² que sostiene que las

¹¹ En la realidad, por su parte, al ejemplo más habitual y paradigmático de los sucesos dañinos causados por vehículos autónomos (al respecto, ver, entre otras referencias: FAHIM, Sadaf, *Ethico-Legal Aspect of AI-driven Driverless Cars*, Springer, Singapore, 2024; KIRPICHNIKOV, Danila; PAVLYUK, Albert; GREBNEVA, Yulia; OKAGBUE, Hilary, “Criminal Liability of the Artificial Intelligence”, *EDP Sciences*, 2020, vol. 159, p. 5; SIBAI, Fadi N., “AI Crimes: A Classification”, en: IEEE (Eds.), *2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, Dublin, Ireland, 2020, p. 8; LA ROSA, Emanuele, “Criminal Law and Artificial Intelligence: Which Spaces for a Criminal Liability of the Robot?”, en: MARINO, D.; MONACA, M. (Eds.), *Artificial Intelligence and Economics: the Key to the Future*, Springer, Cham, 2023, p. 175), pueden agregarse los riesgos que las TIA representan para la comisión de delitos informáticos, como tempranamente lo identificaron BRUNDAGE y sus colegas (BRUNDAGE, Miles, et. al., “The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation”, *ArXiv*, 2018, Vol. abs/1802.0, pp. 19 y ss.) y reiterados por DI NICOLA, Andrea, “Towards digital organized crime and digital sociology of organized crime”, *Trends in Organized Crime*, 2022, entre otros. Asimismo, piénsese en las aplicaciones de IA –disponibles incluso en los smartphones– que permiten manipular contenido audiovisual previamente existente o generar contenido audiovisual (sonidos, imágenes fijas o videos) “sintético” (KAUSHIK, Purva, “Privacy and Other Legal Concerns in the Wake of Deepfake Technology: Comparative Study of India, US and China”, en; DEWANI, N.D.; KHAN, Z.A.; AGARWAL, A.; KHAN, S.A. (Eds.), *Handbook of Research on Cyber Law, Data Protection, and Privacy*, IGI Global, Hershey, 2022; MAHER, Sean, “Deep Fakes. Seeing and Not Believing”, en: FILIMOWICZ, M. (Ed.), *Deep Fakes. Algorithms and Society*, Routledge, London - New York, 2022), esto es, completa o parcialmente fabricado. Son las denominadas *deepfakes* (VAN DER SLOOT, Bart y WAGENVELD, Yvette, “Deepfakes: regulatory challenges for the synthetic society”, *Computer Law & Security Review*, 2022, vol. 46) y sus versiones más perversas, el *deep porn* o el *revenge porn* (DRINNON, Caroline, “When Fame Takes Away the Right to Privacy in One’s Body: Revenge Porn and Tort Remedies for Public Figures”, *William & Mary Journal of Women and the Law*, 2017, Nº 24; DELFINO, Rebecca A., “Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn’s Next Tragic Act”, *Fordham Law Review*, 2019, vol. 88, Nº 3) que se han usado, entre otros ámbitos, en la generación o manipulación de contenido de carácter sexual no consentido. El autor SIBAI (cit. en esta nota) menciona como ejemplos de AI crimes las acciones de espionaje a usuarios realizadas por Amazon y Apple a través de sus asistentes por voz para el entrenamiento de sus aplicaciones. O el caso de la compra en la *darknet* de éxtasis (MDMA), en octubre de 2014, por un *bot* en el contexto de una instalación artística (<https://2016.bitnik.org/r/>). Al respecto, ver ABBOTT, Ryan, *The reasonable robot. Artificial Intelligence and the Law*, Cambridge University Press, Cambridge, 2020, pp. 111-112; CHESTERMAN, Simon, *We, the Robots?*, Cambridge University Press, Cambridge, 2021, p. 57; DSOUZA, cit. (n. 2), p. 253; FOSCH-VILLARONGA, Eduard, *Robots, Healthcare, and the Law. Regulating Automation in Personal Care*, Routledge, Oxon, 2020, p.164; HAYWARD, Keith J.; MAAS, Matthijs, “Artificial intelligence and crime: A primer for criminologists”, *Crime, Media, Culture*, 2020, Vol. 17, Nº 2, p. 9; LAGIOIA, Francesca; SARTOR, Giovanni, “AI Systems Under Criminal Law: A Legal Analysis and a Regulatory Perspective”, *Philosophy and Technology*, 2020, Vol. 33, Nº 3, p. 452); y el de *Alexa*, el asistente conversacional de Amazon, que en diciembre de 2021 le sugirió a una niña de diez años un desafío potencialmente peligroso que el asistente ‘conoció’ en Internet y que consistió en que la chica separara parcialmente el cargador de un teléfono móvil desde el tomacorriente de la pared e insertara una moneda metálica (<https://www.bbc.com/news/technology-59810383>).

¹² DEĞIRMENCI, Olgun. “Sufficiency of Struggling with the Current Criminal Law Rules on the Use of Artificial Intelligence in Crime”, en: KILIÇ, M.; KAHYAOĞLU, S.B. (Eds.), *Algorithmic Discrimination and Ethical Perspective of Artificial Intelligence*, Springer, Gateway East, 2024, pp. 99-100; ABBOTT, Ryan; SARCH, Alex, “Punishing Artificial Intelligence: Legal Fiction or Science Fiction”, *UC Davis Law*

máquinas inteligentes existentes hoy, con su actual grado de desarrollo tecnológico, sí pueden involucrarse materialmente en la ejecución de conductas que revisten los caracteres de delito. Para que las TIA representen desafíos jurídicos no es necesario esperar al desarrollo del “Robot Kleptomaniac”,¹³ una tecnología con plena autonomía capaz de resolver cualquier problema (como una IA general o fuerte, es decir, sistemas “que cobran vida propia o adquieren autoconciencia”)¹⁴ y, con ello, de sustituir –o, al menos, igualar– al ser humano en sus capacidades. Han sido los desarrollos de TIA débiles, esto es, enfocadas en tareas y fines específicos,¹⁵ y sus capacidades ya mostradas, las que, por sí mismas, han hecho surgir preguntas sobre el eventual tratamiento jurídico-penal de los conflictos a que pueden dar lugar.¹⁶ En consecuencia, la traducción al modo indicativo sí se corresponde con la realidad y este es el enfoque que adopto en este trabajo.

II. IDENTIFICAR LA NATURALEZA DE LAS MÁQUINAS INTELIGENTES

La afirmación de que las máquinas inteligentes pueden involucrarse materialmente en la ejecución de conductas constitutivas de delito trae consigo la

Review, 2019, vol. 53, N° 1, pp. 179-180; ABBOTT, Ryan; SARCH, Alex, “Punishing Artificial Intelligence: Legal Fiction or Science Fiction” en: DEAKIN, S.; MARKOU, Ch. (Eds.), *Is Law Computable? Critical perspectives on Law and Artificial Intelligence*, Hart Publishing, Oxford, 2020, pp. 197-199; ABBOTT, cit. (n. 11), p. 13; LA ROSA, cit. (n. 11), p. 175; DEL ROSAL, Bernardo, “¿El modelo de la responsabilidad penal de las personas jurídicas para los daños punibles derivados del uso de la Inteligencia Artificial?”, *Revista de Responsabilidad Penal de Personas Jurídicas y Compliance*, 2023, vol. 2, pp. 6-7.

¹³ REYNOLDS, Carson y ISHIKAWA, Masatoshi, “Robotic Thugs”, en: *Global e-SCM Research Center y Meiji University*, 2007, p. 488; PAGALLO, Ugo, “Robots of Just War: A Legal Perspective”, *Philosophy & Technology*, 2011, vol. 24, N° 3, p. 311; FREITAS, Pedro Miguel; ANDRADE, Francisco; NOVAIS, Paulo, “Criminal liability of autonomous agents: From the unthinkable to the plausible”, en: CASANOVAS, P.; PAGALLO, U.; PALMIRANI, M.; SARTOR, G. (Eds), *AI Approaches to the Complexity of Legal Systems. AICOL 2013. Lecture Notes in Computer Science*, vol 8929, Springer, Berlin - Heidelberg, 2014, p. 154; PAGALLO, Ugo, “Killers, fridges, and slaves: A legal journey in robotics”, *AI and Society*, 2011, vol. 26, N° 4, p. 348; PAGALLO, Ugo, *The Laws of Robots. Crimes, Contracts, and Torts*, 2013, Springer, Dordrecht, p. 52.

¹⁴ HALE, Benjamin, “Technology, the Environment and the Moral Considerability of Artefacts”, en: BERG OLSEN, J.; SELINGER, E.; RIIS, S. (Eds.), *New Waves in Philosophy of Technology*, Palgrave Macmillan, London, 2009, p. 235.

¹⁵ La distinción entre una inteligencia artificial “fuerte” (“strong”) y una “débil” o “cautelosa” (“weak” o “cautious”) fue trazada por SEARLE, John, “Minds, brains, and programs”, *Behavioral and Brain Sciences*, 1980, N° 3.

¹⁶ KÖNIG, Pascal D.; KRAFT, Tobias D.; SCHULZ, Wolfgang; y ZWEIG, Katharina A., “Essence of AI. What Is AI?”, en: DIMATTEO, L.; PONCIBÒ, C.; CANNARSA, M. (Eds.), *The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press, Cambridge, 2022, p. 25.

necesidad de conceptualizar tales dispositivos. El concepto que se construya debe tener su foco en lo jurídico que permite identificar instrumentos de regulación de las TIA. Por ello, es suficiente encontrar una ‘definición de trabajo’ de las máquinas inteligentes, esto es, establecer “una aclaración sobre ‘qué quiero/queremos decir con IA’”.¹⁷ En ese contexto, el punto de partida se encuentra en que las TIA son concretas aplicaciones de sistemas dotados de algoritmos de inteligencia artificial (máquinas inteligentes),¹⁸ esto es, programas informáticos que, ejecutados en aparatos corporeizados (como un computador o un robot) o simplemente como software, cuentan con capacidades (sensores y percepción, razonamiento y toma de decisiones, y actuación)¹⁹ que pueden desplegar de acuerdo con unas determinadas propiedades (autonomía, interactividad y adaptabilidad).²⁰ Esto les permite realizar funciones, resolver problemas, generar contenido o efectuar predicciones.²¹ Ello se logra emulando capacidades de la inteligencia biológica o natural de razonamiento deductivo²² o de aprendizaje.²³ Con valor descriptivo, pero no exento de exageración, se ha llegado a sostener que el desarrollo de estas capacidades de las TIA lo ha sido hasta “un punto en el que claramente podemos decir que el pensar ya no es un privilegio exclusivo de mentes vivientes”.²⁴

¹⁷ WANG, Pei, “On Defining Artificial Intelligence”, *Journal of Artificial General Intelligence*, 2019, vol. 10, N° 2, p. 3.

¹⁸ Lo que implica descartar la atención en la inteligencia artificial como un campo o disciplina científica o una tecnología o método. Campo científico, tecnología y aplicaciones son tres aspectos a los que el uso del concepto de inteligencia artificial puede referirse. KÖNIG, KRAFFT, SCHULZ y ZWEIG, cit. (n. 16), p. 18.

¹⁹ EC HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE (Eds.), “A definition of IA: Main capabilities and scientific disciplines”, European Comission, Bruselas, 2018, disponible en línea: https://ec.europa.eu/futurum/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf; SIBAI, cit. (n. 11).

²⁰ FLORIDI, Luciano; SANDERS, J.W., “On the morality of artificial agents”, *Minds and Machines*, 2004, vol. 14, N° 3, pp. 357-358.

²¹ FINCAN, Müslüm, *Artificial Intelligence and Legal Issues. A Review of AI-based Legal Impasses in Terms of Criminal Law*, Duncker & Humblot, Berlin, 2022, p. 33.

²² En un inicial enfoque adoptado por las ciencias informáticas, conocido como ‘good old-fashioned AI (GOFAI), las TIA pretendían replicar métodos deductivos propios de la lógica. PERIN, Andrea, “La ‘inteligencia artificial’ en la justicia penal ante el principio de responsabilidad penal”, en: ROMEO, C.; RUEDA, M. (Eds.), *Derecho penal, ciberseguridad, cibercrimen e inteligencia artificial. Volumen II: inteligencia artificial y responsabilidad penal*, Comares, Granada, 2023, p. 84.

²³ Como es el caso de los enfoques basados en aprendizaje automático (*machine learning*) que, recurriendo a técnicas estadísticas y probabilísticas (SABOURET, Nicolas, “What Artificial Intelligence Can Do and What It Cannot Do”, en: BOUNFOUR, Ahmed (Ed.), *Platforms and Artificial Intelligence. The Next Generation of Competences*, Springer, Cham 2023, p. 10), les permite a los algoritmos mejorar su rendimiento en alguna tarea a través de la experiencia (GEETHA, T.V.; SENDHILKUMAR, S., *Machine Learning. Concepts, Techniques and Applications*, Chapman-Hall-CRC, Boca Raton, 2023, p. 10).

²⁴ FARHADI, Ashkan, “There is no “I” in “AI””, *AI & Society*, 2021, vol. 36, N° 3, p. 1035.

Las máquinas inteligentes son, antes que todo, dispositivos tecnológicos. En principio, no representarían problema alguno si fueran ontológicamente iguales que las tecnologías convencionales para las que el derecho penal está preparado, toda vez que ha sido construido sobre la idea de que, por sus propiedades y características, las máquinas pueden desenvolver como herramientas, es decir, como objetos, con independencia de su nivel de sofisticación. Esta última afirmación requiere de un breve desarrollo que se expone en el apartado siguiente.

2.1. *Objetos, sujetos y máquinas*

El moderno derecho penal tiene una evidente inspiración humanista sobre la que se ha construido que se expresa, entre otros aspectos, en una clara escisión entre los sujetos como los destinatarios de sus prescripciones, y los objetos, esto es, a los elementos a los que los sujetos recurren para ejecutar las conductas.²⁵ La categoría de los sujetos queda reservada para las personas que exhiben aquellas condiciones que se consideran esenciales para fundamentar la responsabilidad penal, dentro de las que puede reconocerse la decisión de ejecutar el delito; y la de los objetos, por su parte, se integra por todos aquellos elementos utilizados por los sujetos. La noción de instrumento del delito que recoge la legislación chilena seguramente expone con claridad la esencia de esta categoría. Al decir de COUSIÑO MAC IVER, los instrumentos del delito son “las armas, objetos, herramientas u otros objetos semejantes, que se destinan a la ejecución del delito”²⁶.

Los dispositivos tecnológicos convencionales (o no-inteligentes) integran la categoría estructurante de los ‘objetos’. De acuerdo con su naturaleza, el derecho penal considera a las tecnologías convencionales como meros instrumentos, en la medida que, para su funcionamiento, estas requieren de la interacción directa de una persona, esto es, de un agente. Un buen ejemplo es el arma homicida: mientras esta no sea percutida por un agente humano, solo representa un peligro potencial. En la comisión del homicidio, el arma será un instrumento a merced del agente. En consecuencia, el resultado que materialmente cause el arma será imputable tanto objetiva como subjetivamente a la persona que la utilizó, para quien se trata del uso de un mero instrumento para cometer el delito.²⁷

²⁵ INATANI, cit. (n. 3), p. 29.

²⁶ COUSIÑO MAC IVER, Luis, *Derecho Penal chileno*, Tomo I, Editorial Jurídica de Chile, Santiago de Chile, 1975, p. 298.

²⁷ ABBOTT y SARCH, cit. (n. 12), pp. 197-199.

2.2. ¿Máquinas inteligentes como objetos?

No hay obstáculo alguno para que las TIA sean utilizadas por un agente como un mero instrumento para la comisión de un delito,²⁸ caso en el que la máquina inteligente debe ser considerada un objeto; y su utilización hace responsable al sujeto que decidió su uso.²⁹ Pero el panorama del uso y aportación de las máquinas se complejizó a partir del surgimiento de las TIA. En atención a su naturaleza, sostengo que las TIA desbordan la estructura dualista, es decir, se salen de los contornos de la categoría de los ‘objetos’.

Por un lado, porque el esquema de utilización de las TIA es diferente al de las tecnologías convencionales y, dentro de estas, las informáticas. Mientras estas últimas requieren siempre de la interacción con un agente humano, las TIA pueden prescindir de dicho sujeto e, incluso, establecer relaciones de uso entre dispositivos inteligentes. En virtud de esta característica, FLORIDI³⁰ califica a las TIA como tecnologías de tercer orden. Por otro, y como lo explica VERBEEK,³¹ basado en los trabajos de IHDE, que los dispositivos digitales (aplicable, por antonomasia, a los inteligentes) “al mediar nuestra relación sensorial con la realidad, transforman lo que percibimos”, razón por la que “las tecnologías tienen “intenciones” y, por lo tanto, no son instrumentos neutrales, sino que juegan un papel activo en la relación entre los humanos y su mundo”. Estos dispositivos “interactúan con las intencionalidades de los seres humanos que utilizan estos artefactos,³² haciendo surgir una “intencionalidad compuesta: una forma de intencionalidad que resulta de sumar la intencionalidad tecnológica a la intencionalidad humana”.³³

Lo que hace que la categoría de los objetos quede estrecha para las TIA es

²⁸ ABBOTT y SARCH, cit. (n. 12), pp. 197-199.

²⁹ ROMEO, Carlos, “La atribución de responsabilidad penal por los hechos cometidos por sistemas autónomos inteligentes, robótica y tecnologías conexas”, *ULP Law Review*, 2022, vol. 16, N° 1, pp. 10-11.; ROMEO, Carlos, “La atribución de responsabilidad penal por los hechos cometidos por sistemas autónomos inteligentes, robótica y tecnologías conexas”, en: MUÑOZ, J., GARCÍA, O., CEREZO, A. y GARCÍA, E. (Eds.), *Estudios jurídico-penales, criminológicos y político criminales. Libro homenaje al Profesor José Luis Díez Ripollés*, Tirant lo Blanch, Valencia, 2023, pp. 701-703.; ROMEO, Carlos, “La discusión sobre la atribución de responsabilidad penal a los sistemas de inteligencia artificial, en particular a sistemas autónomos”, en: ROMEO, C.; RUEDA, M. (Eds.), *Derecho penal, ciberseguridad, ciberdelitos e inteligencia artificial. Volumen II: inteligencia artificial y responsabilidad penal*. Comares, Granada, 2023, pp. 67-68.

³⁰ FLORIDI, Luciano. *The fourth revolution. How the infosphere is reshaping human reality*. Oxford University Press, Oxford, 2014, pp. 29-30.

³¹ VERBEEK, Peter-Paul, *Moralizing Technology. Understanding and Designing the Morality of Things*, University of Chicago Press, Chicago - London, 2011, p. 9.

³² VERBEEK, cit. (n. 31), p. 145.

³³ VERBEEK, cit. (n. 31), p. 145.

que estas presentan algunos atributos asimilables a los que tienen los ‘sujetos’, integrantes de la otra categoría estructurante. En efecto, las máquinas inteligentes tienen, por un lado, capacidad de actuación en el mundo real, ya que las acciones que estas emprendan o los resultados que de estas se esperan, pueden influir en el contexto en el que tales dispositivos se desenvuelven. Por otro, las máquinas inteligentes tienen capacidad deliberativa que, aunque limitada y en ningún caso comparable en profundidad y complejidad con la de los seres humanos, les permite decidir cursos de acción autónomamente. En este sentido, la autonomía es aquella habilidad técnica de máquinas inteligentes por la que pueden ejecutar una acción sin el control directo de otro usuario humano o de otro dispositivo,³⁴ es decir, es la “capacidad de componer y seleccionar de forma independiente diferentes cursos de acción para lograr objetivos en función de su conocimiento y comprensión del mundo, de sí mismo y de la situación”³⁵.

El involucramiento (material-causal) de las TIA en la ejecución de conductas que revisten caracteres de delito puede presentarse, al menos, en dos niveles distintos.

Sobre el primero ya se ha hecho referencia y consiste en los casos en los que el dispositivo tecnológico es usado como un mero instrumento para la comisión de un delito, como en el ejemplo paradigmático del arma usada para cometer un homicidio. Este primer nivel no ofrece problemas, ya que el derecho penal cuenta con una categoría estructurante –la de los ‘objetos’ (como las cosas, los animales o los dispositivos tecnológicos)– en cuyo diseño se comprenden elementos a los que se les otorga una relevancia mediada por la intervención de un agente, esto es, un sujeto capaz de ser penalmente responsable.³⁶ Esto significa que las estructuras de imputación con las que actualmente cuenta el derecho penal están preparadas para gestionar aquel tipo de conflictos en los que un agente, por sí mismo o usando un mero instrumento (mediado por un sujeto), ejecuta la conducta típica. Como lo afirma DSOUZA, este nivel requiere la utilización deliberada de una TIA como herramienta o mero instrumento y, advierte, aun cuando esta “conserva cierta

³⁴ SCHARRE, Paul; HOROWITZ, Michael C., “An Introduction to Autonomy in Weapon Systems”, 2015, disponible en línea: https://www.files.ethz.ch/isn/188865/Ethical%20Autonomy%20Working%20Paper_021015_v02.pdf; SULLINS, John P., “When is a robot a moral agent?, *International Review of Information Ethics*, 2006, vol. 6, N° 12, p. 28; SULLINS, John P., “When is a robot a moral agent?”, en: ANDERSON, Michael; ANDERSON, Susan (Eds.), *Machine ethics*, Cambridge University Press, Cambridge, 2011, p. 158.

³⁵ DEFENSE SCIENCE BOARD, “DSB Summer Study on Autonomy”, 2016, en línea: <https://dsb.cto.mil/wp-content/uploads/reports/2010s/DSBSS15.pdf>, p. 4.

³⁶ BRUNDAGE, *et al.*, cit. (n. 11); HAYWARD y MAAS, cit. (n. 11), pp. 209-233; SIBAI, cit. (n. 11); BARRESI, Olímpia, “Italian Report on Traditional Criminal Law Categories and AI”, *Revue Internationale de Droit Pénal*, 2023, Vol. 94, N° 1, p. 274.; LA ROSA, cit. (n. 11), pp. 176-177.

autonomía sobre si, cuándo y cómo realiza la conducta específica. Una herramienta impredecible, o no del todo predecible, sigue siendo una herramienta”.³⁷

Pero existe un segundo nivel de involucramiento de las TIA que es posible gracias a sus capacidades y propiedades que marcan las diferencias ontológicas que estas tecnologías presentan con relación a las convencionales. En este, el involucramiento de la máquina inteligente es directo, es decir, no requiere la mediación de un agente. Usando una metáfora, se trata de casos en los que el dispositivo puede decidir cuándo y a quién matar.

El análisis de este segundo nivel requiere algunas distinciones. Una de ellas es la relativa a lo que la doctrina ha denominado ‘delitos-IA’ (*AI crime*),³⁸ esto es, “casos en los que una IA sería penalmente responsable si una persona natural hubiera realizado un acto similar”.³⁹ Ejemplos de laboratorio y reales de delitos-IA han sido mencionados en el § I de este artículo. En general, los delitos-IA plantean desafíos para las tradicionales estructuras de imputación con las que funciona el derecho penal, cuya magnitud es directamente proporcional a la intensidad con la que las características de autonomía, interactividad y adaptabilidad se presenten en el dispositivo inteligente de que se trate. Desde una perspectiva pesimista, pero no exenta de realidad, QUINTERO ha sostenido que para una realidad en la que la intervención de máquinas inteligentes es posible, “el derecho tradicional carece de respuestas coherentes, pues ese derecho, por su propia naturaleza, es inimaginable sin el factor humano”.⁴⁰ En niveles bajos de autonomía, interactividad o adaptabilidad de la máquina inteligente, es posible descubrir una relación material de la acción de la máquina con la de una persona responsable. Esta relación de traduce en que existe causalidad entre la acción o la omisión de la persona y que esa conducta de la aplicación basada en TIA es objetivamente imputable a la persona que ha sido su desarrollador, su entrenador, su distribuidor o su usuario.

En niveles altos de autonomía, interactividad o adaptabilidad de la máquina inteligente, la situación es diferente. Un nivel alto será aquel en el que las características de la máquina inteligente muestren que su actuación no se vincula materialmente con la de un ser humano.⁴¹ Es decir, casos, en los que no exista una relación de causalidad, ni aun hipotética y, por consiguiente, la conducta sea materialmente atribuible solo a la máquina y no a una persona. Estos niveles altos

³⁷ DSOUZA, cit. (n. 2), p. 251.

³⁸ KING, cit. (n. 4); KING, AGGARWAL, TADDEO y FLORIDI, cit. (n. 10), 2020 y 2021; ABBOTT y SARCH, cit. (n. 12) 2019 y 2020.

³⁹ ABBOTT y SARCH, cit. (n. 12, 2019), p. 180.; ABBOTT y SARCH, cit. (n. 12, 2020), p. 332.

⁴⁰ QUINTERO, Gonzalo, “La robótica ante el Derecho penal: el vacío de respuesta jurídica a las desviaciones incontroladas”, *Revista Electrónica de Estudios Penales y de la Seguridad*, 2017, N° 1, p. 7.

⁴¹ Algunos ejemplos se han proporcionado previamente (ver notas n. 10 y 11).

dan lugar a una segunda distinción y que se refiere a lo que la doctrina ha denominado ‘delitos-IA difíciles’ (*hard AI crime*⁴²), esto es, casos en los que conductas descritas en la ley como delito son materialmente ejecutadas por algoritmos inteligentes, pero no hay un sujeto a quien imputarle objetiva y subjetivamente la conducta⁴³ y, por ende, atribuirle responsabilidad. Este sería el caso, por ejemplo, de un algoritmo inteligente que actúa en el mercado de valores y ‘aprende’ que haciendo órdenes masivas de compra, de las que luego se desiste, es una acción que puede hacer que el precio de los activos disminuya y luego comprarlos más baratos. La configuración de un delito-IA difícil no solo debe analizarse con relación a las estructuras dogmáticas, sino también con las posibilidades prácticas de su detección y juzgamiento. Me refiero a que la actividad probatoria y sus reglas pueden también influir en la configuración de un delito-IA difícil. Cuestiones como los déficits de explicabilidad o interpretabilidad de la TIA, su propia complejidad técnica, las capacidades investigativas de los órganos de persecución penal y la comprensión de los operadores del sistema de justicia criminal (jueces, fiscales y abogados) pueden conjugarse para que, en un caso determinado, no sea posible reconocer o acreditar la vinculación de una persona con un comportamiento ejecutado por la TIA. Y aun una tercera distinción es posible: la de los que podrían denominarse ‘proto delitos-IA’ constituidos por una “nueva generación de *actus reus* de IA”,⁴⁴ es decir, conductas inéditas ideadas por máquinas inteligentes y que, por ello, son atípicas e imposibles de indicar porque son conductas que inventarán las máquinas.

Las tres distinciones formuladas en los párrafos anteriores son relevantes para enfrentar la gestión de los conflictos que en estas se originan. El de los ‘proto delitos-IA’ es un caso especial porque el conflicto que ellos pueden originar no es propiamente penal, sino que llama a la acción del legislador para definir si esos comportamientos inéditos deberían o no ser prohibidos penalmente. Esas conductas no generan responsabilidad no porque sean ejecutadas por una máquina inteligente, sino porque son atípicas, de modo que, aunque fueran ejecutadas directamente por un agente capaz de ser responsabilizado, tampoco serían punibles. Pero el grupo de los delitos-IA y el de los delitos-IA difíciles requiere de un análisis.

⁴² ABBOTT y SARCH, cit. (n. 12); KING, cit. (n. 4); KING, AGGARWAL, TADDEO y FLORIDI, cit. (n. 10), 2020 y 2021.

⁴³ Sobre las dificultades de imputación, me remito al exhaustivo trabajo de BARRESI, cit. (n. 36).

⁴⁴ PAGALLO, Ugo; QUATTROCOLO, Serena, “The impact of AI on criminal law, and its twofold procedures”, en: BARFIELD, Woodrow (Ed.), *Research Handbook on the Law of Artificial Intelligence*, Edward Elgar Publishing, Cheltenham, 2018, p. 400.

III. ACTUANTES EN EL DERECHO PENAL

Una posible estrategia a asumir frente a estos casos es la de la atribución directa de responsabilidad penal a las máquinas inteligentes lo que requiere incluir a estos dispositivos en la categoría de ‘sujetos’. No comarto esta estrategia “expansionista”,⁴⁵ y mi rechazo se fundamenta, principalmente, en que ella supone lesionar la dignidad humana. No es posible abordar aquí este importante tema, que constituye por sí mismo una hipótesis para investigaciones posteriores: un derecho penal democrático solo puede comprenderse a través de la idea de la dignidad humana; y es esta la que legitima su forma y contenido. Justifica sus instituciones fundamentales, como la trascendencia que se le otorga a la libertad del sujeto y, junto con ella, a la culpabilidad. Por consiguiente, aplicar el derecho penal directamente a las máquinas implica degradar la esencia de la dignidad, ya que un dispositivo ideado para las personas debería, por ese hecho, quedar reservado para las personas. Basado en el rechazo de la estrategia de la aplicación directa, me centraré en exponer una estrategia alternativa centrada en las máquinas inteligentes y su posición dentro de las estructuras y teoría del derecho penal.

La propuesta puede sintetizarse de la forma que sigue. Los problemas que pueden surgir por el involucramiento de las máquinas inteligentes en conductas que revisten los caracteres de delito se produce porque, como lo sintetiza LA ROSA, en estos casos “las decisiones y evaluaciones que subyacen a la comisión del acto ya no son obra exclusiva del hombre, sino que al menos son compartidas con la máquina, si no delegadas en ella”.⁴⁶ Las decisiones y evaluaciones son posibles gracias a facultades y capacidades que, en el modelito bipartito del derecho penal (sujetos-objetos), se esperan solo de los integrantes de la categoría de los sujetos; o, en sentido contrario, su presencia nunca puede esperarse en los objetos. Y he ahí el origen de los problemas: el derecho penal trata a las máquinas como objetos, pero las inteligentes escapan de ese molde porque tienen algunas cualidades propias de los sujetos. En consecuencia, una forma de hallar una solución pasa por revisar el modelo bipartito, en el sentido de que a partir del surgimiento de las TIA, este necesita ser tripartito. Una nueva categoría intermedia, de los ‘actuantes’, es necesaria para mantener la coherencia de las estructuras y teoría del derecho penal con la realidad sobre la que pretende aplicarse. Atendidos los aspectos tecnológicos de la realidad, en la que la presencia de máquinas inteligentes no deja de crecer y de extenderse, un modelo tripartito es el que ofrece las mejores posibilidades descriptivas y de aplicación. Luego, la distinción de esta nueva categoría intermedia no solo tiene efectos descriptivos; también sirve para identificar la forma en la que

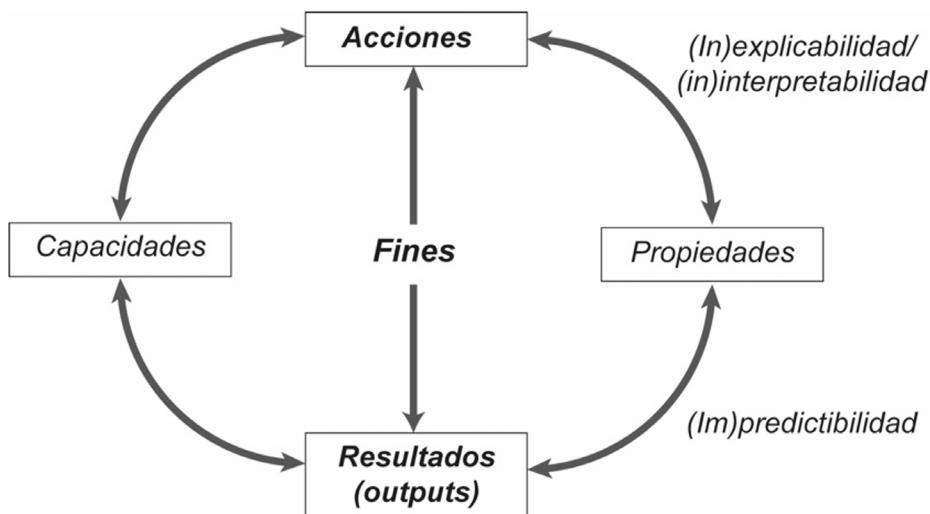
⁴⁵ ABBOTT y SARCH, cit. (n. 12), p. 204.

⁴⁶ LA ROSA, cit. (n. 11), p. 176.

las estructuras del derecho penal deben ajustarse a ella.

Sobre este último aspecto se volverá en el §IV. Por el momento, es necesario delimitar la categoría de los ‘actuantes’. Esa delimitación requiere definir un conjunto de elementos definicionales que permitan darle forma a la categoría y decidir la verdadera pertenencia de una concreta aplicación de TIA, esto es, si integra la categoría de los actuantes o la de los objetos.

Considerando la finalidad que tiene distinguir a los actuantes como una tercera categoría estructurante del derecho penal, sus elementos definicionales (graficados en la Ilustración 1) deben construirse con relación a las características y propiedades de una máquina inteligente ideal en función de su actual desarrollo. En ese sentido, se puede sostener que el conjunto de elementos definicionales está integrado por tres clases de elementos que son las las ‘esferas de actuación’, los ‘componentes’ y las ‘relaciones’.



Fuente: Elaboración Propia (2025).

Ilustración 1. Diagrama de los elementos definicionales de los actuantes.

Las esferas de actuación se refieren, como su nombre lo indica, a las funciones que las máquinas inteligentes pueden desarrollar y comprenden sus fines, las acciones y los resultados (outputs). Estas suponen la definición de una finalidad que es determinada por el diseñador o programador del dispositivo. No es

lógico, ni parece posible aún, pensar en una máquina para la que no se defina una finalidad. Diseñar un dispositivo para que haga lo quiera no tiene sentido; ni parece técnicamente posible aún diseñar una máquina que pueda ejecutar cualquier tipo de función, ya que ello es lo que se espera de una súper inteligencia. Luego, si para el dispositivo basado en TIA se define una finalidad ilícita, este integraría la categoría de los objetos, ya que su utilización para fines ilícitos la convertiría en un mero instrumento. Sería lo mismo que entrenar un perro para que ataque a cualquier persona. Esta hipótesis ya está parcialmente prevista en el Ordenamiento jurídico. En efecto, el delito de abuso de dispositivos del artículo 8° de la Ley N° 21.459 prohíbe la distribución, entre otros elementos, de programas computacionales creados o adaptados principalmente para la perpetración de alguno de los delitos informáticos contenidos en esa misma disposición. Por consiguiente, la distribución de un software inteligente, por ejemplo, basado en técnicas de *machine learning* está prohibido penalmente, siempre que la finalidad del programa sea la comisión de alguno de los delitos informáticos.

Por su parte, las acciones son todas aquellas decisiones que las máquinas pueden llevar a cabo para conseguir la finalidad que les fue propuesta. Estas acciones no solo son interacciones directas y físicas sobre el mundo real (como la que puede ejecutar el brazo mecanizado de un robot controlado por TIA), sino que estas también pueden ejecutarse en el mundo virtual. No debe olvidarse que una de las características de los algoritmos inteligentes es su posibilidad de interactuar directamente con otros algoritmos, prescindiendo de la intervención o mediación humana, lo que puede expresarse gráficamente, como lo ha hecho FLORIDI, afirmando que “las tecnologías de tercer orden (incluido el Internet de las cosas) buscan eliminarnos a nosotros, los engorrosos intermediarios humanos, del circuito”.⁴⁷ Niveles avanzados de autonomía de las máquinas inteligentes para decidir y seleccionar acciones no previamente previstas en su programación inicial, logradas a través de las capacidades de aprendizaje propias de las técnicas de *machine learning*,⁴⁸ hacen surgir la preocupación por obtener de estos dispositivos

⁴⁷ FLORIDI, cit. (n. 30), p. 31.

⁴⁸ El ‘aprendizaje’ en las técnicas de *machine learning* está vinculado con la capacidad de adaptabilidad de los algoritmos inteligentes (FLORIDI, Luciano, “Artificial Agents and Their Moral Nature”, en: FLORIDI, L. (Ed.), *Ethics, Governance, and Policies in Artificial Intelligence*, Springer, Cham, 2021, p. 231) y esta puede ser fácilmente entendida, desde la esfera del profano, con el ejemplo de la “Machine Educable Noughts and Crosses Engine”, diseñada por Donald Michie en 1963 (RUSSELL, Stuart y NORVIG, Peter, *Artificial intelligence: A modern approach*, 4º ed., Pearson, Harlow, 2022, p. 223) y cuya explicación puede encontrarse en FLORIDI y SANDERS, cit. (n. 20), pp. 359–362; en RUSSELL, David. W. (*The BOXES Methodology. Black Box Dynamic Control*, Springer, Cham, 2012, pp. 33 ss), y en FLORIDI, cit. (n. 30), pp. 231–233.

un “comportamiento normativo”,⁴⁹ esto es, ajustados a las normas. Conseguir este objetivo parece ser una tarea compleja y, a la vez, ineludible. Al decir de MALLE, SCHEUTZ Y AUSTERWEIL, “esto se debe a que los humanos exigirán que un robot colaborador comprenda las normas de su comunidad, y retirarán su confianza y cooperación si se dan cuenta de que el robot no se rige por las mismas normas que ellos”.⁵⁰

Y los ‘resultados’ o ‘outputs’, los productos que se esperan de la máquina, esto es, realizar funciones, resolver problemas, generar contenido o efectuar predicciones.⁵¹

Las capacidades y las propiedades de las máquinas inteligentes fueron expuestas previamente en la § I.

La racionalidad que justifique las acciones emprendidas por las máquinas inteligentes o de los resultados por estas generados puede no siempre ser fácilmente reconocida o explicada por operadores humanos,⁵² en las técnicas de *machine learning* y, dentro de esta, de *deep learning* y de redes neuronales. La explicabilidad es la condición que “permite la explicación de modelos de aprendizaje y se centra en por qué el sistema llegó a una decisión particular, explorando sus paradigmas lógicos, a diferencia de la naturaleza inherente de caja negra de la inteligencia artificial. De igual manera, la interpretabilidad del aprendizaje automático permite a los usuarios comprender los resultados de (estos modelos) mediante inteligencia artificial”.⁵³ La explicabilidad y la interpretabilidad configuran uno de los cuatro principios claves de la gobernanza de la IA, junto con la autonomía humana, la prevención de daños y la justicia.⁵⁴ Tanto la inexplicabilidad como la ininterpretabilidad, más probable en sistemas de aprendizaje automático son elementos que, como se verá, tienen impactos en derecho penal.

⁴⁹ OKAL, Billy; ARRAS, Kai O., “Formalizing Normative Robot Behavior”, en: AGAH, Arvin, SALICHES, M., CABIBIHAN, J.J., HE, H. y HOWARD, A.M. (Eds.), *Social Robotics. 8th International Conference, ICSR 2016 Kansas City, MO, USA, November 1–3, 2016 Proceedings*. Springer, Cham, 2016.

⁵⁰ MALLE, B.F.; SCHEUTZ, M.; y AUSTERWEIL, J.L., “Networks of social and moral norms in human and robot agents”, en: ALDINHAS, M.; SILVA, J.; TOKHI, M.; KADAR, E.; y VIRK, G. (Eds.), *A world with robots. International Conference on Robot Ethics: ICRE 2015*, Springer, Cham, 2017 p. 5.

⁵¹ FINCAN, cit. (n. 21), p. 33.

⁵² CREESE, Sadie, “The threat from AI”, en: BAKER, D.J.; ROBINSON, P.H. (Eds.), *Artificial Intelligence and the Law. Cybercrime and Criminal Liability*, Routledge, Oxon, 2021, p. 208.

⁵³ VISHWARUPE, Varad, JOSHI, Prachi M., MATHIAS, Nicole, MAHESHWARI, Shrey, MHAIASKAR, Shweta y PAWAR, Vishal, “Explainable AI and Interpretable Machine Learning: A Case Study in Perspective; *Procedia Computer Science*, 2022, Vol. 204, p. 869.

⁵⁴ BROWNSWORD, Roger, “Artificial Intelligence and Legal Singularity: The Thin End of the Wedge, the Thick End of the Wedge, and the Rule of Law”, en: DEAKIN, S.; MARKOU, Ch. (Eds.), *Is Law Computable? Critical Perspectives on Law and Artificial Intelligence*, Hart Publishing, New York, 2020, p. 142.

Finalmente, la predictibilidad se relaciona con la posibilidad de predecir las acciones o resultados de una máquina inteligente. Como informa BARFIELD, “(m)uchas tecnologías que involucran máquinas inteligentes se están volviendo autónomas y determinan de forma independiente las soluciones que utilizan para resolver problemas y alcanzar objetivos de maneras que a menudo no son predictibles de antemano por los humanos”.⁵⁵

IV. EFECTOS DE LA CATEGORÍA DE LOS ACTUANTES

De acuerdo con el modelo bipartito del derecho penal, las máquinas inteligentes o son objetos o son sujetos. Tal dualismo queda estrecho para las TIA. La consideración de las máquinas inteligentes como objetos hace que sus características produzcan problemas en términos de imputación. Por ejemplo, piénsese en un arma inteligente que, empuñada por el sujeto, decide el momento en el que percutirse y matar a otro. En este caso, ¿quién realmente mató? ¿El sujeto que empuñaba el arma inteligente o esta última? Enfrentados a esta situación, la otra opción sería considerar al arma como un sujeto. Pero esta alternativa se enfrenta a que los sistemas jurídicos aún no han avanzado hasta ese punto, y hay buenas razones para considerar inadecuado que lo hagan. Por eso, el reconocimiento de una tercera categoría intermedia (de los actuantes) presta utilidad. Esta serviría para reconstruir estructuras del derecho penal, como el efecto de los actuantes en la configuración de cursos causales, sus actuaciones como fuentes potenciales de error para el agente, la aportación de la máquina inteligente como instrumento en la autoría mediata o la recomprensión de los criterios de previsibilidad y evitabilidad en la culpa, por mencionar algunos. No es posible aquí analizar cada una de las estructuras para las que la categoría de los actuantes sería útil. Pero la mención de algunas de ellas sí es que necesaria.

La identificación de los actuantes tendría efectos en la base fáctica de la acción como primera categoría dogmática del delito, si esta es entendida como lo hace VAN WEEZEL, citando a HRUSCHKA, “como un “*factum*” en el sentido originario de la palabra, esto es, como algo que ha sido hecho por un sujeto agente”.⁵⁶ A partir del surgimiento de las TIA, estos *factum* ya no solo pueden esperarse de los agentes (personas naturales y, en algunos casos, personas jurídicas), sino también de

⁵⁵ BARFIELD, Woodrow, “Towards a law of artificial intelligence”, en: BARFIELD, W. (Ed.), *Research Handbook on the Law of Artificial Intelligence*, Edward Elgar Publishing, Cheltenham-Northampton, 2018, p. 15.

⁵⁶ VAN WEEZEL, Alex, *Curso de Derecho Penal. Parte general*, Ediciones UC, Santiago de Chile, 2023, p. 150.

dispositivos inteligentes. Es cierto que los animales e incluso la naturaleza, pueden operar esos *factum*; pero también es cierto que solo los dispositivos inteligentes han sido creados por el ser humano y que hemos desarrollado para estos ciertas capacidades que emulan las nuestras. Aceptada que la fuente de los *factum* pueden ser los agentes y las máquinas inteligentes, ello puede tener consecuencias en la forma en la que comprendemos la relación de causalidad. “Cuando la máquina está configurada para tomar sus propias decisiones” –advierte BENÍTEZ–, deberíamos prestar atención a “la ruptura del nexo causal existente entre la actividad del productor, el distribuidor o el usuario y el daño causado por la máquina”.⁵⁷ Ello puede llevar a atribuirle relevancia a los cursos causales iniciados por dichos dispositivos y no interrumpidos por un agente (por ejemplo, un usuario de una aplicación de IA), ampliándose los supuestos de omisiones relevantes; o a los casos de creación o aumento del riesgo, ya que en atención a las propias complejidades técnicas de las máquinas inteligentes, especialmente de las basadas en técnicas de *deep learning*, surgen los déficits de explicabilidad e interpretabilidad y, con ellos, la “caja negra causal”⁵⁸ que hacen imposible vincular un resultado típico con el diseñador o el usuario de una TIA, ya que estos no estarán en condiciones de explicar la decisión de la máquina, ni menos de anticiparla.

También sería útil para reconocer a las máquinas basadas en TIA como instancias creadoras de peligro o de aumento de una ya existente para efectos de la imputación objetiva en las que un agente participa directa o indirectamente sin contrarrestarlo. Un resultado típico se atribuye, entre otros criterios, a quien haya creado o aumentado el riesgo. En el caso del diseño o uso de una TIA, puede ser este dispositivo el que desarrolla esa función. Por ejemplo, el riesgo de una colisión en un vehículo convencional es propio de su conductor que aprieta el acelerador; en un vehículo autónomo, su conductor (o pasajero, en hipotéticos vehículos con niveles superiores de autonomía) no se relaciona funcionalmente con esa velocidad.⁵⁹

Podría abrirse una discusión sobre la necesidad de hacer punible la participación de un agente en un hecho ajeno, pero no iniciado por otro agente, sino por una máquina inteligente que no puede ser considerada autora, es decir, ampliar el concepto de accesoriiedad a hechos ‘pertenecientes’ a los actuantes. También

⁵⁷ BENÍTEZ, Ignacio, “Responsabilidad por el hecho cometido por el robot dotado de inteligencia artificial. ¿Existe un paralelismo con la responsabilidad de la persona jurídica? Breves reflexiones, en: MUÑOZ, J., GARCÍA, O., CEREZO, A. y GARCÍA, E. (Eds.), *Estudios jurídico-penales, criminológicos y político criminales. Libro homenaje al Profesor José Luis Díez Ripollés*, Tirant lo Blanch, Valencia, 2023, p. 503.

⁵⁸ BARRESI, cit. (n. 36), p. 275.

⁵⁹ Y esa esta la razón que ha llevado a algunas legislaciones a la creación de la figura del “user-in-charge”. Ver, al respecto, GIANNINI, Alice; KWIK, Jonathan, “Negligence Failures and Negligence Fixes. A Comparative Analysis of Criminal Regulation of AI and Autonomous Vehicles”, *Criminal Law Forum*, 2023, nº 34, pp. 66 y ss.

podría discutirse la posibilidad de ampliar el catálogo de casos que conforman la autoría mediata, a aquellos en los que el instrumento no sea un agente humano, como lo requiere esa estructura, sino una máquina inteligente, y que corresponde al modelo de responsabilidad de la “perpetración a través de otro” aplicado a las máquinas inteligentes.⁶⁰ Los *factum* de los actuantes también tendría impacto en la forma en la que entendemos y aplicamos la culpa, sobre todo considerando las posibilidades reales del agente humano de predecir la conducta del dispositivo o los resultados que este produzca materialmente.

V. CONCLUSIONES

Entre las tecnologías convencionales y las basadas en IA hay diferencias ontológicas que la funcionalidad y eficacia del derecho penal debe considerar. Las TIA ya no solo son capaces de influir culturalmente (como también lo son las convencionales), sino que también pueden crear cultura (*machine culture*⁶¹). Esta y otras funciones, que eran de monopolio exclusivo de los seres humanos, han dejado de serlo desde el surgimiento de las TIA. Una de las grandes definiciones a las que nuestra especie se enfrenta es sobre la extensión de la transferencia o delegación de competencias a las máquinas inteligentes que estamos dispuestas y dispuestos tolerar.⁶² La delegación de la acción humana en los entes inteligentes artificiales lleva consigo una transferencia de responsabilidad a las máquinas inteligentes que “puede caracterizarse como una reacción tecnológica a la excesiva complejidad de la sociedad moderna, en la que uno no sólo tiene que tomar numerosas decisiones todos los días, sino que también sabe que muchas decisiones tienen el potencial de dañar a otros”.⁶³ Las tecnologías de reconocimiento y cognitivas, mezcladas

⁶⁰ HALLEVY, Gabriel, “The criminal liability of artificial intelligence entities. From science fiction to legal social control”, *Akron Intellectual Property Journal*, 2010, vol. 4, N° 171, pp. 179-181; HALLEVY, Gabriel, “I, Robot I, Criminal: When Science Fiction Becomes Reality: Legal Liability of AI Robots Committing Criminal Offenses”, *Syracuse Science & Technology Law Reporter*, 2010, vol. 22, pp. 9-13; HALLEVY, Gabriel, “Virtual criminal responsibility”, *Original Law Review*, 2010, vol. 6, N° 1, pp. 11-13.; HALLEVY, cit. (n. 5), pp. 69 ss.

⁶¹ BRINKMANN, Levin; BAUMANN, Fabian; BONNEFON, Jean-François; DEREX, Maxime; MÜLLER, Thomas F.; NUSSBERGER, Anne-Marie; CZAPLICKA, Agnieszka; ACERBI, Alberto; GRIFFITHS, Thomas L.; HENRICH, Joseph; LEIBO, Joel Z.; McELREATH, Richard; OUDEYER, Pierre-Yves; STRAY, Jonathan; RAHWAN, Iyad, “Machine culture”, *Nature Human Behaviour*, 2023, Vol. 7, N° 11.

⁶² BEHDADI, Dorna; MUNTHE, Christian, “A Normative Approach to Artificial Moral Agency”, *Minds and Machines*, 2020, Vol. 30, N° 2, p. 214.

⁶³ BECK, Susanne, “Intelligent agents and criminal law—Negligence, diffusion of liability and electronic personhood”, *Robotics and Autonomous Systems*, 2016, Vol. 86, p. 140.

con técnicas de deep learning, “permiten un nivel de autonomía entre sistemas que comienza a cuestionar la doctrina legal establecida en muchas áreas del derecho”.⁶⁴ Como este mismo autor concluye, “esto es previsible, dado que gran parte de la doctrina legal aplicada a sistemas que operan con inteligencia artificial se promulgó en una era de máquinas de baja tecnología y no autónomas, es decir, antes de que las técnicas de inteligencia artificial comenzaran a utilizarse para una amplia gama de aplicaciones”⁶⁵. El acelerado desarrollo de las tecnologías digitales e informáticas les atribuyó un rol importante en la vida social y diaria hasta el punto “que los sistemas de información han pasado de ser una herramienta humana, a funcionar por sí mismos en lugar de los humanos”⁶⁶.

La revisión de las estructuras del derecho en general y, en particular, del derecho penal, surge como una necesidad. Esta revisión pasa atribuirle a las TIA relevancia penal en atención a su naturaleza y capacidades. Aunque estemos lejos de una IA General que pueda reemplazar al ser humano en todas las funciones, debe reconocerse que “cada vez más nos encontramos ante sistemas que adquieren conocimientos o habilidades a partir de su propia interacción con los seres vivos que habitan el entorno que los rodea”,⁶⁷ que se usan como instrumentos en la ejecución de delitos,⁶⁸ o que pueden ‘intervenir’ en la verificación de dichas conductas. La cuestión es la estrategia que es recomendable o viable adoptar para esa revisión.

Aquí se ha propuesto cuestionar el modelo bipartito del derecho penal de la Modernidad, como una solución intermedia entre el desconocimiento de la realidad marcada por la presencia de las TIA y actuar como si nada ocurriera, y otras compatibles con enfoques de carácter post-humanista que confieren a las máquinas la misma posición que la que tienen los seres humanos. La estrategia preferida aquí pasa por reconocer que las máquinas inteligentes son esencialmente diferentes de los objetos. Que, en consecuencia, dan lugar a una tercera categoría, reestructurando el modelo bipartito del derecho penal (basado en la distinción entre sujetos y objetos) a uno tripartito, integrado por sujetos, actuantes y objetos.

Esta estrategia es útil para reconfigurar estructuras del derecho penal, cuyo

⁶⁴ DSOUZA cit. (n. 2), p. 247.

⁶⁵ DSOUZA cit. (n. 2), p. 247.

⁶⁶ DEĞIRMENCI, cit. (n. 12), p. 94.

⁶⁷ PAGALLO, Ugo, “From Automation to Autonomous Systems: A Legal Phenomenology with Problems of Accountability”, en: AA.VV., *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*, AAAI Press, 2017, p. 18.

⁶⁸ BLAUTH, Tais; GSTREIN, Oskar; ZWITTER, Andrej, “Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI”, *IEEE Access*, 2022, Vol. 10; CUSTERS, Bart, “AI in Criminal Law: An Overview of AI Applications in Substantive and Procedural Criminal Law”, en: CUSTERS, B.; FOSCH-VILLARONGA, E. (Eds.), *Law and Artificial Intelligence Regulating AI and Applying AI in Legal Practice*, Asser Press - Springer, Berlin, 2022.

rendimiento se debilita si seguimos pensando en las máquinas inteligentes como meros objetos, desconociendo sus propiedades y capacidades y la posibilidad real que estas tienen de involucrarse en la ejecución de conductas que revisten los caracteres de delito. De qué forma la categoría de los actuantes influye en la reconfiguración de las estructuras del derecho penal es una cuestión que requiere de futuras investigaciones.

REFERENCIAS BIBLIOGRÁFICAS

- ABBOTT, Ryan, *The reasonable robot. Artificial Intelligence and the Law*, Cambridge University Press, Cambridge, 2020. DOI: <https://doi.org/10.1017/9781108631761>.
- ABBOTT, Ryan; SARCH, Alex, “Punishing Artificial Intelligence: Legal Fiction or Science Fiction”, *UC Davis Law Review*, 2019, Vol. 53, pp. 323-384. DOI: <https://doi.org/10.2139/ssrn.3327485>.
- ABBOTT, Ryan; SARCH, Alex, “Punishing Artificial Intelligence: Legal Fiction or Science Fiction”, en: DEAKIN, S.; MARKOU, Ch. (Eds.), *Is Law Computable? Critical perspectives on Law and Artificial Intelligence*, Hart Publishing, Oxford, 2020, pp. 177-204. DOI: <https://doi.org/10.5040/9781509937097.ch-008>.
- BARFIELD, Woodrow, “Towards a law of artificial intelligence”, en: BARFIELD, W. (Ed.), *Research Handbook on the Law of Artificial Intelligence*, Edward Elgar Publishing, Cheltenham-Northampton, 2018, pp. 2-39. DOI: <https://doi.org/10.4337/9781786439055>.
- BARRESI, Olimpia, “Italian Report on Traditional Criminal Law Categories and AI”, *Revue International de Droit Pénal*, 2023, Vol. 94, N° 1, pp. 269-302.
- BECK, Susanne, “Intelligent agents and criminal law—Negligence, diffusion of liability and electronic personhood”, *Robotics and Autonomous Systems*, 2016, Vol. 86, pp. 138-143. DOI: <https://doi.org/10.1016/j.robot.2016.08.028>.
- BEHDADI, Dorna; MUNTHE, Christian, “A Normative Approach to Artificial Moral Agency”, *Minds and Machines*, 2020, Vol. 30, N° 2, pp. 195-218. DOI: <https://doi.org/10.1007/s11023-020-09525-8>.
- BENÍTEZ, Ignacio, “Responsabilidad por el hecho cometido por el robot dotado de inteligencia artificial. ¿Existe un paralelismo con la responsabilidad de la persona jurídica? Breves reflexiones”, en: MUÑOZ, J.; GARCÍA, O.; CEREZO, A.; GARCÍA, E. (Eds.), *Estudios jurídico-penales, criminológicos y político criminales. Libro homenaje al Profesor José Luis Diez Ripollés*, Tirant lo Blanch, Valencia, 2023, pp. 493-503.
- BLAUTH, Tais; GSTREIN, Oskar; ZWITTER, Andrej, “Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI”, *IEEE Access*, 2022, Vol. 10, pp. 77110-77122. DOI: <https://doi.org/10.1109/ACCESS.2022.3191790>.
- BRINKMANN, Levin; BAUMANN, Fabian; BONNEFON, Jean-François; DEREX, Maxime; MÜLLER, Thomas F.; NUSSBERGER, Anne-Marie; CZAPLICKA, Agnieszka; ACERBI, Alberto; GRIFFITHS, Thomas L.; HENRICH, Joseph; LEIBO, Joel Z.; McELREATH, Richard; OUDEYER, Pierre-Yves; STRAY, Jonathan; RAHWAN, Iyad, “Machine culture”, *Nature*

- Human Behaviour*, 2023, Vol. 7, N° 11, pp. 1855-1868. DOI: <https://doi.org/10.1038/s41562-023-01742-2>.
- BROWNSWORD, Roger, “Artificial Intelligence and Legal Singularity: The Thin End of the Wedge, the Thick End of the Wedge, and the Rule of Law”, en: DEAKIN, S.; MARKOU, Ch. (Eds.), *Is Law Computable? Critical Perspectives on Law and Artificial Intelligence*, Hart Publishing, New York, 2020, pp. 135-160.
- BRUNDAGE, Miles; AVIN, Shahar; CLARK, Jack; TONER, Helen; ECKERSLEY, Peter; GARFINKEL, Ben; DAFOE, Allan; SCHARRE, Paul; ZEITZOFF, Thomas; FILAR, Bobby; ANDERSON, Hyrum; ROFF, Heather; ALLEN, Gregory C.; STEINHARDT, Jacob; FLYNN, Carrick; HEIGEARTAIGH, Seán; BEARD, S.J.; BELFIELD, Haydn; FARQUHAR, Sebastian; LYLE, Clare; CROOTOF, Rebecca; EVANS, Owain; PAGE, Michael; BRYSON, Joanna; YAMPOLSKIY, Roman; AMODEI, Dario, “The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation”, *ArXiv*, 2018, Vol. abs/1802.0. DOI: <https://doi.org/10.48550/arXiv.1802.07228>.
- CARROLL, Micah; CHAN, Alan; ASHTON, Henry; KRUEGER, David, “Characterizing Manipulation from AI Systems”, en: AA.VV., *EAAMO '23: Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, ACM, New York, 2023, pp. 1-13. ISBN 9798400703812. DOI: <https://doi.org/10.1145/3617694.3623226>.
- CHESTERMAN, Simon, *We, the Robots?*, Cambridge University Press, Cambridge, 2021. DOI: <https://doi.org/10.1017/9781009047081>.
- COUSÍÑO MAC IVER, Luis, *Derecho Penal chileno*, Tomo I, Editorial Jurídica de Chile, Santiago de Chile, 1975.
- CREESE, Sadie, “The threat from AI”, en: BAKER, D.J.; ROBINSON, P.H. (Eds.), *Artificial Intelligence and the Law. Cybercrime and Criminal Liability*, Routledge, Oxon, 2021, pp. 201-221. DOI: <https://doi.org/10.4324/9780429344015-9>.
- CUSTERS, Bart, “AI in Criminal Law: An Overview of AI Applications in Substantive and Procedural Criminal Law”, en: CUSTERS, B.; FOSCH-VILLARONGA, E. (Eds.), *Law and Artificial Intelligence Regulating AI and Applying. AI in Legal Practice*, Asser Press - Springer, Berlin, 2022, pp. 205-223. DOI: http://dx.doi.org/10.1007/978-94-6265-523-2_11.
- DEFENSE SCIENCE BOARD, “Dsb Summer Study on Autonomy”, 2016, disponible en línea: <https://dsb.cto.mil/wp-content/uploads/reports/2010s/DSBSS15.pdf> [cit. 31.01.2025].
- DEĞIRMENCI, Olgun, “Sufficiency of Struggling with the Current Criminal Law Rules on the Use of Artificial Intelligence in Crime”, en: KILIÇ, M.; KAHYAOĞLU, S.B. (Eds.), *Algorithmic Discrimination and Ethical Perspective of Artificial Intelligence*, Springer, Gateway East, 2024, pp. 93-104. DOI: http://dx.doi.org/10.1007/978-981-99-6327-0_6.
- DELFINO, Rebecca A., “Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn’s Next Tragic Act”, *Fordham Law Review*, 2019, Vol. 88, N°3, pp. 887-938, en línea: https://fordhamlawreview.org/wp-content/uploads/2019/12/Delfino_December_A_2.pdf (cit. 30.01.2025).
- DI NICOLA, Andrea, “Towards digital organized crime and digital sociology of organized crime”, *Trends in Organized Crime*, 2022. DOI: <https://doi.org/10.1007/s12117>.

- 022-09457-y.
- DRINNON, Caroline, "When Fame Takes Away the Right to Privacy in One's Body: Revenge Porn and Tort Remedies for Public Figures", *William & Mary Journal of Women and the Law*, 2017, N° 24, pp. 209-233.
- DSOUZA, Mark, "Don't panic. Artificial intelligence and Criminal Law", en: BAKER, D.J.; ROBINSON, P.H. (Eds.), *Artificial intelligence and the Law. Cybercrime and criminal liability*, Routledge, Oxon, 2020, pp. 247-264. DOI: <https://doi.org/10.4324/9780429344015>.
- DEL ROSAL, Bernardo, "¿El modelo de la responsabilidad penal de las personas jurídicas para los daños punibles derivados del uso de la Inteligencia Artificial?", *Revista de Responsabilidad Penal de Personas Jurídicas y Compliance*, 2023, Vol. 2, pp. 1-49, disponible en línea: <https://www.redepec.com/wp-content/uploads/pdf/-el-modelo-de-la-responsabilidad-penal-de-las-personas-jur-dicas-para-los-da-os-punibles-derivados-del-uso-de-la-inteligencia-artificial--3098.pdf> (cit. 21.01.2025).
- DE LA CUESTA, Paz, "Inteligencia artificial y responsabilidad penal", *Revista Penal México*, 2020, n° 16-17, pp. 51-62.
- EC HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE (Eds.), "A definition of IA: Main capabilities and scientific disciplines", European Comission, Bruselas, 2018, en línea: https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf (cit. 28.01.2025).
- FAHIM, Sadaf, *Ethico-Legal Aspect of AI-driven Driverless Cars*, Springer, Singapore, 2024. DOI: <https://doi.org/10.1007/978-981-97-6883-7>.
- FARHADI, Ashkan, "There is no "I" in 'AI'", *AI & Society*, 2021, Vol. 36, N° 3, pp. 1035-1046. DOI: <https://doi.org/10.1007/s00146-020-01136-2>.
- FINCAN, Müslüm, *Artificial Intelligence and Legal Issues. A Review of AI-based Legal Impasses in Terms of Criminal Law*, Duncker & Humblot, Berlin, 2022. DOI: <https://doi.org/10.3790/978-3-428-58716-2>.
- FOSCH-VILLARONCA, Eduard, *Robots, Healthcare, and the Law. Regulating Automation in PersonalCare*, Routledge, Oxon, 2020. DOI: <https://doi.org/10.4324/9780429021930>.
- FLORIDI, Luciano, *The fourth revolution. How the infosphere is reshaping human reality*. Oxford University Press, Oxford, 2014.
- FLORIDI, Luciano, "Artificial Agents and Their Moral Nature", en: FLORIDI, L. (Ed.), *Ethics, Governance, and Policies in Artificial Intelligence*, Springer, Cham, 2021, pp. 221-249. DOI: https://doi.org/10.1007/978-3-030-81907-1_12.
- FLORIDI, Luciano; SANDERS, J. W., "On the morality of artificial agents", *Minds and Machines*, 2004, Vol. 14, N° 3, pp. 349-379. DOI: <https://doi.org/10.1023/B:MIND.0000035461.63578.9d>.
- FREITAS, Pedro Miguel; ANDRADE, Francisco; NOVAIS, Paulo, "Criminal liability of autonomous agents: From the unthinkable to the plausible", en: CASANOVAS, P.; PAGALLO, U.; PALMIRANI, M.; SARTOR, G. (Eds), *AI Approaches to the Complexity of Legal Systems. AICOL 2013. Lecture Notes in Computer Science*, Vol 8929, Springer, Berlin - Heidelberg, 2014, pp. 145-156. DOI: <https://doi.org/10.1007/978-3-662-45960-7>.
- GEETHA, T.V.; SENDHILKUMAR, S., *Machine Learning. Concepts, Techniques and Applications*, Chapman-Hall-CRC, Boca Raton, 2023.

- GIANNINI, Alice; KWIK, Jonathan, “Negligence Failures and Negligence Fixes. A Comparative Analysis of Criminal Regulation of AI and Autonomous Vehicles”, *Criminal Law Forum*, 2023, n° 34, pp. 43-85.
- HALE, Benjamin, “Technology, the Environment and the Moral Considerability of Artefacts”, en: BERG OLSEN, J.; SELINGER, E.; RIIS, S. (Eds.), *New Waves in Philosophy of Technology*, Palgrave Macmillan, London, 2009, pp. 216-240. DOI: https://doi.org/10.1057/9780230227279_11.
- HALLEVY, Gabriel, “I, Robot I, Criminal: When Science Fiction Becomes Reality: Legal Liability of AI Robots Committing Criminal Offenses”, *Syracuse Science & Technology Law Reporter*, 2010, Vol. 22, pp. 1-37.
- HALLEVY, Gabriel, “The criminal liability of artificial intelligence entities. From science fiction to legal social control”, *Akron Intellectual Property Journal*, 2010, Vol. 4, N° 171, pp. 171-201.
- HALLEVY, Gabriel, “Virtual criminal responsibility”, *Original Law Review*, 2010, Vol. 6, N° 1, pp. 6-27.
- HALLEVY, Gabriel, *When Robots kill: Artificial intelligence under criminal law*, Northeastern University Press, Lebannon NH, 2013.
- HAYWARD, Keith J.; MAAS, Matthijs, “Artificial intelligence and crime: A primer for criminologists”, *Crime, Media, Culture*, 2020, Vol. 17, N° 2, pp. 209-233. DOI: <https://doi.org/10.1177/1741659020917434>.
- INATANI, Tatsuhiko, “‘Moralizing Technology’ and Criminal Law Theory”, en: BORGES, G., SORGE, C. (Eds.), *Law and Technology in a Global Digital Society. Autonomous Systems, Big Data, IT Security and Legal Tech*, Springer, Cham, 2022, pp. 27-49. DOI: https://doi.org/10.1007/978-3-030-90513-2_2.
- KAUSHIK, Purva, “Privacy and Other Legal Concerns in the Wake of Deepfake Technology: Comparative Study of India, US and China”, en: DEWANI, N.D.; KHAN, Z.A.; AGARWAL, A.; KHAN, S.A. (Eds.), *Handbook of Research on Cyber Law, Data Protection, and Privacy*, IGI Global, Hershey, 2022, pp. 37-49. DOI: <https://doi.org/10.4018/978-1-7998-8641-9.ch003>.
- KING, Thomas C., “Projecting AI-Crime: A Review of Plausible Threats”, en: ÖHMAN, C.; WATSON, D. (Eds.), *The 2018 Yearbook of the Digital Ethics Lab*, Springer, Cham, 2019, pp. 65-84. DOI: https://doi.org/10.1007/978-3-030-17152-0_6.
- KING, Thomas C.; AGGARWAL, Nikita; TADDEO, Mariarosaria; FLORIDI, Luciano, “Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions”, *Science and Engineering Ethics*, 2020, Vol. 26, n° 1, pp. 89-120. DOI: <https://doi.org/10.1007/s11948-018-00081-0>.
- KING, Thomas C.; AGGARWAL, Nikita; TADDEO, Mariarosaria; FLORIDI, Luciano, “Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions”, en: COWLS, J.; MORLEY, J. (Eds.), *The 2020 Yearbook of the Digital Ethics Lab*, Springer, Cham, 2021, pp. 195-227. DOI: https://doi.org/10.1007/978-3-030-80083-3_14.
- KIRPICHNIKOV, Danila; PAVLYUK, Albert; GREBNEVA, Yulia; OKAGBUE, Hilary, “Criminal Liability of the Artificial Intelligence”, *EDP Sciences*, 2020, Vol. 159, pp. 1-10. DOI: <https://doi.org/10.1051/e3sconf/202015904025>.
- KÖNIG, Pascal D.; KRAFFT, Tobias D.; SCHULZ, Wolfgang; y ZWEIG, Katharina A., “Essence

- of AI. What Is AI?", en: DIMATTEO, L.; PONCIBÒ, C.; CANNARSA, M. (Eds.), *The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press, Cambridge, 2022, pp. 18-34. DOI: <https://doi.org/10.1017/9781009072168.005>.
- LAGIOIA, Francesca; SARTOR, Giovanni, "AI Systems Under Criminal Law: A Legal Analysis and a Regulatory Perspective", *Philosophy and Technology*, 2020, vol. 33, N° 3, pp. 433-465. DOI: <https://doi.org/10.1007/s13347-019-00362-x>.
- LA ROSA, Emanuele, "Criminal Law and Artificial Intelligence: Which Spaces for a Criminal Liability of the Robot?", en: MARINO, D.; MONACA, M. (Eds.), *Artificial Intelligence and Economics: the Key to the Future*, Springer, Cham, 2023, pp. 173-187. DOI: https://doi.org/10.1007/978-3-031-14605-3_14.
- MAHER, Sean, "Deep Fakes. Seeing and Not Believing", en: FILIMOWICZ, M. (Ed.), *Deep Fakes. Algorithms and Society*, Routledge, London - New York, 2022, pp. 1-22. DOI: <https://doi.org/10.4324/9781003173397-1>.
- MALLE, B.F.; SCHEUTZ, M.; AUSTERWEIL, J.L., "Networks of social and moral norms in human and robot agents", en: ALDINHAS, M.; SILVA, J.; TOKHI, M.; KADAR, E.; y VIRK, G. (Eds.), *A world with robots. International Conference on Robot Ethics: ICRE 2015*, Springer, Cham, 2017, pp. 3-17. DOI: https://doi.org/10.1007/978-3-319-46667-5_1.
- MARTINEZ-MIRANDA, Enrique; MCBURNETT, Peter; HOWARD, Matthew, "Learning unfair trading: A market manipulation analysis from the reinforcement learning perspective", en: AA.VV., *2016 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, Natal, Brasil, 2016, pp. 103-109. DOI: <https://doi.org/10.1109/EAIS.2016.7502499>.
- MANDEL, Gregory N., "History Lessons for a General Theory of Law and Technology", *Minnesota Journal of Law, Science & Technology*, 2007, Vol. 8, n° 2, pp. 551-570. Disponible en <https://scholarship.law.umn.edu/mjlst/vol8/iss2/10>.
- MIZUTA, Takanobu, "Can an AI perform market manipulation at its own discretion? – A genetic algorithm learns in an artificial market simulation", en: AA.VV., *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, ACT, Canberra, 2020, pp. 407-412. DOI: <https://doi.org/10.1109/SSCI47803.2020.9308349>.
- OKAL, Billy; ARRAS, Kai O., "Formalizing Normative Robot Behavior", en: AGAH, Arvin; SALICHES, M.; CABIBIHAN, J.J.; He, H.; y HOWARD, A.M. (Eds.), *Social Robotics. 8th International Conference, ICSR 2016 Kansas City, MO, USA* (November 1–3, 2016) *Proceedings*, Springer, Cham, 2016, pp. 62-71. DOI: https://doi.org/10.1007/978-3-319-47437-3_7.
- PAGALLO, Ugo, "From Automation to Autonomous Systems: A Legal Phenomenology with Problems of Accountability", en: AA.VV., *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*, AAAI Press, Melbourne, 2017, pp. 17-23, en línea: <https://doi.org/10.24963/ijcai.2017/3>.
- PAGALLO, Ugo, "Killers, fridges, and slaves: A legal journey in robotics", *AI and Society*, 2011, Vol. 26, N° 4, pp. 347-354. DOI: <https://doi.org/10.1007/s00146-010-0316-0>.
- PAGALLO, Ugo, "Robots of Just War: A Legal Perspective", *Philosophy & Technology*, 2011, Vol. 24, N° 3, pp. 307-323. DOI: [10.1007/s13347-011-0024-9](https://doi.org/10.1007/s13347-011-0024-9).
- PAGALLO, Ugo, *The Laws of Robots. Crimes, Contracts, and Torts*, Springer, Dordrecht, 2013. DOI: <https://doi.org/10.1007/978-94-007-6564-1>.

- PAGALLO, Ugo; QUATTROCOLO, Serena, “The impact of AI on criminal law, and its twofold procedures, en: BARFIELD, Woodrow (Ed.), *Research Handbook on the Law of Artificial Intelligence*, Edward Elgar Publishing, Cheltenham, 2018, pp. 385-409. DOI: <https://doi.org/10.4337/9781786439055.00026>.
- PERIN, Andrea, “La “inteligencia artificial” en la justicia penal ante el principio de responsabilidad penal”, en: ROMEO, C.; RUEDA, M. (Eds.), *Derecho penal, ciberseguridad, ciberdelitos e inteligencia artificial. Volumen II: inteligencia artificial y responsabilidad penal*, Comares, Granada, 2023, pp. 83-118.
- QUINTERO, Gonzalo; “La robótica ante el Derecho penal: el vacío de respuesta jurídica a las desviaciones incontroladas”, *Revista Electrónica de Estudios Penales y de la Seguridad*, 2017, N° 1, pp. 1-23, disponible en línea: www.ejc-reeps.com.
- REAL ACADEMIA ESPAÑOLA, *Gramática descriptiva de la lengua española* (Bosque, I. y Demonte, V., Eds.), Espasa, Madrid, 1999.
- REAL ACADEMIA ESPAÑOLA, *Nueva gramática de la lengua española, Morfología y sintaxis I*, Espasa, Madrid, 2009.
- REYNOLDS, Carson; ISHIKAWA, Masatoshi, “Robotic Thugs”, en AA.VV., *ETHICOMP2007*, Global e-SCM Research Center y Meiji University, Tokyo, 2007, pp. 487-492.
- ROMEO, Carlos, “La atribución de responsabilidad penal por los hechos cometidos por sistemas autónomos inteligentes, robótica y tecnologías conexas”, *ULP Law Review*, 2022, vol. 16, N° 1, pp. 7-16.
- ROMEO, Carlos, “La atribución de responsabilidad penal por los hechos cometidos por sistemas autónomos inteligentes, robótica y tecnologías conexas”, en: MUÑOZ, J.; GARCÍA, O.; CEREZO, A.; y GARCÍA, E. (Eds.), *Estudios jurídico-penales, criminológicos y político-criminales. Libro homenaje al Profesor José Luis Díez Ripollés*, Tirant lo Blanch, Valencia, 2023, pp. 695-707.
- ROMEO, Carlos, “La discusión sobre la atribución de responsabilidad penal a los sistemas de inteligencia artificial, en particular a sistemas autónomos”, en: ROMEO, C.; RUEDA, M. (Eds.), *Derecho penal, ciberseguridad, ciberdelitos e inteligencia artificial. Volumen II: inteligencia artificial y responsabilidad penal*. Comares, Granada, 2023, pp. 57-81.
- RUSSELL, David W., *The BOXES Methodology. Black Box Dynamic Control*, Springer, Cham, 2012. DOI: <https://doi.org/10.1007/978-1-84996-528-6>.
- RUSSELL, Stuart; NORVIG, Peter, *Artificial intelligence: A modern approach*, 4a ed., Pearson, Harlow, 2022.
- SABOURET, Nicolas, “What Artificial Intelligence Can Do and What It Cannot Do”, en: BOUNFOUR, Ahmed (Ed.), *Platforms and Artificial Intelligence. The Next Generation of Competences*, Springer, Cham 2023, pp. 183-192. DOI: https://doi.org/10.1007/978-3-030-90192-9_8.
- SCHARRE, Paul; HOROWITZ, Michael C., “An Introduction to Autonomy in Weapon Systems”, 2015, disponible en línea: https://www.files.ethz.ch/isn/188865/Ethical%20Autonomy%20Working%20Paper_021015_v02.pdf (citada: 05.01.2025).
- SEARLE, John, “Minds, brains, and programs”, *Behavioral and Brain Sciences*, 1980, N° 3, pp. 417-424.
- SIBAI, Fadi N., “AI Crimes: a classification”, en: IEEE (Eds.), *2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, Dublin, Ireland,

- 2020, pp. 1-8. DOI: <https://doi.org/10.1109/CyberSecurity49315.2020.9138891>.
- SULLINS, John P., "When is a robot a moral agent?", *International Review of Information Ethics*, 2006, Vol. 6, N° 12, pp. 23-30. DOI: <https://doi.org/10.29173/irie136>.
- SULLINS, John P., "When is a robot a moral agent?", en: ANDERSON, M.; ANDERSON, Susan (Eds.), *Machine ethics*, Cambridge University Press, Cambridge, 2011, pp. 151-161. DOI: <https://doi.org/10.1017/cbo9780511978036.013>.
- URBINA, Fabio; LENTZOS, Filippa; INVERNIZZI, Cédric; y EKINS, Sean, "Dual use of artificial-intelligence-powered drug Discovery", *Nature Machine Intelligence*, 2022, N° 4, pp. 189-191. DOI: <https://doi.org/10.1038/s42256-022-00465-9>.
- VERBEEK, Peter-Paul, *Moralizing Technology. Understanding and Designing the Morality of Things*, University of Chicago Press, Chicago - London, 2011. DOI: <https://doi.org/10.7208/chicago/9780226852904.001.0001>.
- VISHWARUPE, Varad; JOSHI, Prachi M.; MATHIAS, Nicole; MAHESHWARI, Shrey; MHAISALKAR, Shweta; y PAWAR, Vishal, "Explainable AI and Interpretable Machine Learning: A Case Study in Perspective", *Procedia Computer Science*, 2022, Vol. 204, pp. 869-876. DOI: <https://doi.org/10.1016/j.procs.2022.08.105>.
- WANG, Pei, "On Defining Artificial Intelligence", *Journal of Artificial General Intelligence*, 2019, Vol. 10, N° 2, pp. 1-37. DOI: <https://doi.org/10.2478/jagi-2019-0002>.
- VAN WEEZEL, Alex, *Curso de Derecho Penal. Parte general*, Ediciones UC, Santiago de Chile, 2023.
- VAN DER SLOOT, Bart; WAGENVELD, Yvette, "Deepfakes: regulatory challenges for the synthetic society", *Computer Law & Security Review*, 2022, Vol. 46, p. 105716. DOI: <https://doi.org/10.1016/j.clsr.2022.105716>.



El contenido de la *Revista de Derecho Universidad de Concepción* se publica bajo la licencia Creative Commons Atribución 4.0 Internacional, y puede usarse gratuitamente, dando los créditos a los autores y a la revista, conforme a esta licencia.