El uso de las armas autónomas: ¿una zona de impunidad penal?[1]

The use of autonomous weapons: a zone of criminal impunity?

Ana Isabel GARCÍA ALFARAZ
Profesor Ayudante Doctor de Derecho penal de la Universidad de Salamanca

Abstract

The development of lethal autonomous weapons represents one of the greatest challenges currently facing both the military and legal spheres. These systems, capable of selecting and attacking targets without meaningful human control, are transforming the nature of armed conflict by introducing a new technological dimension into decision-making. Unlike conventional weapons, whose performance ultimately depends on human oversight, autonomous weapons grant artificial intelligence and learning algorithms the capacity to make rapid decisions and to modify their behaviour in ways that may be unpredictable to programmers or operators.

Their use raises significant questions under international humanitarian law, particularly with regard to the principles of distinction and proportionality, which aim to limit the effects of war and to protect the civilian population. Likewise, from the perspective of criminal law, difficulties arise in attributing criminal responsibility for war crimes or violations of international norms when operations are adopted and carried out by autonomous systems rather than directly identifiable individuals. This phenomenon calls into question the foundations of liberal criminal law, which is premised on the attribution of criminal responsibility to natural persons, and makes it necessary to reconsider existing legal frameworks in order to prevent impunity.

Keywords: artificial intelligence, criminal liability, criminal law, lethal autonomous weapons (LAWS), responsibility gap, plutophilia

[1] Este artículo fue escrito durante la estancia de investigación en el Dipartimento di Giurisprudenza dell'Università degli Studi di Palermo.

Resumen

El desarrollo de las armas autónomas constituye uno de los mayores desafíos actuales en el ámbito militar y jurídico. Estos sistemas, capaces de seleccionar y atacar objetivos sin intervención humana directa, están transformando la naturaleza de los conflictos armados al introducir una nueva dimensión tecnológica en la toma de decisiones. A diferencia de las armas convencionales, cuya actuación depende en última instancia del control humano, las armas autónomas confieren a la inteligencia artificial y a los algoritmos de aprendizaje la capacidad de decidir rápidamente, pudiendo modificar su comportamiento y que éste sea impredecible para el programador o el operador.

La utilización de las armas autónomas plantea interrogantes en torno al Derecho internacional humanitario, especialmente respecto a los principios de distinción y proporcionalidad, los cuales persiguen limitar los efectos de la guerra y proteger a la población civil. Asimismo, desde la perspectiva del Derecho penal, surge la dificultad de atribuir responsabilidad penal por los crímenes de guerra o las violaciones de las normas internacionales cuando las operaciones son adoptadas y ejecutadas por sistemas autónomos y no por individuos directamente identificables. Este fenómeno cuestiona los cimientos del Derecho penal liberal basado en la atribución de la responsabilidad penal a las personas físicas, siendo preciso, por tanto, reflexionar sobre los marcos jurídicos existentes para evitar en último término la impunidad.

Palabras clave: armas autónomas, atribución de la responsabilidad penal, inteligencia artificial, Derecho penal, impunidad, plutofilia

SUMARIO: I. INTRODUCCIÓN: DE LOS CONFLICTOS ARMADOS EN LA ERA DE INTELIGENCIA ARTIFICIAL. II. LAS ARMAS AUTÓNOMAS Y EL DERECHO INTERNACIONAL HUMANITARIO. III. EL DERECHO PENAL ANTE LAS ARMAS AUTÓNOMAS: LA ATRIBUCIÓN DE RESPONSABILIDAD PENAL. 1.- Atribución de la responsabilidad a los propios sistemas de inteligencia artificial autónomos. a) Las armas autónomas: ¿un sujeto jurídico responsable penalmente? b) La capacidad de acción. c) La culpabilidad de los entes artificiales. d) La punibilidad de los entes artificiales. 2.- Atribución de la responsabilidad penal a personas físicas por los daños provocados por sistemas de inteligencia artificial autónomos. a) La acción penalmente relevante y el problema del dominio. b) Acción sin dominio: la autoría mediata frente a algoritmos autónomos. c) Autoría mediata por dominio de la organización tecnológica. d) Gestión corporativa de riesgos y *compliance*. e)

La omisión impropia en contextos algorítmicos. f) La responsabilidad por el producto. g) Recapitulando: a modo de balance general. V. REFLEXIONES FINALES.

I. INTRODUCCIÓN: DE LOS CONFLICTOS ARMADOS EN LA ERA DE INTELIGENCIA ARTIFICIAL

A lo largo de la historia se constata que la convivencia de diferentes grupos plantea conflictos, evidenciándose en cierta forma su inevitabilidad (Olasolo Alonso, 2015, p. 27). Así, junto a los acuerdos o la propuesta de tratados de paz los Estados invierten recursos intelectuales y materiales en las estrategias de defensa o de ataque. Evidentemente, los conflictos armados, al igual que cualquier otro aspecto social no es ajeno a los avances de la ciencia. La evolución de los conflictos armados ha estado y está marcada por la incorporación de las "nuevas" tecnologías, dirigidas al diseño, la fabricación y el despliegue, ya sea virtual o real, de armas cada vez más sofisticadas (Battistelli, 2023b) para mejorar la precisión y eficacia de las armas así como para demostrar el poderío bélico de los Estados (Meza Rivas, 2022, p. 50), reescribiendo así el desarrollo de los conflictos armados (Areola García, 2022, p. 188) y transformando radicalmente el escenario bélico.

La irrupción de la inteligencia artificial ha supuesto el paso a guerras tecnológicas en cuanto que adquieren una mayor relevancia los ciberataques, se asiste a un uso masivo de drones armados y vigilancia digital, se emplean robots militares no letales como apoyo logístico o labores de reconocimiento hasta llegar al estallido de las armas autónomas[2], entendidas como sistemas equipados con inteligencia artificial capaces de razonar y actuar en ausencia del ser humano que los entrenó[3], considerando simultáneamente múltiples variables para tomar la mejor decisión en cada situación (Scharre & Horowitz, 2015, p. 5). Decisión que se adopta prácticamente de forma inmediata, porque otra característica es que estos sistemas no sólo consideran múltiples factores, sino que los tiempos de análisis y de respuesta son mucho más

[2] Existen diferentes expresiones en la literatura para hacer referencia a esta realidad: LAWS (*lethal autonomous weapon systems*), FAWS (*fully autonomous weapon systems*), *lethal autonomous weapons*, LAR (*lethal autonomous robots*) o *killer robots*, entre otras.

[3] En este sentido, se considera que ya en 2020 se utilizó el primer dron con inteligencia artificial autónoma en combate (el Dron STM Kargu-2) en Libia (UN. Panel of Experts Established pursuant to Security Council Resolution 1973 (2011), 8, p. 17).

reducidos que los que emplearía el ser humano, acortándose así significativamente el clásico proceso de las operaciones militares de observación-decisión-acción (Siroli, 2023, p. 62). Esta constante evolución nos acerca a una utopía: participar o intervenir en conflictos armados sin que éstos comporten la pérdida de vidas humanas entre su ejército. Este ha sido el sueño desde los albores de la historia de la humanidad: un *deus ex machina*, un "alguien" o un "algo" dotado de poderes excepcionales que intervendría en el conflicto con todo su poder e inclinaría la balanza a favor de su ejército y en detrimento de aquel del enemigo (Battistelli, 2023a, p. 21).

Obviamente, todo progreso exige una profunda reflexión sobre lo que es lícito hacer y lo que no se debe hacer y, evidentemente, la inteligencia artificial no escapa de estas consideraciones. Es más, la aparición de la inteligencia artificial no es simplemente una evolución más, sino que por primera vez estamos ante un avance que puede implicar la superación del ser humano, de estar incluso por encima de su control. En este sentido, Kurzweil predice su singularidad, esto es, que las máquinas serán lo suficientemente inteligentes como para programarse y mejorarse a sí mismas llegando a independizarse de sus programadores humanos (Kurzweil, 2015).

Las aplicaciones de la inteligencia artificial están generando nuevos problemas éticos y jurídicos ligados a la protección de datos, la propia identidad del ser humano o los derechos humanos. En el campo de los conflictos armados se plantean igualmente importantes cuestiones como, por ejemplo, ¿cuántas muertes de civiles constituyen daños colaterales aceptables? ¿matar o no matar a personas que parecen combatientes enemigos, pero que quizás no lo son?, etc. (Parisi, 2023, p. 14). Sin duda, estos desafíos se agravan por el actual contexto de modernidad líquida como indica Bauman, en cuanto que ésta es cambiante e incierta, en las que nada es estable (Bauman, 2007), donde los usos y aplicaciones de la inteligencia artificial constituyen un campo en constante y vertiginosa evolución. Pero, además, a diferencia de la modernidad clásica o sólida, la racionalidad científica en la modernidad líquida no sólo indica los medios sino también los fines. Así, los avances tecnológicos en el ámbito militar han logrado hacer fluctuar la responsabilidad y, por lo tanto, despersonalizarla o deshumanizarla. Basta pensar en el hecho de que los misiles y drones inteligentes han sustituido tanto a las tropas como a los mandos del ejército en la selección de objetivos y en la toma de decisiones, consiguiendo

así una neutralización de la evaluación moral. Se desconecta de los hechos a la par que suprime el sentimiento de culpa o la responsabilidad moral para poder así justificar comportamientos que, de otra forma, serían inaceptables. Además, la justificación también se puede encontrar en el propio funcionamiento de los drones. Cuando un dron envía al operador una ingente cantidad de información que no puede procesar en tiempo real, no sólo le proporciona la información, sino también le exonera de la culpa, de la responsabilidad moral que sufriría si tuviera toda la información y tuviera que decidir, por ejemplo, la ejecución de determinadas personas, pero además le garantiza que, si comete un error, no será acusado de inmoralidad (Bauman & Lyon, 2013, pp. 76-78).

En pocas palabras, el empleo de armas autónomas supone un importante desafío que precisa la atención del Derecho (penal) no sólo por constituir una amenaza para bienes jurídicos fundamentales, sino también por su capacidad de intervenir sin control humano, la opacidad algorítmica, la variabilidad conductual, la imprevisibilidad derivada del autoaprendizaje, la dependencia de datos y la dificultad para reconstruir la cadena causal. Aspectos todos ellos que dificultan aún más la atribución de responsabilidad penal ante la posible afectación de bienes jurídicos como consecuencia de las conductas llevadas a cabo por las máquinas, por las armas autónomas.

II. LAS ARMAS AUTÓNOMAS Y EL DERECHO INTERNACIONAL HUMANITARIO

No existe un único concepto de armas autónomas, ni tampoco consenso sobre el contenido y alcance de los componentes que la integran, sino infinidad de definiciones muy dispares entre sí (Meza Rivas, 2022, pp. 82 ss). En este artículo se asumirá la definición elaborada por el Comité Internacional de la Cruz Roja que considera sistemas de armas autónomos “las armas que seleccionan y aplican la fuerza a objetivos sin intervención humana” (Comité Internacional de la Cruz Roja, 2021, p. 5). De acuerdo con este concepto las armas autónomas una vez activadas, pueden detectar, decidir y ejecutar un ataque basándose en algoritmos de inteligencia artificial, sensores y software, sin necesidad de que una persona apruebe cada acción concreta.

La decisión la toma la propia máquina siguiendo la programación establecida, pero sin que exista un mandato humano en esa situación concreta (*man-off-the-loop*); es decir, el operador no decide cada acción (*man-in-the-loop*) y tampoco supervisa la acción pudiendo intervenir (*man-on-the-loop*) (Battistelli, 2023b) (Scharre & Horowitz, 2015) (Siroli, 2023, p. 61) (Van Severen & Vander Maelen, 2021). Así, las armas autónomas que integran inteligencia artificial y que poseen la capacidad de aprender y mejorar a partir de la experiencia (Russell & Norvig, 2004), pueden basarse tanto en técnicas de *machine learning* tradicionales como en arquitecturas de *deep learning*, especialmente cuando requieren procesos avanzados de percepción, clasificación o reconocimiento. En estos sistemas, el aprendizaje automático permite que la propia tecnología ajuste y optimice su comportamiento con base en nuevas entradas de información, lo que implica que no sólo actúan sin control humano directo, sino que también pueden modificar sus pautas de actuación de forma dinámica y, por ende, dificultar que quienes las utilicen puedan predecir su comportamiento, complicando, como se analizará, el proceso de atribución de responsabilidad.

Estas armas presentan un enorme potencial de destrucción, pero su singularidad no reside tanto en este aspecto, sino en su autonomía, en la rapidez en el análisis de los datos y la toma de la decisión, o en una mayor eficacia y precisión de las operaciones militares, lo cual se traduce a su vez en una fácil accesibilidad a zonas inaccesibles u hostiles, una menor probabilidad de pérdidas humanas (al menos para quien utiliza estas armas), un menor coste o la posibilidad de operar en enjambre, en conjunto con otros sistemas, coordinando ataques múltiples. No obstante, su uso no está exento de sombras como la posible agravación o difusión de los conflictos armados, las crisis humanitarias derivadas, los eventuales errores de cálculo, las violaciones de los principios del Derecho internacional humanitario o la dificultad para atribuir responsabilidad en caso de errores o daños colaterales.

La irrupción de la inteligencia artificial con fines civiles y/o militares forma parte de nuestra vida diaria (Meza Rivas, 2022, p. 74) (Miró Llinares, 2018) e incluso se promueve el desarrollo y empleo de las armas autónomas en los actuales conflictos armados (Sehrawat, 2017, p. 38).

De hecho, son muchos los Estados que las utilizan en la actualidad[4], si bien esencialmente con fines defensivos y con un comportamiento predecible (Bertieri & Iaria, 2023, p. 142). Sin embargo, pese a su difusión no existe un marco jurídico que aborde el uso de las armas autónomas.

La UE en su reciente Reglamento (UE) 2024/1689, conocido como la Ley de Inteligencia Artificial, excluye expresamente en su art. 2.3 a los sistemas de inteligencia artificial con fines exclusivamente militares (también a los de defensa o de seguridad nacional). En principio, está exclusión estaría justificada sobre la base de que están sujetas al Derecho internacional público, un marco jurídico que se considera más adecuado para regular los sistemas de inteligencia artificial en el marco de operaciones militares o de defensa (Costa, 2023). Previamente, el Parlamento Europeo, en la Resolución del Parlamento Europeo, 12 de septiembre 2018, sobre los sistemas armamentísticos autónomos (2018/2572(RSP)), insistía en "la importancia fundamental de impedir el desarrollo y la producción de sistemas armamentísticos autónomos letales desprovistos de control humano con respecto a funciones críticas como las de seleccionar y atacar objetivos" (aptdo. 4). Se trata de una postura claramente en contra de las armas autónomas, abogando exclusivamente por la existencia de armas sometidas al control humano. Bajo esta perspectiva se defiende por tanto que sólo las personas humanas pueden ser responsables de los daños ocasionados por el uso de las armas autónomas.

Igualmente, en otra Resolución del Parlamento Europeo, de 20 de enero de 2021, sobre inteligencia artificial (2020/2013(INI)), el Parlamento Europeo reconoce el impacto que la inteligencia artificial está ejerciendo "en la forma de operar de los ejércitos debido, principalmente, a la integración y el uso de nuevas tecnologías y capacidades autónomas" (Considerando B) a la par que puntualiza que estas aplicaciones deben estar centradas en el ser humano, al servicio de la humanidad (Considerando E). Por ello, el Parlamento Europeo incide de nuevo en que los sistemas de inteligencia artificial deben estar sujetos al control humano garantizando en todo momento "los medios para corregir su curso, detenerla o desactivarla en

[4] Sirvan como, por ejemplo, el *Iron Dome*, desarrollado por Israel, el *Goalkeeper Close-In Weapon System* de los Países Bajos, o el sistema *Nächstbereichschutz (NBS) MANTIS*, utilizado por Alemania.

caso de comportamiento imprevisto, intervención accidental, ciberataque o interferencia de terceros con tecnología basada en la IA, o cuando terceros adquieran dicha tecnología" (aptdo. 3). Asimismo, restringe su ámbito de actuación estableciendo que los sistemas de armas autónomos letales "deben emplearse únicamente como último recurso y solo son lícitos si están sujetos a un estricto control humano" (aptdo. 34) y "en casos especificados claramente y conforme a procedimientos de autorización establecidos de antemano y de forma detallada" (aptdo. 37). En principio, esta postura humanista de la UE desconoce la realidad: de una parte, el desarrollo y la difusión de las armas autónomas en los actuales conflictos armados; y de otra, el hecho de que a nivel internacional no existe una prohibición en el uso de las armas autónomas.

A nivel internacional, destacan las iniciativas de la ONU, dentro de las reuniones de la Convención sobre Ciertas Armas Convencionales (conocida por su sigla en inglés, CCW) para fijar el marco jurídico de las armas autónomas. Sin embargo, todavía no han tenido éxito. Es más, ni siquiera existe un consenso sobre qué es un sistema autónomo, cuáles son sus características técnicas o qué se entiende por control humano (UN, 2025). Los Estados se muestran divididos -defendiendo desde una regulación más o menos estricta hasta la prohibición total- y muchos de ellos no están interesados en autolimitarse, conscientes de que quien domine la inteligencia artificial, dominará el mundo como indicaba Vladimir Putin en un discurso en 2017 (Gigova, 2017).

La ausencia de una regulación vinculante específica de las armas autónomas no implica que se sustraiga del marco regulador general de los conflictos armados. El Derecho internacional humanitario se aplica "plenamente a todos los sistemas de armas, incluido el posible desarrollo de sistemas de armas autónomos letales" (ONU, 2019, p. 11). Debe advertirse que para la aplicación del Derecho internacional humanitario las armas autónomas se consideran armas y no combatientes. No obstante, también se defiende la postura opuesta, que las armas autónomas sean consideradas combatientes con base en su autonomía y capacidad de tomar decisiones sin control humano (Shilo, 2018), es decir, que se está ante sujetos y no ante meros instrumentos.

Los conflictos armados suponen una mutación del ordenamiento jurídico. Ya Cicerón indicaba: *inter arma enim silent leges*, durante un conflicto armado, las leyes civiles pueden suspenderse o ignorarse, se pueden interrumpir o abolir determinados derechos considerados inalienables en tiempos de paz, imponiéndose por la fuerza y la amenaza de las armas la voluntad de una de las partes (Liñán Lafuente, 2017, p. 265). Los Estados, por tanto, pueden recurrir a armas, medios y métodos de combate para usar legítima y legalmente su fuerza militar en las hostilidades o conflictos armados (Meza Rivas, 2022, p. 133). Pero, los Estados no tienen un poder ilimitado para hacer daño al enemigo en un conflicto armado, sino que están limitados por principios y reglas de obligado cumplimiento. Así, por ejemplo, en los conflictos armados se permite matar a otras personas (enemigos) si se respetan las normas del combate, el *ius in bello* o el Derecho internacional humanitario. En cambio, cuando en un conflicto armado se llevan a cabo conductas que constituyen una violación grave del Derecho internacional humanitario y han sido tipificadas como delito en virtud de tratados internacionales o del Derecho internacional consuetudinario, estamos ante crímenes de guerra (Morello, 2022, p. 185).

El Derecho internacional humanitario es:

> un conjunto normativo que persigue controlar jurídicamente el fenómeno bélico -reglamentando los métodos y los medios de combate, distinguiendo entre personas y bienes civiles y objetivos militares, protegiendo a las víctimas y a quienes las asisten-, con vistas a limitar en la mayor medida posible los ingentes males que el mismo causa en los seres humanos (Pérez González, 2017, p. 33).

La esencia del Derecho internacional humanitario es el principio de distinción, que consiste en la diferenciación entre civiles y combatientes y entre objetivos civiles y militares (art. 48 PA I), prohibiéndose, por tanto, los ataques indiscriminados. Este principio es determinante para decidir si los ataques a las personas o bienes pueden ser calificados como crímenes de guerra en un conflicto armado. No obstante, en la práctica resulta difícil delimitar entre bienes civiles y objetivos militares debido al doble uso que puede tener una infraestructura cibernética, la interconectividad entre los sistemas informáticos militares y civiles (Ambos, 2015, p. 15) o, la interrelación con el contexto, por ejemplo, si una escuela o una iglesia se utiliza como protección por los combatientes se convertiría en objetivo militar.

El otro pilar fundamental del Derecho internacional humanitario es el principio de proporcionalidad, el cual prohíbe aquellos medios y métodos de guerra que causen males superfluos o sufrimientos innecesarios, es decir, daños colaterales excesivos en relación con la ventaja militar concreta y directamente prevista.

Junto a estos dos principios básicos, aparece el principio de precaución que persigue minimizar los daños civiles en la mayor medida posible, presentando dos vertientes: precauciones en el ataque (art. 57 PA I) y precauciones contra los efectos de los ataques (art. 58 PA I). Así, la omisión en la adopción de estas cautelas puede transformar en crimen de guerra lo que de otro modo sería un ataque permitido contra una persona o bien de carácter civil.

Presentado de modo general el Derecho internacional humanitario, resulta conveniente analizar su aplicación cuando se usan armas autónomas. Como se ha apuntado, las principales características de las armas autónomas son su alta precisión, su alto poder computacional que permite un ágil funcionamiento y su autonomía. Estas armas operan a partir de los datos proporcionados, considerando múltiples variables y fuera del control humano directo por lo que actúan de forma neutral y objetiva, desapareciendo el error humano. Error éste que puede ser relevante en situaciones en las que las personas estén bajo presión, estrés, situaciones de vida o muerte, fatiga, etc. No obstante, la autonomía no implica ausencia de error, sino simplemente que las armas autónomas pueden estar sujetas a sus propios errores.

En principio, teniendo en cuenta dichas características, las armas autónomas favorecerían el cumplimiento efectivo de los principios de Derecho internacional humanitario, al tratarse de sistemas de armas objetivos, que toman en consideración todos los datos y las precauciones necesarias, más proclives a reducir las muertes de civiles, a minimizar los daños colaterales, etc., brindando, por ende, mejores resultados desde un punto de vista humanitario (Umbrello et al., 2020) (Costa, 2023). No obstante, frente a esta postura también es posible considerar que el empleo de armas autónomas puede ser contraproducente desde la perspectiva humanitaria. En primer lugar, fomenta la carrera armamentista entre los Estados, pero también entre actores no estatales como organizaciones y grupos criminales.

En segundo lugar, contribuye a la difusión de los conflictos armados, puesto que estos sistemas resultan más atractivos dado que no exigen el despliegue de soldados (humanos), disminuyen el riesgo de pérdidas humanas entre el ejército, etc., favoreciendo, en último término, su aceptación social y/o política. Los drones se presentan como una herramienta muy ventajosa ya que protegen las vidas propias a la par que consiguen exterminar la de los enemigos. De esta manera, la llegada de los drones o la creciente autonomía de las armas han provocado cambios trascendentales en la percepción de los conflictos armados a todos los niveles: micro (soldados individuales), meso (fuerzas armadas) y macro (la sociedad). A nivel macro se asiste a una desconexión o disociación entre la opinión pública y los conflictos armados. Estas transformaciones son capaces de invisibilizar los conflictos armados, de reducir los daños colaterales, haciéndolos más fáciles y atractivos en términos políticos y sociales (Bauman & Lyon, 2013, pp. 23 ss).

En tercer lugar, posibilitan el aumento del riesgo de escalada del conflicto, dada la rapidez de las operaciones al reducirse considerablemente el tiempo de análisis y respuesta de estos sistemas. Y, en cuarto lugar, incrementan significativamente riesgo de discriminación algorítmica. Los sistemas de inteligencia artificial reproducen o incluso amplifican sesgos presentes en los datos con los que fue entrenado (Van Severen & Vander Maelen, 2021), así como los sesgos derivados de una distribución desigual de las variables (Miró Llinares, 2018). Así, si el entrenamiento contiene sesgos, el arma autónoma puede identificar objetivos basándose en la raza, la etnia, el género, etc., pudiendo, por ejemplo, identificar hombres como objetivos más probables por estadística militar, aumentando el riesgo de ataques indiscriminados; confundir civiles con combatientes si su apariencia no coincide con los patrones con los que fue entrenada o si, por ejemplo, el arma fue entrenada en entornos urbanos occidentales, podría malinterpretar los comportamientos en zonas rurales de otros países, a saber, podría confundir una ceremonia local con una actividad hostil, etc.

La discriminación algorítmica implica que las armas autónomas en vez de presentar un comportamiento objetivo o neutral refuerzan los sesgos y los prejuicios existentes,

convirtiéndose en una amenaza para la población civil y un mayor desafío para la aplicación y el respeto del Derecho internacional humanitario.

En general, las normas del Derecho internacional humanitario pueden ofrecer una respuesta válida cuando estamos, por ejemplo, ante un ataque indiscriminado en el que el operador de forma intencional, dolosa, utiliza un arma, un sistema de inteligencia artificial, con el objetivo de matar población civil de forma indiscriminada, cometiendo claramente un crimen de guerra. Asimismo, se podría atribuir la responsabilidad penal a título de imprudencia si a pesar de existir razones para dudar de la capacidad de la máquina para diferenciar entre combatientes y civiles no se adoptaron las cautelas necesarias.

Los problemas de imputación surgen cuando estamos ante sistemas de armas autónomas. De una parte, esta última etapa de la evolución de las tecnologías aplicadas a los conflictos armados no viene acompañada de regulación. Estamos ante una situación de alegalidad con conocimiento. Los Estados, la comunidad internacional y los ciudadanos saben que las armas autónomas se están utilizando en los conflictos armados; pero existe falta de voluntad por parte de los Estados a comprometerse en la protección de los derechos humanos. De otra parte, los sistemas de armas autónomos tienen la capacidad de aprender y, por lo tanto, pueden modificar o ajustar su comportamiento durante la misión. La autonomía implica que el sistema analiza el entorno, evalúa opciones, selecciona y ejecuta la acción sin intervención humana. La máquina introduce un elemento decisional que altera la estructura de la acción humana (Pagallo, 2013). En cambio, en la automatización el resultado es consecuencia directa del programa. Por ejemplo, en los misiles guiados tradicionales el ser humano mantiene dominio completo del proceso causal y, por ende, se puede imputar la conducta a los operadores o programadores sin mayores dificultades dogmáticas.

La conducta del sistema inteligente (del arma autónoma) se vuelve impredecible, no sólo cuando se encuentra en una situación para la que no ha sido programado con una respuesta adecuada, sino también cuando decide de forma autónoma una determinada acción basándose en su experiencia. De este modo, el sistema inteligente intenta ejecutar la tarea asignada por los seres humanos, por los programadores, incluso cuando esto exige adecuarse al contexto. Esto

puede considerarse una "ventaja" en su desarrollo, pero también implica que los humanos no pueden anticipar ni controlar plenamente el comportamiento del sistema inteligente en estas situaciones "imprevistas". No se puede predecir si un arma autónoma malinterpretará una señal o un contexto basándose en su experiencia, tampoco cuándo lo hará, etc. Entonces, si ni el programador ni el operador tienen el suficiente control sobre el comportamiento de la máquina, no asumirán o no deberían asumir la responsabilidad por dichas conductas (Matthias, 2004, p. 177). Además, el propio entorno o la interrelación con otros factores pueden contribuir a dificultar o incluso a imposibilitar el control o la propia predicción del ser humano respecto al comportamiento de la máquina (Magro, 2014).

Estos continuos avances en robótica militar exigen un abordaje holístico e internacional en el que se tomen en consideración los beneficios y los riesgos de emplear las armas autónomas. En el ámbito penal, emerge un nuevo paradigma en torno a la atribución de la responsabilidad cuando un sistema inteligente provoca un resultado penalmente relevante: ¿quién debe responder? ¿quién actuó? ¿el que lo ordenó? ¿el programador? ¿el propio sistema inteligente? ¿nadie? ¿se deben asumir estos riesgos como, por ejemplo, los del tráfico rodado, en aras del desarrollo y los avances tecnológicos?

Estos interrogantes se deben abordar con urgencia. Los medios de comunicación registran las continuas afectaciones a la población civil en el marco de los recientes conflictos armados, difundiendo la percepción de que tales vulneraciones resultan impunes y favoreciendo, por ende, su continuidad. Este tratamiento jurídico (internacional) transmite la idea de que los crímenes de guerra pasan a integrar un Derecho penal simbólico, en cuanto que las conductas, *a priori* constitutivas de crímenes de guerra, no son castigadas. En este sentido, apuntaba Beccaria que la eficacia del Derecho penal o de las penas residía en su certeza. Así, la seguridad de que la pena será aplicada (más que su severidad o intensidad) es lo que tiene un verdadero efecto disuasorio sobre los delincuentes. Sólo si se tiene la certeza de que el delito será castigado, desalentará su comisión (Beccaria, 2015).

Lamentablemente, esta situación de impunidad lejos de cambiar parece que se va a acentuar como consecuencia de dos fenómenos: el desarrollo tecnológico y la plutofilia. Los constantes

avances tecnológicos derivados de la inteligencia artificial o la robótica militar se centran en la reducción del control humano directo o el uso de las armas cada vez más autónomas, que erosionan o difuminan la responsabilidad y la rendición de cuentas por las decisiones adoptadas (Sehrawat, 2017), dificultándose así claramente el proceso de atribución de responsabilidad penal y aumentando a la par las posibilidades de que los actos ilícitos o comportamientos delictivos realizados por las armas autónomas queden impunes debido a su independencia e impredecibilidad.

Igualmente, la presencia de rasgos de plutofilia en este contexto favorece la impunidad. Esencialmente, la impunidad de los crímenes de guerra se observa cuando los ejecutan actores poderosos. Este fenómeno pondría de manifiesto una doble injusticia: se protege a los poderosos, a los que concentran el poder (las grandes potencias, los miembros del Consejo de Seguridad, etc.) y se castiga selectivamente a los más débiles: los Estados pequeños, sin respaldo internacional, grupos rebeldes, etc.

En este contexto de plutofilia, las armas autónomas pueden convertirse en una herramienta que amplifique la impunidad, ya que contribuyen a difuminar la responsabilidad individual y aprovechan las estructuras de poder, las cuales ya favorecen de por sí a los más poderosos en el orden internacional. Además, conviene tener presente que las armas autónomas se concentran en manos de los actores más poderosos, que ante las violaciones del Derecho internacional humanitario intentarán escudarse en “fallos técnicos” de las máquinas para eludir la responsabilidad.

Asimismo, la plutofilia y la selectividad del Derecho penal existentes implican decantarse a favor de quien detenta el poder en detrimento de los más débiles, como es la población civil en los conflictos armados, que, de una parte, sufren las consecuencias de la violencia y de otra, contemplan cómo los responsables poderosos escapan de la sanción.

III. EL DERECHO PENAL ANTE LAS ARMAS AUTÓNOMAS: LA ATRIBUCIÓN DE RESPONSABILIDAD PENAL

Para Stephen Hawking, la inteligencia es la capacidad de adaptarse al cambio (*Professor Stephen Hawking*, 2016). El Derecho penal, al igual que la inteligencia, debe ser capaz de adaptarse. Es más, el Derecho penal en cuanto medio de control social no puede permanecer ajeno a los avances tecnológicos y a las transformaciones sociales que se producen, exigiéndose su adecuación. Estamos ante un proceso de adaptación del Derecho penal en un contexto poco favorable, en el que las reglas ya no preceden a la técnica, sino que van por detrás y luchan por posicionarse frente a conflictos siempre nuevos y siempre impredecibles. Son muchos los problemas sociales que plantea la sociedad global y que no encuentran una respuesta en las legislaciones nacionales ni en la normativa internacional. Así, frente a un Derecho penal "inamovible", en el que la teoría del delito es el puerto seguro donde uno se puede refugiar cuando el mar de la vida y el viento de los hechos sociales agitan nuevos problemas y anuncian tormenta (Mangione, 2024, p. 97), se le exige que cambie y que lo haga rápidamente, al ritmo de la tecnología, la cual en poco tiempo deja obsoletos los objetos o los avances, pero, además que lo haga otorgando certeza. Se demanda la presencia de un Derecho penal que dé certezas a la técnica, a los avances tecnológicos.

De este modo, ante la realidad del uso de armas autónomas, ¿"quién" es el responsable en caso de daños derivados de la conducta realizada por la máquina? ¿a "quién" se le debe atribuir la responsabilidad penal?

Obviamente, conviene tener presente que los problemas sólo surgirán cuando efectivamente estemos ante armas autónomas, porque si se trata de sistemas de inteligencia artificial que están bajo el control y supervisión humana y no tienen autonomía, el Derecho penal ofrece respuestas válidas. Se trata simplemente de una evolución de la ciberdelincuencia tradicional (Lledó Benito, 2022, p. 100).

La primera, y siendo la más razonable, se podría atribuir la responsabilidad al usuario o al operador por las acciones o daños derivados, dado que los entes artificiales se encontrarían bajo su control, siendo posible imputarlos tanto a título de dolo como de imprudencia, e igualmente,

cabría la posibilidad de que respondiera el operador por una autoría mediata, cuando el operador ejecuta el hecho delictivo por medio de otro (en este caso, la máquina inteligente), del que se sirve como instrumento. Evidentemente, el recurso a la estructura de la autoría mediata resulta necesario para quienes defienden el reconocimiento de la responsabilidad al sistema de inteligencia artificial.

No plantea ninguna dificultad la atribución de responsabilidad a título de dolo a quien usa dolosamente el sistema de inteligencia artificial para cometer un delito. Igualmente, se podrá atribuir responsabilidad por dolo eventual, cuando los resultados no sean deseados de forma principal, pero se encuentren necesariamente ligados. Situaciones que pueden ser muy habituales en el ámbito de los conflictos armados. Piénsese, por ejemplo, en un comandante que ordena desplegar un dron en una zona densamente poblada con combatientes y civiles. Tanto los ingenieros como el mando militar saben que el algoritmo del dron tiene errores de precisión del 30 %, especialmente para diferenciar civiles de combatientes, por ejemplo, confunde a personas con herramientas agrícolas con combatientes armados. El comandante, a pesar de conocer ese alto riesgo, autoriza el uso del dron, el dron identifica erróneamente a un grupo de civiles como combatientes y lanza un ataque, causando decenas de muertes.

Igualmente, también serán frecuentes las actuaciones a título de imprudencia. Por ejemplo, un Estado despliega un sistema de torretas en la frontera para detectar y neutralizar intrusos armados. El comandante responsable ordena activar el sistema en una zona donde se sabe que transita población civil refugiada que huye del conflicto. Sin embargo, el comandante no revisa los parámetros del software, que estaban configurados de forma muy sensible, de manera que cualquier objeto metálico era interpretado como un arma. El comandante confía en la tecnología, no verifica ni supervisa el sistema y la torreta abre fuego contra un grupo de refugiados que portaban objetos metálicos: cazuelas y cubiertos consigo y mueren varios civiles.

Igualmente, también existe la posibilidad de atribuir la responsabilidad al programador o fabricante del sistema de inteligencia artificial a título de dolo (cuando con el sistema de

inteligencia artificial creado pretende cometer un delito) o de imprudencia por vulnerar el deber objetivo de cuidado, como ya sucede con la responsabilidad por el producto defectuoso.

A este respecto, resultará esencial definir el umbral de riesgo permitido. En ausencia de una norma vinculante (tanto internacional como comunitaria) que regule los sistemas de inteligencia con fines exclusivamente militares, se podría utilizar como modelo el Reglamento (UE) 2024/1689, de Inteligencia Artificial que establece cuatro categorías de sistemas de inteligencia artificial atendiendo al nivel de riesgo: inaceptable (y, por lo tanto, prohibidos), alto, limitado y mínimo o sin riesgo, estableciéndose para cada nivel determinadas exigencias, obligaciones o cautelas para su uso, pudiendo incluir la supervisión humana desde un control indirecto a uno directo. La inobservancia de estas normas de cuidado determinará en buena medida la posibilidad de atribuir responsabilidad al operador, al programador o al mando[5].

En este punto, adquiere también una vital importancia el uso apropiado o adecuado del sistema de inteligencia artificial (Valls Prieto, 2022). De este modo, cuando el sujeto se desvía del uso adecuado y acepta, es consciente de dicho desvío en su actuación, responderá dolosamente. En cambio, cuando el sujeto no es consciente de la desviación serán posibles dos situaciones: imprudencia o caso fortuito, siendo preciso, por ende, valorar el grado de conocimiento, la probabilidad de producción del resultado o la aceptación del mismo.

Lógicamente, la presencia de un caso fortuito determinará la ausencia de responsabilidad penal. Así, cuando el daño producido por el sistema de inteligencia artificial no sea atribuible a título de dolo o imprudencia a su programador u operador, no serán responsables. Piénsese, por ejemplo, en un dron que está configurado únicamente para identificar y seguir objetivos, pero no para atacar. Su *software* ha sido probado rigurosamente y no existían antecedentes de errores. Sin embargo, durante una misión, una tormenta geomagnética solar altera de forma súbita las señales de GPS y los sensores del dron y esto provoca que el sistema interprete la interferencia como una amenaza, se descontrole y lance un proyectil contra un grupo de civiles, causando su muerte.

[5] No obstante, debe advertirse la posibilidad de que las normas de cuidado se puedan ampliar peligrosamente en virtud del principio de precaución.

De la exposición anterior se observa con claridad cómo el Derecho penal puede ofrecer una respuesta válida a los supuestos de daños derivados de sistemas de inteligencia artificial siempre y cuando éstos se encuentren bajo la supervisión o el control humano. Los problemas para imputar penalmente aparecen cuando nos encontramos ante sistemas de inteligencia artificial autónomos, con capacidad de aprendizaje autónomo (Busato, 2022, p. 358). Cuando estos sistemas producen daños para los bienes jurídico-penales surge lo que se denomina *responsibility gap*: una brecha de responsabilidad, ya que ninguna persona podría ser considerada claramente responsable por las acciones de un sistema autónomo, especialmente de sistemas de inteligencia artificial avanzados. Este concepto fue introducido por Matthias (2004), para quien son tres las condiciones que generan un *responsibility gap*. Primera, la delegación de decisiones relevantes a un sistema autónomo (dejan de estar bajo el control humano). Segunda, la imprevisibilidad del comportamiento del sistema. Esta imprevisibilidad no se debe a ignorancia o negligencia humana, sino a la propia esencia del aprendizaje, que genera modelos que se modifican constantemente. En tales casos, nadie tiene el control suficiente como para ser considerado responsable, ni siquiera a nivel moral. Y tercera, la ausencia de criterios satisfactorios para imputar la conducta a sus diseñadores, usuarios o supervisores. Por consiguiente, el desarrollo de sistemas de inteligencia artificial autónomos plantea la imposibilidad de imputar responsabilidad directa o mediata por los resultados lesivos generados por un agente técnico.

Los sistemas de inteligencia artificial autónomos desafían las concepciones tradicionales del Derecho penal. Así, cuando estamos ante resultados derivados de la actuación de sistemas de inteligencia artificial autónomos, ¿quién responde? Existen, en principio, dos posibilidades. De una parte, los propios sistemas de inteligencia artificial, y de otra, las personas físicas que las han usado, programado o fabricado. No obstante, ambos escenarios no están exentos de obstáculos, como se analizará a continuación.

1.- Atribución de la responsabilidad a los propios sistemas de inteligencia artificial autónomos

Esta propuesta propone tratar a los sistemas de inteligencia artificial como sujetos de imputación penal, reconociéndoles algún tipo de personalidad jurídica. En teoría, esta solución permitiría cerrar la brecha de responsabilidad cuando, de una parte, ningún ser humano tiene control efectivo o previsibilidad sobre la conducta del sistema y, de otra parte, el daño proviene de decisiones autónomas no atribuibles claramente a diseñadores u operadores. Sin embargo, esta opción plantea importantes desafíos de cara a su viabilidad conforme a los principios del Derecho penal actual como se analizará en los siguientes subapartados.

a) Las armas autónomas: ¿un sujeto jurídico responsable penalmente?

En primer lugar, es necesario plantearse si un algoritmo puede ser un sujeto jurídico responsable penalmente. En este sentido, conviene tener presente que la teoría del delito está construida sobre una noción antropomorfa de la responsabilidad que implica la capacidad de realizar acciones humanas (Danaher, 2016). Tradicionalmente, en virtud de la máxima de Derecho romano *societas delinquere non potest*, sólo podían ser sujetos activos las personas físicas[6]. No obstante, los continuos cambios económicos y sociales evidenciaron que las personas jurídicas (empresas, partidos políticos, sindicatos, etc.) no sólo pueden cometer delitos, sino que los cometen y, de hecho, en el ámbito económico, por ejemplo, se erigen como los principales sujetos activos.

Los sistemas de inteligencia artificial cometen "delitos" (Abbott, 2020), pueden llevar a cabo comportamientos tipificados como delitos, pero ¿puede un sistema de inteligencia artificial ser un sujeto jurídico responsable penalmente? Esta cuestión exige plantearse dónde se traza la línea divisoria entre máquina y persona física o humana. Tarea cada vez más compleja si se tienen en cuenta los continuos progresos en el ámbito de la biotecnología, la inteligencia

[6] Si bien en la Edad Media los animales y no los dueños respondían de sus actos dañinos. Lógicamente, con la llegada de la Ilustración y las teorías preventivas de la pena desaparece el Derecho penal contra los animales.

artificial o la nanotecnología que dificultan cada vez más establecer dónde comienza y dónde termina el cuerpo humano[7].

Igualmente, con carácter previo a la posible atribución de personalidad jurídica es necesario reflexionar, aunque sea de forma breve, sobre distintas cuestiones: ¿Qué se entiende por persona o entidad artificial? ¿Dónde se sitúa a este ente inteligente? ¿Cuál es su naturaleza jurídica? Aspectos todos ellos esenciales para el Derecho penal pero que no son de su competencia directa, sino de otras ramas del ordenamiento jurídico.

En primer lugar, conviene tener en cuenta que no existe unanimidad respecto al concepto de inteligencia artificial (Comité Económico y Social Europeo, 2017) (Hallevy, 2015) y como se ha indicado previamente, tampoco respecto al sentido y alcance de las armas autónomas.

El art. 3.1) del Reglamento (UE) 2024/1689, de inteligencia artificial, establece que se entenderá por sistema de inteligencia artificial:

> un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales.

Esta definición establece, de una parte, que cada sistema de inteligencia artificial tiene un objetivo determinado en su función la cual ha sido definida previamente por los seres humanos (Morillas Fernández, 2023), ya que "no son capaces de adivinar lo imprevisto ni de crear algo de la nada" (Valls Prieto, 2022, p. 10) y, de otra, que existen diferentes grados de autonomía. Siendo necesario, por tanto, abordar en segundo lugar, qué nivel de autonomía tienen los sistemas de inteligencia artificial. En este sentido, la inteligencia artificial se tiende a clasificar en dos categorías: fuerte y débil (Burchard, 2020); (Burgstaller et al., 2019) (de la Cuesta Aguado, 2019) (Fateh-Moghadam, 2019) (Gaede, 2019). Los sistemas de inteligencia artificial débil integrarían aquellos sistemas que aplican técnicas de inteligencia artificial a problemas concretos o específicos para los que se le ha asignado sin que tengan la posibilidad de actuar

[7] Sirva como ejemplo el caso de Neil Harbisson que se define a sí mismo como un *cyborg* y así está reconocido jurídicamente en el Reino Unido. Este reconocimiento radica en el hecho de tener implantada en la cabeza una antena que le permite escuchar y percibir los colores. Antena que se considera como una parte más de su cuerpo y no como un dispositivo electrónico.

por sí solos. En cambio, los sistemas de inteligencia artificial fuerte son capaces de operar fuera del perfil que se le ha fijado y, por ende, no existirían límites a sus posibilidades de uso y/o desarrollo, pudiendo realizar cualquier y múltiples tareas de forma que igualen o superen a la inteligencia humana, puesto que se autoprograman, son proactivos[8] y es en esta categoría, en principio, en la que se englobarían las armas autónomas, definidas en esta investigación como cualquier arma que pueda seleccionar y atacar objetivos sin intervención humana.

Por último, sería necesario plantearse la cuestión relativa a su naturaleza jurídica. Esto es, si es posible integrarla en alguna categoría existente: cosas, animales, personas físicas o personas jurídicas o bien, si es necesario crear una nueva categoría: entes artificiales o una personalidad jurídica electrónica, con derechos y obligaciones (capacidad jurídica) y con la capacidad para ejercer derechos y cumplir deberes (capacidad de obrar), al igual que se le atribuyen a las personas físicas y jurídicas (Hernández Giménez, 2019) (van den Hoven van Genderen, 2018), como preludio para decidir si lo son o no a efectos penales (Blanco Cordero, 2019).

Por lo común, si se admite que existen máquinas capaces de tomar decisiones autónomas similares a las que adoptan los humanos, deberían establecerse también consecuencias jurídicas (Benítez Ortúzar, 2020, p. 80). O en otras palabras, el siguiente paso lógico sería reconocer su personalidad jurídica y, de este modo, poderles atribuir la responsabilidad penal por las acciones llevadas a cabo (Pagallo & Quattrocolo, 2018).

Esta postura fue inicialmente defendida por la UE. En el Informe con recomendaciones a la Comisión sobre las normas de derecho civil en materia de robótica (2015/2103 INL) insta a la Comisión en el punto 59(f) a explorar:

> el establecimiento de un estatuto jurídico específico para los robots a largo plazo, de modo que al menos los robots autónomos más sofisticados puedan considerarse

[8] De hecho, el punto 59(f) del Informe con recomendaciones a la Comisión sobre las normas de derecho civil en materia de robótica (2015/2103 INL) insta a la Comisión a explorar «el establecimiento de un estatuto jurídico específico para los robots a largo plazo, de modo que al menos los robots autónomos más sofisticados puedan considerarse personas electrónicas responsables de indemnizar por cualquier daño que causen, así como la posibilidad de reconocer la personalidad electrónica de los robots que toman decisiones autónomas o interactúan de forma independiente con terceros». Sin embargo, en 2018 a través de la *Open Letter to the European Commission on Artificial Intelligence and Robotic*, más de 200 expertos se opusieron al considerar que la creación de un estatuto jurídico de «persona electrónica» permitiría a los productores eludir la responsabilidad por los daños causados.

personas electrónicas responsables de indemnizar por cualquier daño que causen, así como la posibilidad de reconocer la personalidad electrónica de los robots que toman decisiones autónomas o interactúan de forma independiente con terceros.

Posteriormente, mediante el Dictamen del Comité Económico y Social Europeo, aprobado el 31 de mayo de 2017, la UE expresó una posición totalmente opuesta. Pronunciándose en contra de otorgar cualquier tipo de personalidad jurídica a los sistemas de inteligencia artificial, sobre la base de que tal reconocimiento permitiría a los productores eludir la responsabilidad por los daños causados o supondría un riesgo moral inadmisible, ya que implicaría trasladar la responsabilidad desde los humanos hacia las máquinas, agravándose la irresponsabilidad o contribuyendo a una *abrogation of our own responsibility* (Bryson, 2010, p. 73).

b) La capacidad de acción

La atribución de la responsabilidad penal a un sistema inteligente encuentra en la teoría del delito diversos obstáculos. El primero radicaría en reconocer su capacidad de acción, si el sistema de inteligencia artificial actúa (es sujeto) o si se actúa sobre él (es un mero instrumento). Si un arma autónoma hiere o mata a civiles, en principio, podríamos indicar que estamos ante acciones que ha realizado ese sistema inteligente, que además actúa autónomamente, con independencia del control humano, puesto que es capaz de aprender y tomar decisiones con base en su experiencia, pero ¿realmente estamos ante una acción penalmente relevante? Resultará difícil si acudimos a la concepción de acción como acción final o a la social de Roxin, ya que no alcanza el valor ontológico (Lledó Benito, 2022, p. 88). No obstante, si se permanece en las teorías causalistas es mucho más factible. Desde el causalismo naturalista bastaría con constatar un movimiento corporal que provoque una modificación en el mundo exterior para considerar que estamos ante una acción penalmente relevante. La única precisión a tener en cuenta sería una restricción de los sistemas de inteligencia artificial. De este modo, para que un sistema de inteligencia artificial fuera capaz de una acción penalmente relevante sería necesario que presentara una forma corpórea y se moviera, desechándose aquellos sistemas incorpóreos o los consistentes en un mero *software*.

Para el causalismo valorativo, la acción humana debe ser voluntaria. Esto es, cualquier movimiento corporal voluntario podría ser considerado como una acción penalmente relevante.

Esto exigiría, de una parte, una interpretación amplia de la expresión "acción humana", como ya se hace en los supuestos de las acciones realizadas por las personas jurídicas y no circunscribirla exclusivamente a la acción llevada a cabo por las personas físicas; y, de otra parte, que el ente inteligente actúe voluntariamente, con el deseo de causar un resultado en el exterior. Considerando, que los sistemas de inteligencia artificial autónomos planifican los medios para alcanzar un objetivo, por ejemplo, un dron selecciona el trayecto más eficiente para alcanzar un objetivo. Se podría afirmar su carácter voluntario. Esta postura centra la atención en la autonomía de los sistemas de inteligencia, en su capacidad de aprendizaje, en su capacidad de modificar los comportamientos de acuerdo con la experiencia, factores externos, etc.; ya que los entes artificiales autónomos toman decisiones sin ninguna intervención humana (posterior a su activación). Decisiones que pueden ser impredecibles para su programador y su operador.

Asimismo, existe también la posibilidad de negar la relevancia de la acción penal cuando concurren determinadas causas: movimientos reflejos, estados de inconsciencia o fuerza irresistible. En el caso de sistemas de inteligencia artificial sería posible encontrarse con estas causas de exclusión de la acción penal. En primer lugar, el movimiento reflejo se podría equiparar a una respuesta automática programada, sin decisión autónoma real. Por ejemplo, un dron programado para estabilizarse dispara accidentalmente al detectar un cambio brusco en su sistema de sensores. En segundo lugar, los actos de inconsciencia podrían asimilarse a un fallo del sistema, un apagón que lleve al sistema de inteligencia artificial a ejecutar acciones sin coherencia. Y, en tercer lugar, la fuerza irresistible podría asemejarse con una interferencia externa irresistible, por ejemplo, el sistema sufre un hackeo y el que toma el control lo utiliza para atacar a un grupo de civiles. En principio, nada impediría trasladar estos supuestos de exclusión de la acción penal cuando estemos ante un sistema de inteligencia artificial.

Desde esta perspectiva, conforme a las teorías causalistas se podría afirmar su capacidad de acción. Las dificultades aparecerían cuando se recurre al concepto final de acción que pivota en torno a la intencionalidad de la acción humana. Las acciones tienen que ser finales, orientadas a un fin, controladas. Exige, por tanto, conciencia y voluntad. La acción es fruto de

la voluntad intencional del sujeto activo y sólo si ésta se configura como expresión de la voluntad humana puede ser considerada penalmente relevante. No se trataría simplemente de realizar un movimiento corporal, sino un acto orientado a un fin. Evidentemente, un arma autónoma puede identificar y atacar determinados objetivos, pero toma la decisión basándose en su programación o, realmente, transmite su propia evaluación, es decir, emite un juicio. Si se afirma que toma la decisión a partir de unos parámetros determinados por la programación, entonces la finalidad es atribuida por el programador y no por el ente inteligente. Considerando, de esta forma, que la inteligencia artificial no puede reproducir la actividad de una conciencia críticamente autocontrolada (Moro, 2014) (Aires de Sousa, 2020, p. 69).

Si, por el contrario, se sostiene que el ente artificial emite un juicio, se defiende la postura de que el ente artificial tiene voluntad (Hallevy, 2015, pp. 93 y ss), una autoconciencia, pero aparece una nueva dificultad: demostrarlo. Tarea que será tremendamente complicada de realizar si se tiene presente que su funcionamiento es opaco y resulta prácticamente imposible precisar cómo los sistemas de inteligencia artificial llegan a una conclusión o decisión (Van Severen & Vander Maelen, 2021) y, por consiguiente, también si se trata de un juicio propio o simplemente de la voluntad del programador.

En resumen, si se pretende afirmar la capacidad de acción de un ente inteligente bastaría con considerar desde una perspectiva funcional o utilitarista que presenta una intencionalidad artificial. Ahora bien, si se asume que los sistemas inteligentes pueden llevar a cabo acciones penalmente relevantes y seguimos avanzando en las categorías de la teoría del delito, aparece un nuevo escollo: la culpabilidad.

c) La culpabilidad de los entes artificiales

En la actualidad, se considera que los sistemas inteligentes carecen de autodeterminación moral, de libre albedrío. En principio y de forma general, se considera que los sistemas de inteligencia, los entes artificiales o inteligentes carecen de los elementos esenciales de la personalidad. Si bien es cierto que pueden aprender y tomar decisiones (ya que, gozan de autonomía), no son conscientes de su propia libertad, carecen de una conciencia jurídica o de la capacidad de ofrecer una respuesta emocional o ética (Lledó Benito, 2022, p. 57). Hoy por hoy no cumplen

con los requisitos mínimos de culpabilidad, pues carecen de conciencia de sí mismos, de una voluntad libre y consciente, así como de la capacidad para comprender la ilicitud de su conducta y de los medios necesarios para ajustarla a la ley (Blanco Cordero, 2019, p. 78).

No obstante, este impedimento no es óbice para que puedan ser consideradas dentro de una nueva categoría (entes artificiales o inteligentes), similar a la de las personas jurídicas y, por ende, poderles atribuir la responsabilidad penal. Es más, no tiene sentido permitir la atribución de la responsabilidad penal a las personas jurídicas y negarla a ciertas máquinas o sistemas inteligentes, ya que ambas entidades carecen del libre albedrío moral (Abbott, 2020) (Hallevy, 2010) (Hu, 2018) (Magro, 2014), aunque, como precisa Posada Maya, no estamos ante un problema jurídico sino filosófico (Posada Maya, 2019). De hecho, las personas jurídicas no tienen un pensamiento y voluntad como el ser humano, pero ello no es óbice para que se construyan conceptos como el de la voluntad de las personas jurídicas y otros similares que permitan superar el obstáculo de la culpabilidad. Es más, el Derecho penal puede crear ficciones si lo necesita y así lo ha hecho. Piénsese, por ejemplo, en el concepto de "hombre medio" usado en culpabilidad (Quintero Olivares, 2017). Simplemente bastaría con abandonar la concepción de la responsabilidad antropocentrista o antropomorfa y estructurar responsabilidades penales isomórficas (Posada Maya, 2019).

De este modo, con base a la capacidad autónoma de las redes neuronales de los robots y la inteligencia artificial sería posible atribuir responsabilidad penal a los sistemas de inteligencia artificial (Hallevy, 2010, p. 191) por los daños derivados de la conducta ejecutada por un sistema de inteligencia artificial. Así, si estamos ante un sistema de inteligencia artificial que ofrece respuestas autónomas e imprevisibles (incluso para el propio programador o fabricante), parece excesivo hacer responder al productor o al operador (Piergallini, 2020)(del Rosal Blasco, 2023). La solución más adecuada debería ser la de atribuir directamente responsabilidad a la máquina (Pagallo & Quattrocolo, 2018). La clave reside, por tanto, en su nivel de autonomía a partir de la cual primeramente se debería reconocer su personalidad jurídica, una personalidad electrónica, para así posteriormente poderles atribuir la responsabilidad penal por las acciones ejecutadas por los entes inteligentes autónomos.

Para Hallevy si se verifican los presupuestos de la responsabilidad penal en una entidad, ésta debe responder, con independencia de que sea una persona física, jurídica o una entidad artificial. Defiende, por tanto, una concepción claramente utilitarista de la responsabilidad penal, que exigiría simplemente verificar en la actuación de la inteligencia artificial los elementos externos (*actus reus*) y mentales (*mens rea*) requeridos por la responsabilidad penal (Hallevy, 2010).

Una vez expuestos los problemas relativos a la admisión de la culpabilidad, es necesario abordar los interrogantes relativos a la punibilidad: ¿se puede castigar a un sistema de inteligencia? ¿cuál sería la sanción aplicable? ¿la misma sanción que a las personas físicas o que a las personas jurídicas? ¿sanciones específicas? ¿cómo se cumplirían?

d) La punibilidad de los entes artificiales

Tradicionalmente, la pena se centraba en los seres humanos y en concreto en aquellos imputables. No obstante, en las últimas décadas los ordenamientos jurídicos han ido dando cabida a las sanciones a las personas jurídicas, esencialmente de contenido económico dirigidas a castigar económicamente a la persona jurídica, pero también a los socios o accionistas de la empresa. Asimismo, sigue todavía presente una cierta reticencia a la expansión del término pena para hacer referencia a cualquier sanción penal. No en vano nuestro CP utiliza la expresión medidas aplicables a las personas jurídicas.

En principio, tomando como referencia las sanciones previstas para las personas jurídicas, se podrían contemplar sanciones de naturaleza económica contra el ente artificial. Sin embargo, existe un importante inconveniente a tener en cuenta: cómo se puede sancionar económicamente a un ente si éste no tiene reconocidos derechos y obligaciones y, por ende, no tiene bienes o activos con los que poder hacer frente a la multa. Sería necesario, por tanto, el reconocimiento previo de una personalidad jurídica, del estatus de personalidad, denominémosla electrónica o inteligente en los sistemas de inteligencia artificial, concretando exactamente cuáles son los bienes jurídicos o derechos de los que son titulares para determinar así su restricción o privación (Blanco Cordero, 2019, p. 79).

Junto a la posible sanción económica, también se podrían considerar otras sanciones similares a las previstas en nuestro Código penal en el art. 33.7 CP para las personas jurídicas. Así, frente a la disolución de la persona jurídica se podrían prever como sanciones equivalentes: la desactivación definitiva o la destrucción física con la pérdida de la personalidad jurídica, electrónica atribuida, siendo posible una destrucción del *hardware* o del *software*. Igualmente, podrían ser adecuadas sanciones como la reprogramación -garantizando su "resocialización"-, la prohibición de realizar en el futuro las actividades relacionadas con el delito cometido, la imposición de sistemas de trazabilidad o el sometimiento a la supervisión y control humano directo de todas o parte de sus actividades, limitando o incluso eliminando su autonomía. A pesar del alarde creativo en la previsión de las sanciones aplicables a los entes inteligentes existe un escollo insalvable: la valoración de todas estas sanciones no sería la misma que la efectuada por una persona física o incluso jurídica. Decididamente la motivación cambia, difiere la percepción del castigo, puesto que parece difícil pensar que los entes artificiales están dotados de una voluntad de supervivencia o de una libre determinación de la personalidad que podrían verse afectadas con la amenaza de la imposición de las correspondientes sanciones y que actuarían como factores preventivos de la comisión de delitos (Lledó Benito, 2022, p. 88) (del Rosal Blasco, 2023, p. 18). No obstante, podría ser suficiente con adaptar las penas para que tuvieran una función y un sentido (Hallevy, 2015, pp. 185 y ss). Asimismo, sancionar directamente a la inteligencia artificial puede tener efectos preventivos en cuanto que la destrucción del sistema de inteligencia artificial, a modo de *robot death penalty*, privaría a los desarrolladores, propietarios y usuarios de los beneficios del sistema que de otro modo obtendrían, incentivándolos así a modificar su comportamiento hacia modos socialmente deseables (Abbott, 2020).

Igualmente, autores como Mulligan consideran que castigar a los sistemas de inteligencia artificial cumple un fin muy concreto: la venganza, generando una satisfacción psicológica en las víctimas (como se cita en Blanco Cordero, 2019, p. 78).

2.- Atribución de la responsabilidad penal a personas físicas por los daños provocados por sistemas de inteligencia artificial autónomos

Este modelo constituye la alternativa dominante para evitar el *responsability gap*. Así, evitar la impunidad requiere insistir en que la responsabilidad última siempre recaiga en los seres humanos que construyen, son dueños u operan las máquinas, y que el daño causado por una herramienta es culpa del operador, programador o fabricante. Esta propuesta parte de la premisa de que el Derecho penal se dirige a los seres humanos y no a las máquinas por muy inteligentes que sean. Se defiende, por tanto, que la responsabilidad siga recayendo en seres humanos (usuarios, programadores, fabricantes), incluso cuando la actuación del sistema de inteligencia artificial sea impredecible. Esta postura se basa en dos aspectos. De una parte, que todo sistema de inteligencia artificial es un artefacto creado por el ser humano[9] y, por lo tanto, siempre existe algún grado de responsabilidad por su diseño, entrenamiento, implementación o uso. Y, de otra parte, incluso si el comportamiento concreto es impredecible, puede exigirse responsabilidad ampliando la noción de imprudencia sobre la base de deberes reforzados de diligencia en programación, pruebas y control; protocolos de seguridad o evaluación de riesgos antes de su uso; o bien fijando una responsabilidad por riesgos tecnológicos, esto es, si alguien decide utilizar o poner en circulación un sistema con potencial de causar daños, asume la carga de responder por ellos, aunque no pudiera prever exactamente cómo se producirían.

En general, se podría atribuir la responsabilidad a cuatro "cargos" o roles: usuario u operador, programador, fabricante y mando o superior jerárquico, siempre que, como indica el art. 5 CP, estemos ante una acción u omisión dolosa o imprudente penada por la ley.

En primer lugar, se puede atribuir responsabilidad penal cuando se trata de una acción u omisión voluntaria dolosa. La persona quiere el resultado o acepta conscientemente que ocurrirá. En otras palabras, usa intencionadamente el sistema autónomo como "instrumento" para cometer el delito. Por ejemplo, un operador dirige intencionadamente el arma autónoma hacia un grupo

[9] Aunque conviene tener presente que los propios algoritmos de aprendizaje automático son capaces de diseñar otros artefactos. Los seres humanos ya no son los únicos que pueden diseñar o crear artefactos (Lledó Benito, 2022, p. 72).

de civiles, sabiendo que el sistema disparará, por lo que el operador responderá por un homicidio doloso en autoría o autoría mediata.

En segundo lugar, si existe imprudencia grave en el diseño, fabricación, despliegue o uso. La persona no quiere el resultado, pero vulnera de forma evidente y grave deberes de cuidado, profesionales o de seguridad, generando un riesgo relevante y previsible. Piénsese, por ejemplo, en el fabricante que omite un test crítico de reconocimiento de objetivos para ahorrar tiempo, pese a saber que el modelo confunde objetivos civiles con militares, pero confiando en que los daños no se producirán. Tras el despliegue se produce un ataque sobre un objetivo civil, por lo que se podría considerar que estamos ante un homicidio imprudente imputable a los responsables técnicos.

En principio, atribuir la responsabilidad a las personas físicas se presenta cómo la vía más coherente con el Derecho penal actual. Sin embargo, igualmente, existen importantes problemas dogmáticos para poderles atribuir la responsabilidad cuando los sistemas de inteligencia artificial son autónomos como se esboza a continuación.

a) La acción penalmente relevante y el problema del dominio

Welzel concibió la acción como un comportamiento voluntario finalista, es decir, orientado hacia un fin mediante el conocimiento del curso causal. Sin embargo, los sistemas de inteligencia artificial autónomos estiran esta concepción, porque el operador no conoce el razonamiento interno del sistema, no puede prever de forma fiable cómo actuará el sistema, y tampoco puede corregir la actuación una vez iniciado el proceso autónomo. Por consiguiente, la conexión entre la voluntad humana y el resultado final se debilita y, en principio, no podría existir acción humana que abarque el resultado producido por la máquina.

Asimismo, si se considera el concepto social de acción de Roxin, que establece que la acción es una conducta significativa en el mundo exterior, que es dominada o dominable por la voluntad (Roxin, 2008, p. 194), también surgen dificultades. El problema radica en que, en las armas autónomas, de una parte, existe una interrupción estructural entre el acto humano (diseñar, activar el sistema) y el resultado (la selección autónoma del blanco), y de otra, la conducta de la máquina no es socialmente significativa en términos de voluntad humana. Es

cierto que Roxin admite que la acción puede comprender procesos mediatos complejos, pero exige siempre dominio del proceso. Y claramente, ante las armas autónomas, ese dominio desaparece o se reduce a niveles insuficientes para fundamentar la autoría.

Una posible solución sería reinterpretar la acción humana como control sobre los riesgos y la supervisión, no sobre cada resultado. Como se ha indicado, la autonomía de las armas autónomas rompe la configuración causal clásica entre acción humana y resultado lesivo. En las armas autónomas, el operador, el programador o el mando no determinan de forma directa y continua la conducta final del sistema, sino que solo establecen parámetros iniciales cuya ejecución se desarrolla mediante procesos algorítmicos no completamente previsibles. Esta discontinuidad causal impide fundamentar la imputación penal en la tradicional exigencia de dominio del hecho o capacidad de evitar el resultado concreto, dado que puede que nadie ejerza un control efectivo en el momento en el que se produce el daño.

Ante esta brecha, la categoría de acción podría reinterpretarse como ejercicio de deberes de creación, limitación o supervisión de los riesgos. De este modo, la imputación penal se reconduce a la infracción de deberes de diligencia ligados al diseño, entrenamiento, autorización o despliegue del sistema en las que lo jurídicamente relevante es la gestión del riesgo y no el control material del resultado. Esta "objetivación" de la acción permitiría evitar brechas de responsabilidad derivadas de la autonomía, y al mismo tiempo mantendría la coherencia del sistema penal al exigir a los intervinientes que aseguren, mediante medidas de diseño y supervisión, que el funcionamiento del sistema no sobrepase los límites del Derecho Internacional Humanitario y la protección de bienes jurídicos.

b) Acción sin dominio: la autoría mediata frente a algoritmos autónomos

Según Roxin, el autor mediato domina el hecho a través de un instrumento humano que actúa sin capacidad de autodeterminación (error, coacción, etc.). La idea principal es el dominio del hecho, es decir, el autor controla el curso causal mediante su poder sobre el instrumento (Roxin, 2014, pp. 84 y ss). Pero ¿puede un arma autónoma ser un instrumento en el sentido formulado por Roxin?

En principio, las armas autónomas plantean dos problemas. El primer problema radica en que el instrumento no es un instrumento pasivo, sino un sistema con autonomía. Y el segundo hace referencia a que no existe dominio del hecho, porque la conducta se genera mediante procesos internos inaccesibles y no determinados por la voluntad del operador, programador, etc., motivados por el *machine learning* o incluso el *deep learning*.

Es más, incluso si se acepta una analogía con el instrumento automatizado, el arma autónoma presenta variabilidad conductual que rompe la estructura de dominio exigida por la autoría mediata (Lledó Benito, 2022, p. 96).

Igualmente, se podría flexibilizar esta teoría y admitir que, aunque las armas autónomas no son instrumentos en sentido clásico, lo son desde una interpretación funcional del dominio del hecho, reinterpretando el dominio del hecho sobre riesgos, no sobre cada resultado concreto. En otras palabras, el operador, el mando o el diseñador controlan indirectamente el riesgo jurídico global, aunque no el resultado específico y se considera autor mediato a quien crea, activa o mantiene un sistema autónomo sin controles mínimos, generando un riesgo no permitido.

c) Autoría mediata por dominio de la organización tecnológica

Esta concepción implica un cambio de paradigma en la imputación penal: no se busca atribuir responsabilidad por un resultado específico concreto, sino establecer mecanismos para asegurar la prevención de riesgos inherentes a sistemas tecnológicos autónomos que operan con alta complejidad y autonomía.

De este modo, se posibilita sostener una responsabilidad penal colectiva o institucional, compatible con los principios del Derecho penal actual, que reconoce la importancia de la estructura y organización en la comisión de hechos delictivos. Esta responsabilidad colectiva fortalece la gobernanza y el control, al enfatizar la obligación de las organizaciones y entidades públicas de implementar sistemas de *compliance*, evaluación continua y rendición de cuentas.

Esta propuesta se basa en el dominio de la organización (*Organisationsherrschaft*) de Roxin, es decir, en el supuesto de un control efectivo y previsible de la organización sobre los medios y la conducta que produce el resultado delictivo (Roxin, 2014, pp. 111 y ss). Sin embargo, las

armas autónomas introducen una complejidad inédita. Se trata de sistemas que pueden mostrar comportamientos imprevisibles, adaptativos, con variabilidad conductual y opacidad algorítmica, lo que limita la posibilidad de un control absoluto y lineal.

Para solventar las posibles dificultades, en el contexto de las armas autónomas, el dominio de la organización debería entenderse como un dominio sobre la estructura organizacional que habilita y controla el sistema, aunque el comportamiento específico del sistema sea en parte autónomo y no totalmente previsible. Este dominio implica la responsabilidad sobre el diseño, la supervisión y la gestión de riesgos, más que sobre cada decisión puntual del sistema.

En conclusión, la autoría por organización tecnológica representa un instrumento clave para abordar la responsabilidad penal en contextos donde la acción humana directa es difusa o limitada, permitiendo responder jurídicamente por los riesgos emergentes de la inteligencia artificial autónoma, de las armas autónomas, bajo un enfoque preventivo y estructural.

d) Gestión corporativa de riesgos y *compliance*

El enfoque más reciente para prevenir la impunidad, particularmente en el ámbito corporativo, se centra en exigir a las empresas la gestión activa y rigurosa de los riesgos generados por la inteligencia artificial. El Departamento de Justicia de los Estados Unidos, a través de sus *Guidelines* actualizadas (*Evaluation of Corporate Compliance Programs*, ECCP) de septiembre de 2024, ha convertido la gestión del riesgo de la inteligencia artificial en un factor clave para evaluar los programas de *compliance* de las compañías.

Este sistema busca evitar la impunidad asegurando que las empresas tengan mecanismos internos para detectar y corregir desviaciones, manteniendo la responsabilidad en la persona jurídica o en los individuos dentro de la corporación (Gómez-Jara Díez et al., 2024).

e) La omisión impropia en contextos algorítmicos

La figura de la omisión impropia o comisión por omisión adquiere especial relevancia como criterio para atribuir responsabilidad a aquellas personas físicas que, teniendo un deber jurídico específico de intervención, permiten que el daño se produzca mediante su inactividad.

Desde una perspectiva dogmática, la omisión impropia aparece cuando el sujeto, aun sin realizar materialmente la conducta típica, omite una acción jurídicamente exigible que habría

evitado el resultado, de modo que la inacción se equipara normativamente a la causación activa del hecho. Para poder exigir responsabilidad penal son necesarios tres requisitos. El primero consiste en la existencia de una posición de garante, derivada de una ley, de un contrato, de la asunción voluntaria de una función protectora o del dominio sobre una fuente de riesgo; el segundo se basa en la posibilidad real y concreta de evitar el resultado mediante la acción omitida; y el tercero reside en que la omisión debe ser la causa del resultado, es decir, que el resultado se produzca precisamente por no actuar. En teoría, la omisión impropia permitiría atribuir responsabilidad penal cuando un arma autónoma causa un daño si las personas físicas responsables del diseño, supervisión, operación o despliegue no realizan acciones exigibles para evitar resultados lesivos previsibles y evitables.

Pero, de nuevo, en la práctica, en el caso de las armas autónomas, el deber puede existir, puede estar contemplado en una norma, pero el control fáctico se reduce drásticamente y, además, la capacidad real de evitar el resultado puede no existir o ser técnicamente imposible cuando estamos ante armas autónomas (*man-off-the-loop*).

Además, habría que probar el dolo o la imprudencia en el momento en el que diseñó, programó, activó, etc. el arma autónoma y no va a ser fácil, debido a la opacidad de su funcionamiento, su autonomía o ausencia del control humano significativo sobre la decisión final, la dificultad para establecer un nexo causal entre la omisión y el resultado final o porque el momento del diseño sucede normalmente años antes del uso bélico.

f) La responsabilidad por el producto

La responsabilidad por el producto es un régimen jurídico según el cual el fabricante, distribuidor o vendedor puede ser considerado responsable penal cuando un producto que ha puesto en el mercado presenta un defecto (de diseño, fabricación o información) y causa daños a personas o bienes. Su lógica central es que, si alguien obtiene un beneficio al introducir un producto en el mercado, también debe asumir los riesgos derivados de que ese producto sea defectuoso.

Aunque el marco de responsabilidad por el producto es útil, las armas autónomas presentan dificultades particulares (del Rosal Blasco, 2023). En primer lugar, la trazabilidad del daño

resulta especialmente compleja. Estos sistemas suelen operar mediante algoritmos opacos y procesos de toma de decisiones difícilmente reconstruibles, lo que impide identificar con precisión en qué punto se produjo el fallo: si en el diseño del *software*, en los sensores, en la integración del sistema o en la interacción con el entorno. Esta falta de transparencia complica seriamente atribuir el defecto a un responsable concreto.

A ello se suma la intervención de múltiples actores en el ciclo del arma autónoma. El fabricante del *hardware*, los desarrolladores del *software*, las empresas que integran los componentes, los operadores militares y quienes establecen las reglas de despliegue, todos ellos contribuyen al funcionamiento final del sistema. La coexistencia de tantos niveles de responsabilidad diluye la posibilidad de identificar un único sujeto responsable, como suele hacerse en los casos clásicos de responsabilidad por productos defectuosos.

Otro desafío proviene de la capacidad de los sistemas autónomos para actualizarse o aprender tras su puesta en circulación. Cuando un arma modifica su comportamiento con base en nuevas experiencias o datos, puede producir daños derivados de patrones no previstos por el fabricante en el momento de su comercialización. Esta evolución posterior dificulta determinar si el defecto que originó el daño estaba presente desde el diseño inicial o surgió como consecuencia del proceso de aprendizaje del sistema, lo que tensa la estructura jurídica tradicional, basada en un producto estático.

Finalmente, el contexto impredecible del combate introduce un grado adicional de incertidumbre. Los escenarios en los que operan las armas autónomas pueden variar de forma abrupta y contener innumerables variables no anticipadas durante el diseño y la prueba del producto. Esta volatilidad puede influir en el desempeño del sistema y generar resultados que no responden únicamente a defectos del producto, sino a condiciones extremas del entorno. Distinguir entre un daño atribuible al arma y uno derivado de la situación táctica se convierte así en una tarea difícil o incluso imposible, lo que complica aún más la aplicación del régimen de responsabilidad por el producto.

g) Recapitulando: a modo de balance general

El desarrollo y utilización de sistemas de inteligencia artificial plantea un importante desafío jurídico en torno a la delimitación del riesgo permitido, especialmente en sectores en los que la actuación de la máquina puede tener consecuencias graves o irreversibles. En este sentido, conviene apuntar que no existe una regulación sobre las armas autónomas en las que se establezca su marco regulador o la gestión de los riesgos como efectúa, por ejemplo, el Reglamento (UE) 2024/1689, sobre inteligencia artificial, en un ámbito general, excluyendo los sistemas usados exclusivamente con fines militares, de defensa o de seguridad nacional.

Tradicionalmente, el Derecho penal admite ciertos márgenes de riesgo cuando la actividad realizada es socialmente útil, se ejerce conforme a los estándares técnicos vigentes y se aplican medidas adecuadas de control. Este aspecto se revela hoy imprescindible para comprender el espacio de actuación de la inteligencia artificial, en tanto que estos sistemas participan de procesos decisionales con repercusiones directas sobre bienes jurídicos fundamentales.

La atribución de responsabilidad penal derivada de daños causados por sistemas de inteligencia artificial autónomos depende de una distinción central: si el resultado era previsible *ex ante* o si, por el contrario, resultaba científicamente imposible de prever**.**

Cuando la desviación del sistema o el daño constituye una posibilidad conocida y previsible *ex ante*, aunque con un grado variable de probabilidad, la imputación recae en el ámbito de la responsabilidad subjetiva, pudiendo adoptar la forma de imprudencia o, en determinados casos, de dolo eventual. De este modo, si la persona física (el programador, el usuario, el mando) infringe una norma de cuidado, pero confía en que el resultado posible no se producirá, estamos ante un supuesto de imprudencia consciente. Aquí opera el criterio del riesgo permitido, que establece cuándo la conducta supera el umbral del riesgo permitido, de lo admisible socialmente. En cambio, cuando el agente conoce la posibilidad del resultado y, aun así, continúa, actúa, aceptando su eventual producción, la imputación será a título de dolo eventual, puesto que asume el riesgo de que el daño se concrete. En ambos supuestos la clave es la previsibilidad del resultado y la actitud subjetiva de la persona física frente al riesgo que introduce.

Evidentemente, la situación es distinta cuando el resultado lesivo es científicamente imprevisible debido a la falta de conocimiento suficiente sobre el funcionamiento o las posibles desviaciones del sistema. En situaciones de incertidumbre científica, donde el resultado no es cognoscible o es imposible prever científicamente, la responsabilidad penal queda excluida. Si el daño deriva de una desviación causada por factores totalmente inesperados, estaremos ante un acontecimiento fortuito. No siendo admisible la incorporación del principio de precaución al ámbito del Derecho penal para invocar en ausencia de previsibilidad, la existencia de responsabilidad penal (Quintero Olivares, 2025, pp. 38-40) (del Rosal Blasco, 2023, pp. 14-15).

En la era de la sociedad del riesgo, ahora especialmente tecnológico, se asiste a una importante transformación y el Derecho penal se orienta cada vez más hacia los riesgos con el objetivo de anticipar la protección de los bienes jurídicos y, como no, otorgar la ansiada seguridad. En este sentido, surge el dilema de cómo el Derecho penal puede reforzar los deberes de cuidado.

En el ámbito de los sistemas de inteligencia artificial autónomos podemos encontrarnos claramente con supuestos en los que se actúa sin el cuidado debido, se infringe una norma que sí se podía cumplir, pero resulta difícil demostrar que se produce el resultado como consecuencia de esa vulneración, y que éste era previsible. Para sortear los problemas de causalidad y previsibilidad que exige la imputación, el Derecho penal acude a la técnica de tipificación de los delitos de peligro, especialmente de peligro abstracto o a una responsabilidad objetiva, vulnerando claramente los principios y garantías propios del Derecho penal.

Así, resulta habitual brindar la protección penal exigiendo simplemente la presencia de un peligro real (delito de peligro concreto), la idoneidad de la conducta para poner en peligro los bienes jurídicos (delitos de peligro abstracto-concreto) o presumiendo la peligrosidad de la conducta (delitos de peligro abstracto). En los delitos de peligro, lo que importa es si la conducta crea un riesgo objetivamente relevante (o simplemente se presume, en los delitos de peligro abstracto), pero carece de importancia si el autor era capaz de preverlo o lo imaginó. Situaciones en las que claramente se adelanta la intervención del Derecho penal.

Igualmente, en aras a esa pretendida eficacia, el Derecho penal lamentablemente acude a la responsabilidad objetiva, es decir, la imputación del resultado lesivo se efectúa al margen de un juicio de culpabilidad, prescindiendo de la constatación de dolo o imprudencia y tomando como base exclusivamente la mera producción del resultado o la infracción formal de un deber. Este modelo de atribución es incompatible con un Derecho penal del hecho, pues desplaza el análisis desde la conducta subjetivamente reprochable hacia consecuencias que pueden exceder el ámbito de dominio del autor.

Así, desde una perspectiva garantista se aprecia con claridad cómo la expansión del ámbito de actuación del Derecho penal provoca el quebrantamiento de sus principios legitimadores. En primer lugar, se produce la vulneración del principio de culpabilidad, erigido en límite infranqueable del poder punitivo, al permitir sanciones sin un reproche personal fundado. Igualmente, se produce una afectación de los principios de legalidad y de proporcionalidad, al ampliar el ámbito de punición más allá del dolo y la imprudencia; así como de la presunción de inocencia, facilitando imputaciones que presumen indebidamente la responsabilidad por el solo acaecimiento del resultado (o su presunción *iure et de iure*). En resumen, conlleva una desnaturalización de la función de garantía del Derecho penal, que exige vincular la pena a un hecho típico, antijurídico y culpable, evitando la punición por riesgos no asumidos o resultados no dominables.

Dadas las importantes dificultades que plantean los delitos de peligro abstracto o la responsabilidad objetiva, quizás lo más adecuado sería trasladar el problema fuera del Derecho penal, hacia otras ramas del ordenamiento jurídico como el Derecho administrativo (del Rosal Blasco, 2023, p. 44); ya que así se evita distorsionar o manipular la teoría del delito, permite estándares dinámicos y facilita la prevención y la represión sin exigir culpabilidad, sin verificar la existencia de dolo o imprudencia.

En definitiva, el panorama actual pone de manifiesto que no existe una solución única, coherente y plenamente satisfactoria para afrontar el *responsibility gap* que generan los sistemas de inteligencia artificial autónomos. El Derecho penal no puede imputar penalmente al sistema de inteligencia artificial porque no es sujeto, carece de personalidad jurídica, pero tampoco

puede imputar sin más a personas físicas por los resultados derivados de las armas autónomas sin flexibilizar o funcionalizar el Derecho penal.

Las alternativas analizadas ofrecen respuestas parciales, cada una con sus pros y sus contras impidiendo su plena recepción dogmática. Todas ellas presentan algún tipo de tensión conceptual con los principios estructurales del Derecho penal, autoría, acción, culpabilidad, previsibilidad y responsabilidad personal, de modo que su implementación exigiría optar por renuncias o ajustes teóricos cuya justificación excede el marco de una aproximación preliminar de este artículo.

Este artículo no pretende ofrecer soluciones definitivas, sino delimitar el problema y mostrar la complejidad de su tratamiento dogmático, destacando que la respuesta adecuada no podrá ser monolítica ni inmediata. Al contrario, deberá surgir de un proceso de elaboración teórica más detenida, que evalúe con rigor las diversas alternativas y sus implicaciones político-criminales, sin olvidar que la adaptación del Derecho penal a los retos de la inteligencia artificial debe realizarse sin perder de vista sus fundamentos garantistas.

No obstante, si lo que se pretende es conseguir eliminar la brecha de responsabilidad existente, será necesario corregir la situación de alegalidad existente, promover la existencia de una regulación de las armas autónomas que fije el umbral del riesgo permitido y exigir la supervisión humana para garantizar la atribución de responsabilidad penal.

El desarrollo y uso de los sistemas de armas autónomas plantea un importante reto ético y jurídico. Si bien su avance se justifica frecuentemente por criterios de eficacia militar o reducción de riesgos para tropas propias, la ausencia de un control humano significativo tensa los límites del riesgo permitido. Como plantea Quintero Olivares (Quintero Olivares, 2025, p. 34), ¿se puede ceder el poder de decidir sobre la vida a una máquina? La autonomía introduce un grado de imprevisibilidad incompatible con el estándar mínimo de seguridad exigible en actividades que afectan a la vida, la integridad física o la paz internacional. La clave reside en que la delegación de decisiones letales a algoritmos no solo puede incrementar la probabilidad de daños no previstos, sino que dificulta radicalmente la atribución de responsabilidad, rompiendo el fundamento mismo de la imputación penal. Desde esta perspectiva, la supervisión

humana significativa no puede entenderse como un simple requisito técnico, sino también como un presupuesto normativo indispensable para mantener la actividad dentro de los márgenes del riesgo permitido. Solo la presencia de personas físicas con capacidad real y efectiva de supervisar, intervenir y, en su caso, detener el funcionamiento del sistema permite afirmar que la actuación del arma "autónoma" es jurídicamente controlable. La supervisión humana garantiza que las decisiones letales no se desvinculen de la esfera de responsabilidad que corresponde a las personas físicas y a las instituciones que los despliegan, ya que siempre existe un sujeto que puede ser considerado responsable. Asegura que cada decisión (letal) pueda ser rastreada hasta un operador o mando concreto, fortaleciendo la posibilidad de determinar la existencia de dolo o imprudencia. Asimismo, al mantener supervisión humana significativa, se disminuye la probabilidad de resultados completamente inesperados, facilitando la imputación de responsabilidad a individuos o estructuras organizativas.

V. REFLEXIONES FINALES

Con la irrupción de la inteligencia artificial se asiste a un verdadero cambio de paradigma, un nuevo mundo que hasta hace poco se consideraba utópico, propio de la ciencia ficción, pero que ahora forma parte de nuestra vida.

La inteligencia artificial aplicada al campo militar es una realidad incuestionable con un alto impacto y, por lo tanto, es necesario conocer y analizar sus consecuencias sociales, políticas, económicas, éticas y jurídicas, así como fijar una regulación en la que se garantice el respeto a los derechos humanos y a las disposiciones del Derecho internacional humanitario.

Obviamente, la aparición y empleo de armas autónomas implica desafíos adicionales, puesto que el grado de autonomía en las funciones de los sistemas de armas es proporcional al nivel de control humano que se ejerce sobre ellos. Por consiguiente, a mayor autonomía en las funciones, menor será el control humano, aumentando, por ende, el nivel de sustitución o reemplazo del ser humano en la realización de las operaciones por parte de las máquinas y, por consiguiente, resultará más complejo atribuir la responsabilidad penal por las acciones de las máquinas, de las armas autónomas. Así, la aparición de las armas autónomas evidencia una

brecha de responsabilidad. La solución más fácil para superarla sería afirmar el principio del control humano significativo. Para ello, sería necesario. Primero, una moratoria inmediata sobre el desarrollo y despliegue de estos sistemas. Y segundo, la negociación de un tratado internacional que prohíba completamente las armas autónomas, tal como se hizo con las minas antipersonales. La exigencia de una supervisión o control humano significativo de los sistemas de inteligencia artificial aseguraría la presencia de un ser humano responsable conforme a los principios y categorías del Derecho penal actual.

Sin embargo, los Estados no están interesados en limitarse, máxime cuando éstos son los que poseen las armas autónomas y el poder para salir impunes por los crímenes de guerra cometidos. En este contexto, el Derecho penal deberá afrontar importantes retos.

En primer lugar, deberá dar respuesta a los riesgos sociales y daños cada vez más frecuentes como consecuencia de las innovaciones tecnológicas producidas. Esto es, se debe encontrar un equilibrio entre Ciencia y Derecho, pero desde una perspectiva internacional e interdisciplinar. Este abordaje resulta imprescindible no sólo por las características de la sociedad actual, sino también por el campo de estudio que nos ocupa: las armas autónomas y su uso en conflictos armados (internacionales). Los daños provocados exigen una intervención penal, una respuesta dirigida a evitar la impunidad en el marco de los conflictos armados. Ámbito en el que las continuas violaciones del Derecho internacional humanitario y la ejecución de las conductas tipificadas como crímenes de guerra parecen integrar un Derecho penal simbólico, el cual crea la ilusión de protección de los civiles o de sus bienes, pero que no sanciona a los responsables de la vulneración de los bienes jurídicos. Esta impunidad existente en la actualidad se puede magnificar ante la irrupción del uso de las armas autónomas en los conflictos armados, dado que su autonomía e impredecibilidad difuminan y dificultan la atribución de la responsabilidad penal.

Y, en segundo lugar, el Derecho penal debe valorar si las categorías dogmáticas existentes sirven, es decir, si la formulación actual de la teoría del delito continúa siendo válida para atribuir responsabilidad penal a los entes artificiales, o si es necesaria la formulación de un nuevo concepto de delito para facilitar la atribución de responsabilidad penal en este contexto.

Resulta esencial impedir que los avances tecnológicos decidan la impunidad por las consecuencias derivadas del uso de las armas autónomas. La lucha contra la impunidad ayuda a evitar la reincidencia, pero también a combatir comportamientos plutofílicos que desprotegen a las víctimas, la población civil, pobre y desamparada que permanece en la zona del conflicto y benefician claramente a los Estados poderosos, que raramente vienen sancionados.

BIBLIOGRAFÍA

Abbott, R. (2020). *The reasonable robot: Artificial Intelligence and the Law*. Cambridge University Press.

Aires de Sousa, S. (2020). «Não fui eu, foi a máquina»: Teoría do crime, responsabilidade e inteligência artificial. En A. Miranda Rodrigues (Ed.), *Inteligência artificial no direito penal* (pp. 59-93). Almedina.

Ambos, K. (2015). Responsabilidad penal en el ciberespacio. *Indret*, *2*, 1-33.

Areola García, A. (2022). Armas Autónomas Letales—Medios de Defensa y Ataque. *Revista Seguridad y Poder Terrestre*, *1*(2), 177-186.

Battistelli, F. (2023a). Armi non umane. Miti, sogni e incubi dell'autonomia delle armi. En F. Farruggia (Ed.), *Dai droni alle armi autonome* (pp. 21-42). FrancoAngeli.

Battistelli, F. (2023b). Introduzione. En F. Farruggia (Ed.), *Dai droni alle armi autonome* (pp. 17-20). FrancoAngeli.

Bauman, Z. (2007). *Liquid Times. Living in an Age of Uncertainty*. Polity Press.

Bauman, Z., & Lyon, D. (2013). *Liquid Surveillance*. Polity Press.

Beccaria, C. (2015). *Tratado de los delitos y las penas*. Universidad Carlos III de Madrid.

Benítez Ortúzar, I. F. (2020). Reflexiones sobre robótica y Derecho. Especial referencia al vehículo autónomo. *Foro galego. Revista xurídica xeral de Galicia*, *208*, 75-96.

Bertieri, S., & Iaria, A. (2023). Il diritto internazionale umanitario e la sfida delle armi autonome all'intus-legere. En F. Farruggia (Ed.), *Dai droni alle armi autonome* (pp. 142-159). FrancoAngeli.

Blanco Cordero, I. (2019). Homo sapiens y ¿machina sapiens?: Un Derecho penal para los robots dotados de inteligencia artificial. En C. Mallada Fernández (Ed.), *Nuevos retos de la ciberseguridad en un contexto cambiante* (pp. 63-80). Thomson Reuters Aranzadi.

Bryson, J. J. (2010). Robots should be slaves. En Y. Wilks (Ed.), *Close Engagements with Artificial Companions* (pp. 63-74). John Benjamins.

Burchard, C. (2020). Artificial intelligence as the end of criminal law? On the algorithmic transformation of society. En A. Miranda Rodrigues (Ed.), *Inteligência artificial no direito penal* (pp. 59-93). Almedina.

Burgstaller, P., Hermann, E., & Lampesberger, H. (2019). *Künstliche Intelligenz: Rechtliches und technisches Grundwissen*. Manz.

Busato, P. C. (2022). De máquinas y seres vivos: ¿Quién actúa en los resultados derivados de decisiones cibernéticas? En E. Demetrio Crespo (Ed.), *Derecho penal y comportamiento humano. Avances desde la neurociencia y la inteligencia artificial* (pp. 349-376). Tirant lo Blanch.

Comité Económico y Social Europeo. (2017). *Dictamen C-288 sobre la "inteligencia artificial: Las consecuencias de la inteligencia artificial para el mercado único (digital), la producción, el consumo, el empleo y la sociedad", de 31 de mayo y 1 de junio de 2017*.

Comité Internacional de la Cruz Roja. (2021). *Posición del CICR sobre los sistemas de armas autónomos*. https://www.icrc.org/sites/default/files/document_new/file_list/4550_003-ebook.pdf

Costa, M. J. (2023). Sistemas de armas autónomas e respectiva regulamentação. En S. Aires de Sousa (Ed.), *A proposta de Regulamento Europeu sobre Inteligência Artificial: Algumas questões jurídicas* (pp. 65-78). Instituto Jurídico Faculdade de Direito da Universidade de Coimbra.

Danaher, J. (2016). Robots, Law and the Retribution Gap. *Ethics and Information Technology*, 299-309.

de la Cuesta Aguado, P. M. (2019). Inteligencia artificial y responsabilidad penal. *Revista Penal México*, *16-17*, 51-62.

del Rosal Blasco, B. (2023). ¿El modelo de responsabilidad penal de las personas jurídicas para los daños punibles derivados del uso de la inteligencia artificial? *Revista de Responsabilidad Penal de las Personas Jurídicas y Compliance*, *2*, 1-49.

Fateh-Moghadam, B. (2019). Innovationsverantwortung im Strafrecht: Zwischen strict liability, Fahrlässigkeit und erlaubtem Risiko – Zugleich ein Beitrag zur Digitalisierung des Strafrechts. *ZStW*, *131*, 863-887.

Gaede, K. (2019). *Künstliche Intelligenz – Rechte und Strafen für Roboter?* Nomos.

Gigova, R. (2017, septiembre 2). *Who Putin thinks will rule the world*. CNN. https://www.cnn.com/2017/09/01/world/putin-artificial-intelligence-will-rule-world

Gómez-Jara Díez, C., Feijoo Sánchez, B., & Tejada Plana, D. (2024). La irrupción de la inteligencia artificial en elas guidelines del compliance del Departamento de Justicia Americano. *Revista de Responsabilidad Penal de las Personas Jurídicas y Compliance*, *5*, 1-6.

Hallevy, G. (2010). The criminal liability of artificial intelligence entities – From science fiction to legal social control. *Akron Intellectual Property Journal*, *4*(2), 171-201.

Hallevy, G. (2015). *Liability for crimes involving artificial intelligence systems*. Springer.

Hernández Giménez, M. (2019). Inteligencia artificial y Derecho penal. *Actualidad Jurídica Iberoamericana*, *10 bis*, 792-843.

Hu, Y. (2018). Hu, Ying, Robot Criminal Liability Revisited. En J. S. Yoon, S. H. Han, & S. J. Ahn (Eds.), *Dangerous Ideas in Law* (pp. 494-509). Bohmunsa.

Kurzweil, R. (2015). *La Singularidad está cerca: Cuando los humanos transcendamos la biología*. Lola Books.

Liñán Lafuente, A. (2017). Crímenes de guerra. *Eunomía. Revista en Cultrua de la Legalidad*, 264-272.

Lledó Benito, I. (2022). *El Derecho Penal, Robots, IA y Cibercriminalidad: Desafíos éticos y jurídicos. ¿Hacia una distopía?* Dykinson.

Magro, M. B. (2014). Biorobotica, robotica e diritto penale. En D. Provolo, S. Riondato, & F. Yenisey (Eds.), *Genetics, robotic, law, punishment* (pp. 499-516). Padova University Press.

Mangione, A. (2024). *Intelligenza artificiale, attività d'impresa e diritto penale. La «funzione di garanzia» nell'organizzazione e dell'organizzazione per la «sorveglianza dell'AI»*. Giappichelli.

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics Inf Technol*, *6*, 175-183. https://doi.org/10.1007/s10676-004-3422-1

Meza Rivas, M. J. (2022). *Las armas autónomas letales: Un desafío para el Derecho internacional humanitario, los derechos huanos, laa seguridad y el desarme internacionales*. Tirant lo Blanch.

Miró Llinares, F. (2018). Inteligencia artificial y justicia penal: Más allá de los resultados lesivos causados por robots. *Revista de Derecho Penal y Criminología*, *20*, 87-130.

Morello, R. (2022). Crímenes de guerra. En *Derecho Penal Internacional* (pp. 183-262). Thomson Reuters Aranzadi.

Morillas Fernández, D. L. (2023). Implicaciones de la inteligencia artificial en el ámbito del Derecho Penal. En *Derecho penal, inteligencia artificial y neurociencias* (pp. 59-91). Roma Tre-Press.

Moro, P. (2014). Biorobotica e diritti fondamentali. Problemi e limiti dell'intelligenza artificiale. En D. Provolo, S. Riondato, & F. Yenisey (Eds.), *Genetics, robotic, law, punishment* (pp. 533-544). Padova University Press.

Olasolo Alonso, H. (2015). *Introducción al Derecho Internacional Penal*. Tirant lo Blanch.

ONU. (2019). *CCW/MSP/20119/9, Reunión de las Altas Partes Contratantes en la Convención sobre prohibiciones o resptricciones del empleo de ciertas armas convencionales que puedan considerarse excesivamente nocivas o de efectos indiscriminados*. https://docs.un.org/es/CCW/MSP/2019/9

Pagallo, U. (2013). *The Laws of Robots: Crimes, Contracts, and Torts*. Springer.

Pagallo, U., & Quattrocolo, S. (2018). The impact of AI on criminal law, and its twofold procedures. En W. Barfield & U. Pagallo (Eds.), *Research Handbook on the Law of Artificial Intelligence* (pp. 385-409). Edward Elgar.

Parisi, G. (2023). Prefazione. En F. Farruggia (Ed.), *Dai droni alle armi autonome* (pp. 13-15). FrancoAngeli.

Pérez González, M. (2017). El Derecho internacional humanitario frente a la violencia bélica: Una apuesta por la humanidad en situaciones de conflicto. En *Derecho Internacional Humanitario* (3ª ed., pp. 25-52). Tirant lo Blanch.

Piergallini, C. (2020). *Intelligenza artificiale: Da «mezzo» a «autore» del reato?* https://u-pad.unimc.it/handle/11393/282100

Posada Maya, R. (2019). La responsabilidad penal de los agentes de inteligencia artificial: Entre la ficción y una realidad que se aproxima. En *Un juez para la democracia. Libro Homenaje a Perfecto Andrés Ibáñez* (pp. 561-581). Dykinson.

Professor Stephen Hawking: 13 of his most inspirational quotes. (2016, enero 8). The Telegraph. https://www.telegraph.co.uk/news/science/stephen-hawking/12088816/Professor-Stephen-Hawking-13-of-his-most-inspirational-quotes.html

Quintero Olivares, G. (2017). La robótica ante el Derecho penal: El vacío de la respuesta jurídica a las desviaciones incontroladas. *Revista Electrónica de Estudios Penales y de la Seguridad*, *1*, 1-23.

Quintero Olivares, G. (2025). Los límites de la aportación de la Inteligencia Artificial al derecho penal. En M. L. Noya Ferreiro & M. Á. Catalina Benavente (Eds.), *La inteligencia artificial y su aplicación en el sistema de justicia penal* (pp. 21-46). Aranzadi.

Roxin, C. (2008). *Derecho Penal. Parte General. Tomo I. Fundamentos. La estructura de la Teoría del Delito*. Thomson Reuters Civitas.

Roxin, C. (2014). *Derecho Penal. Parte General. Tomo II. Especiales formas de aparición del delito*. Thomson Reuters Civitas.

Russell, S. J., & Norvig, P. (2004). *Inteligencia artificial: Un enfoque moderno*. Prentice - Hall Hispanoamericana.

Scharre, P., & Horowitz, M. (2015). *An Introduction to Autonomy in Weapon Systems*. Center for a New American Security, CNAS. https://www.cnas.org/publications/reports/an-introduction-to-autonomy-in-weapon-systems

Sehrawat, V. (2017). Autonomous weapon system: Law of armed conflict (LOAC) and other legal challenges. *Computer Law & Security Review*, *33*(1), 38-56. https://doi.org/10.1016/j.clsr.2016.11.001

Shilo, L. (2018). When Turing met Groitus AI, Indeterminism, and Responsibility. *SSRN*. https://doi.org/Shilo,%2520Liron,%2520When%2520Turing%2520Met%2520Grotius%2520AI,%2520Indeterminism,%2520and%2520Responsibility%2520(April%25209,%25202018).%2520Available%2520http://dx.doi.org/10.2139/ssrn.3280393

Siroli, G. P. (2023). Vulnerabilità delle tecnologie informatiche, Intelligenza Artificiale e LAWS. En F. Farruggia (Ed.), *Dai droni alle armi autonome* (pp. 61-75). FrancoAngeli.

Umbrello, S., Torres, P., & De Bellis, A. F. (2020). The future of war: Could lethal autonomous weapons make conflict more ethical? *AI & Society*, *35*, 273-282.

UN. (2025). *CCW/GGE.1/2025/WP.1, Report Convention on prohibitions or restrictions on the use of certain conventional weapons which may be deemed to be excessively injurious or to have indiscriminate effects*. https://meetings.unoda.org/ccw/convention-on-certain-conventional-weapons-group-of-governmental-experts-on-lethal-autonomous-weapons-systems-2025

UN. Panel of Experts Established pursuant to Security Council Resolution 1973 (2011) (Ed.). (8). *Letter dated 8 March 2021 from the Panel of Experts on Libya Established pursuant to Resolution 1973 (2011) addressed to the President of the Security Council*. UN. https://digitallibrary.un.org/record/3905159

Valls Prieto, J. (2022). Sobre la responsabilidad penal por la utilización de sistemas inteligentes. *Revista Electrónica de Ciencia Penal y Criminología*, *24-27*, 1-35.

van den Hoven van Genderen, R. (2018). Legal personhood in the age of artificially intelligent robots. En W. Barfield & U. Pagallo (Eds.), *Research Handbook on the Law of Artificial Intelligence* (pp. 213-250). Edward Elgar.

Van Severen, S., & Vander Maelen, C. (2021). Killer Robots: Lethal Autonomous Weapons and International Law. En J. De Bruyne & C. Vanleenhove (Eds.), *Artificial Intelligence and the Law* (pp. 151-172). Intersentia.