Cuadernos de RES PUBLICA en derecho y criminología
Nº en prensa ISSN: 2990-0697 DOI: 10.46661/respublica.13605.

Del riesgo algorítmico al daño irreversible: inteligencia artificial de alto riesgo, posición de garante y conflictos armados

From Algorithmic Risk to Irreversible Harm: High-Risk Artificial Intelligence, Guarantor Position and Armed Conflicts

Diana Matar Khalil
Fundación Universitaria Juan de Castellanos. Boyacá (Colombia)
Dmatar@jdc.edu.co
ORCID: 0009-0003-3017-7608

Resumen

El presente artículo analiza los desafíos jurídicos derivados del empleo de sistemas de inteligencia artificial de alto riesgo en contextos de conflicto armado, a partir de la teoría jurídica del riesgo, la posición de garante, el Derecho Internacional Humanitario y el Derecho Internacional de los Derechos Humanos. El estudio sostiene que la identificación de riesgos algorítmicos previsibles genera deberes jurídicos de prevención orientados a evitar su transformación en daños incompatibles con los principios de humanidad, distinción, proporcionalidad y precaución. Asimismo, se analiza la relevancia de la dignidad humana, el control humano significativo y la posición de garante como categorías jurídicas aplicables a tecnologías emergentes. Se concluye que los marcos normativos vigentes proporcionan herramientas suficientes para abordar los desafíos derivados de la inteligencia artificial en los conflictos armados y que los Estados tienen la responsabilidad de adoptar medidas regulatorias orientadas a prevenir riesgos y proteger los derechos fundamentales.

Palabras clave: Inteligencia artificial; riesgo algorítmico; conflictos armados; Derecho Internacional Humanitario; posición de garante.

Abstract

This article examines the legal challenges arising from the use of high-risk artificial intelligence systems in armed conflict scenarios through the lens of legal risk theory, the guarantor position doctrine, International Humanitarian Law and International Human Rights Law. The article argues that the identification of foreseeable algorithmic risks generates legal duties of prevention aimed at avoiding their transformation into harms incompatible with the principles of humanity, distinction, proportionality and precaution. It further explores the relevance of human dignity, meaningful human control and the guarantor position as legal categories applicable to emerging technologies. The study concludes that existing legal frameworks provide sufficient tools to address the challenges posed by artificial intelligence in armed conflicts and that States have a responsibility to adopt regulatory measures aimed at preventing risks and protecting fundamental rights..

Keywords: Artificial intelligence; algorithmic risk; armed conflicts; International Humanitarian Law; guarantor position.

Cómo citar este trabajo: *Matar Khalil*, Diana. (2026). Del riesgo algorítmico al daño irreversible: inteligencia artificial de alto riesgo, posición de garante y conflictos armados *Cuadernos de RES PUBLICA en derecho y criminología, (5), 01–27.* https://doi.org/10.46661/respublica.13605

Recepción: 10.06.2026 **Aceptación:** 17.06.2026 **Publicación:** en prensa

1. Introducción

La guerra ha sido históricamente una actividad asociada a la incertidumbre, al riesgo y a la posibilidad permanente de pérdida. Desde los planteamientos clásicos de Carl von Clausewitz sobre la niebla y la fricción de la guerra hasta los desarrollos contemporáneos del Derecho Internacional Humanitario, la comunidad internacional ha reconocido que los conflictos armados constituyen escenarios en los que la protección absoluta frente al daño resulta imposible. Sin embargo, el reconocimiento de la existencia del riesgo no implica su aceptación irrestricta ni exonera a los Estados del deber jurídico de prevenir, mitigar y controlar aquellos daños previsibles que puedan afectar a la población civil y a las personas protegidas.

En la actualidad, la incorporación de sistemas de inteligencia artificial en operaciones militares plantea nuevos desafíos para el derecho. Estas tecnologías prometen optimizar la toma de decisiones, mejorar la precisión operacional y reducir determinados niveles de incertidumbre en el campo de batalla. No obstante, también generan riesgos inéditos derivados de la opacidad algorítmica, los sesgos de programación, las deficiencias en los procesos de entrenamiento de datos y la creciente autonomía de ciertos sistemas utilizados en contextos de conflicto armado.

La discusión jurídica contemporánea se ha concentrado principalmente en aspectos éticos, tecnológicos y regulatorios de la inteligencia artificial. Sin embargo, son todavía limitados los estudios que abordan esta problemática desde la teoría jurídica del riesgo y desde los deberes de prevención que recaen sobre los Estados, las organizaciones militares y los demás actores involucrados en el diseño, desarrollo, validación y utilización de sistemas de inteligencia artificial aplicados a operaciones militares.

El presente artículo parte de la premisa según la cual la guerra nunca ha sido una actividad libre de riesgos y que el Derecho Internacional Humanitario, lejos de desconocer esta realidad, ha construido un conjunto de normas orientadas precisamente a limitar los efectos de la violencia armada y reducir los daños previsibles derivados de las hostilidades. Bajo esta perspectiva, se sostiene que los sistemas de inteligencia artificial de alto riesgo empleados en conflictos armados generan riesgos jurídicos, operacionales y humanitarios cuya identificación activa deberes de prevención y mitigación incluso antes de la materialización del daño. Si bien, se ha demostrado su efectivad y destaca su apoyo operacional, los riesgos siempre son latentes y se deben prever.

Se argumenta, además, que el riesgo algorítmico no surge exclusivamente durante la ejecución de una operación militar, sino desde las fases previas de diseño, programación, entrenamiento, validación y despliegue de los sistemas. En consecuencia, la ausencia de un régimen internacional específico no impide que los Estados adopten medidas regulatorias orientadas a la identificación, administración y reducción de tales riesgos en cumplimiento de las obligaciones derivadas del Derecho Internacional Humanitario, los Derechos Humanos y los principios generales de prevención.

La circunstancia de que gran parte del marco normativo aquí estudiado haya sido formulado con anterioridad al desarrollo de la inteligencia artificial no implica su obsolescencia frente a los desafíos tecnológicos actuales. Las transformaciones en los medios y métodos empleados en los conflictos armados no alteran la vigencia de principios jurídicos fundamentales como la humanidad, la distinción, la proporcionalidad, la precaución, la prevención del daño y la responsabilidad. Antes bien, tales principios proporcionan criterios interpretativos indispensables para evaluar las implicaciones jurídicas derivadas del uso de sistemas de inteligencia artificial de alto riesgo en escenarios contemporáneos.

2. El riesgo como categoría jurídica

La noción de riesgo ha adquirido una importancia creciente dentro de los sistemas jurídicos contemporáneos como consecuencia del desarrollo tecnológico, la complejidad social y la aparición de actividades capaces de generar afectaciones masivas a bienes jurídicos individuales y colectivos. A diferencia de los modelos tradicionales de responsabilidad construidos alrededor del daño consumado, las tendencias actuales del derecho han evolucionado hacia esquemas preventivos orientados a la identificación, administración y mitigación de riesgos antes de la materialización efectiva del perjuicio.

Desde la perspectiva sociológica, Ulrich Beck sostiene que las sociedades contemporáneas han transitado hacia una "sociedad del riesgo", caracterizada por la producción permanente de amenazas derivadas del propio desarrollo científico y tecnológico. Según este autor, la modernidad ya no se enfrenta exclusivamente a peligros naturales, sino a riesgos manufacturados por la actividad humana, cuya complejidad dificulta su identificación y control (Beck, 1998). En similar sentido, Anthony Giddens advierte que el avance tecnológico genera nuevas formas de incertidumbre que exigen mecanismos institucionales capaces de gestionar riesgos cada vez más sofisticados (Giddens, 1999).

En el ámbito jurídico colombiano, el riesgo ha sido entendido como una posibilidad abstracta y aleatoria de daño. Esta aproximación permite diferenciarlo de otras categorías cercanas como la amenaza, el peligro y el daño. El riesgo se caracteriza por la existencia de una probabilidad de afectación futura sobre determinados bienes jurídicos, sin que exista todavía una manifestación concreta e inmediata de lesión.

Al respecto, la jurisprudencia de la Corte Constitucional, ha establecido que en el riesgo "existe una posibilidad abstracta y aleatoria de que el daño a la vida o a la integridad personal se produzca. Este nivel se divide en dos categorías: a) riesgo mínimo: categoría hipotética en la que la persona sólo se ve amenazada por la muerte y la enfermedad naturales y; b) riesgo ordinario: se refiere a aquel riesgo que proviene tanto de factores internos como externos a la persona y que se deriva de la convivencia en sociedad. En este nivel de la escala, los ciudadanos deben soportar los riesgos que son inherentes a la existencia humana y a la vida en sociedad. (Sentencia T – 399 de 2018).

La amenaza constituye un estadío diferente. Puede entenderse como la existencia de hechos reales y verificables que generan peligro para la integridad, la libertad o los derechos fundamentales de una persona. Mientras el riesgo se mueve en el terreno de la probabilidad abstracta, la amenaza supone la existencia de circunstancias concretas que permiten prever una afectación específica. De allí que toda amenaza implique la existencia de un riesgo, aunque no todo riesgo alcance la entidad suficiente para constituirse en amenaza. Siguiendo el criterio jurisprudencial mencionado, en la amenaza "existen hechos reales que, de por sí, implican la alteración del uso pacífico del derecho a la tranquilidad y que hacen suponer que la integridad o la libertad de la persona corren verdadero peligro. En efecto, la amenaza de daño conlleva el inicio de la alteración y la merma del goce pacífico de los derechos fundamentales, debido al miedo razonable que produce visualizar el inicio de la destrucción definitiva del derecho. Por eso, a partir de este nivel, el riesgo se convierte en amenaza. Dependiendo de su intensidad, este nivel se divide en dos categorías: ordinaria y extrema".

Por su parte, el peligro representa una situación más próxima a la producción del daño. Se manifiesta como una condición latente o potencial de afectación física, económica, social o jurídica. Finalmente, el daño corresponde a la materialización efectiva de la lesión sobre los bienes jurídicos protegidos por el ordenamiento.

Así las cosas, la jurisprudencia constitucional colombiana ha reconocido la necesidad de diferenciar los diversos niveles de riesgo a los que pueden estar sometidas las personas. En términos generales, se ha distinguido entre riesgos ordinarios, que corresponden a las contingencias normales derivadas de la vida en sociedad, y riesgos extraordinarios o excepcionales, caracterizados por una intensidad superior capaz de comprometer de manera significativa derechos fundamentales. Esta clasificación reviste especial interés para el análisis propuesto porque determina la intensidad de los deberes de protección que recaen sobre el Estado.

Esta diferenciación conceptual posee especial relevancia para el análisis de tecnologías emergentes. La regulación jurídica no puede esperar necesariamente a la ocurrencia del daño para intervenir. Por el contrario, la identificación de riesgos previsibles permite activar mecanismos preventivos orientados a evitar que situaciones de riesgo evolucionen hacia escenarios de amenaza, peligro y posterior daño.

La existencia de riesgos extraordinarios genera obligaciones positivas de actuación derivadas de la posición de garante que corresponde a las autoridades públicas. En tales eventos, el Estado no se limita a abstenerse de causar daños, sino que asume deberes concretos de prevención, protección y mitigación orientados a reducir las probabilidades de afectación de los derechos fundamentales.

La noción de posición de garante permite comprender por qué la identificación de riesgos previsibles genera deberes jurídicos concretos de actuación. En el derecho colombiano, esta figura ha sido desarrollada ampliamente por la jurisprudencia penal y constitucional como fundamento de aquellas situaciones en las cuales una persona se encuentra jurídicamente obligada a impedir la producción de un resultado lesivo cuando posee la capacidad real de evitarlo.

La Sala de Casación Penal de la Corte Suprema de Justicia ha señalado que la posición de garante corresponde a la situación en la que una persona tiene el deber jurídico específico de actuar para impedir la ocurrencia de un resultado típico evitable, de manera que la omisión de dicha actuación puede generar responsabilidad cuando el resultado finalmente se produce (Corte Suprema de Justicia, Sala de Casación Penal, Rad. 25.536, 2006).

Este entendimiento encuentra fundamento normativo inmediato en el artículo 25 de la Ley 599 de 2000, disposición que regula las hipótesis de comisión por omisión y la posición de garante dentro del ordenamiento penal colombiano. No obstante, la construcción jurisprudencial de esta figura también ha sido vinculada a principios constitucionales orientados a la protección efectiva de los bienes jurídicos fundamentales, entre ellos el principio de solidaridad previsto en el artículo 95 de la Constitución Política, conforme al cual toda persona debe actuar con responsabilidad frente a situaciones que comprometan la vida o la salud de otros individuos.

A su vez, la Corte Constitucional ha precisado que la posición de garante surge en aquellos eventos en los cuales una persona posee una obligación constitucional o legal de intervenir para proteger un bien jurídico y, pese a tener la capacidad de hacerlo, permanece inactiva (Sentencia SU-1184 de 2001). De esta forma, la responsabilidad jurídica no se limita a las conductas activas que generan daño, sino que puede extenderse a la omisión de medidas necesarias para evitar riesgos previsibles.

Esta aproximación resulta especialmente útil para el análisis de los sistemas de inteligencia artificial de alto riesgo. Si los riesgos asociados a una determinada tecnología pueden ser identificados desde las etapas de diseño, programación, entrenamiento, validación o despliegue, los distintos actores que participan en su desarrollo no pueden asumir una posición de indiferencia frente a sus posibles consecuencias. La previsibilidad del riesgo, activa deberes correlativos de evaluación, supervisión y mitigación.

En este contexto, la posición de garante adquiere una dimensión preventiva. Los Estados, las autoridades militares, los organismos reguladores y, en determinados ámbitos, los desarrolladores y fabricantes de sistemas tecnológicos, ejercen distintos niveles de control sobre fuentes potenciales de riesgo. En consecuencia, la obligación jurídica no surge únicamente cuando el daño ya se ha materializado, sino desde el momento en que resulta razonablemente posible identificar escenarios de afectación que puedan comprometer bienes jurídicos fundamentales.

La interpretación realizada por la Corte Constitucional respecto del artículo 25 del Código Penal ofrece elementos particularmente relevantes para esta discusión. La Corporación ha explicado que la posición de garante puede originarse, entre otros supuestos, en la asunción de protección sobre una fuente de riesgo dentro del propio ámbito de dominio, en el desarrollo de actividades riesgosas o en la creación previa de situaciones capaces de generar peligro para bienes jurídicos protegidos (Sentencia SU-1184 de 2001).

Trasladada al ámbito de la inteligencia artificial militar, esta lógica permite sostener que quienes diseñan, autorizan o utilizan sistemas de alto riesgo participan en distintos niveles de administración de una fuente tecnológica potencialmente lesiva. Por ello, la identificación de riesgos algorítmicos previsibles no constituye una simple advertencia técnica, sino un elemento capaz de activar deberes jurídicos de prevención compatibles tanto con la teoría general del riesgo como con los principios de precaución y protección que subyacen al Derecho Internacional Humanitario y los Derechos Humanos.

Desde esta perspectiva, el tratamiento jurídico de la inteligencia artificial militar no puede construirse exclusivamente sobre esquemas de responsabilidad posteriores al daño. La verdadera función de la posición de garante consiste en desplazar la atención hacia la prevención, exigiendo que quienes controlan fuentes tecnológicas de riesgo adopten medidas razonables destinadas a evitar que riesgos identificables evolucionen hacia amenazas concretas, peligros inminentes o daños incompatibles con la protección de la persona humana.

Esta lógica preventiva también puede identificarse en la concepción contemporánea de la seguridad, la cual resulta entendida por la jurisprudencia en una obligación de medio y no de resultado. Adicionalmente, la Corte resaltó que la noción de seguridad se proyecta en tres dimensiones a saber, (i) como un valor constitucional, (ii) como un derecho colectivo y (iii) como un derecho fundamental. La Corte ha señalado que el derecho a la seguridad personal no se ciñe únicamente a los eventos en los que esté comprometida la libertad individual (protección de las personas privadas de la libertad), sino que comprende todas aquellas garantías que por cualquier circunstancia pueden verse afectadas y que necesitan protección por parte del Estado; concretamente, la vida y la integridad personal como derechos básicos para la existencia misma de las personas. (Sentencia T – 224 de 2014).

Desde una perspectiva jurídica, la seguridad no implica la eliminación absoluta de todos los riesgos ni la desaparición total de las pérdidas potenciales. Por el contrario, supone la existencia de mecanismos institucionales destinados a administrar, reducir y controlar riesgos dentro de niveles razonablemente aceptables. El reconocimiento de la posibilidad de pérdidas constituye precisamente uno de los fundamentos que justifican la adopción de medidas preventivas y de protección.

Esta reflexión adquiere especial importancia frente a los sistemas de inteligencia artificial de alto riesgo utilizados en contextos de conflicto armado. En estos escenarios, el problema jurídico no consiste en determinar si el riesgo puede eliminarse completamente, pues toda operación militar implica inevitablemente determinados niveles de

incertidumbre. La cuestión central consiste en establecer qué riesgos son previsibles, quién tiene la obligación de identificarlos y cuáles son las medidas razonables que deben adoptarse para impedir que dichos riesgos evolucionen hacia amenazas, peligros o daños incompatibles con las exigencias del Derecho Internacional Humanitario y de los Derechos Humanos.

En consecuencia, la teoría jurídica del riesgo ofrece un marco conceptual particularmente útil para analizar la utilización de sistemas de inteligencia artificial en operaciones militares. Desde esta perspectiva, la regulación de estas tecnologías no debe construirse exclusivamente a partir de los daños ya producidos, sino principalmente desde la identificación temprana de riesgos previsibles y desde la adopción de mecanismos orientados a su prevención, mitigación y control.

2.1. Riesgo, amenaza, peligro y daño como categorías aplicables a los sistemas de inteligencia artificial de alto riesgo

La utilización de sistemas de inteligencia artificial en ámbitos sensibles de la actividad humana ha reabierto el debate sobre la capacidad del derecho para anticipar y gestionar riesgos derivados de tecnologías emergentes. Aunque la inteligencia artificial suele presentarse como una herramienta orientada a optimizar procesos de toma de decisiones y aumentar los niveles de eficiencia, su creciente incorporación en sectores críticos demuestra que los beneficios potenciales coexisten con riesgos cuya magnitud aún no ha sido completamente determinada.

La preocupación por estos riesgos ha llevado a diversos organismos internacionales y reguladores a desarrollar modelos normativos basados en la clasificación de sistemas según su nivel de peligrosidad. La aproximación regulatoria más notable es el Reglamento de Inteligencia Artificial de la Unión Europea (AI Act), instrumento que adopta un modelo escalonado basado en la gestión del riesgo y distingue entre sistemas de riesgo inaceptable, alto riesgo, riesgo limitado y riesgo mínimo. Este modelo parte de una premisa fundamental: no toda inteligencia artificial genera las mismas consecuencias jurídicas ni requiere idénticos niveles de control regulatorio.

La categoría de alto riesgo resulta especialmente relevante para el presente estudio debido a que comprende sistemas capaces de producir afectaciones significativas sobre derechos fundamentales, seguridad pública, infraestructuras críticas y procesos decisorios de especial sensibilidad. En estos casos, la regulación jurídica no se fundamenta exclusivamente en los daños ya producidos, sino en la existencia de una probabilidad razonable de afectación futura que justifica la adopción de medidas preventivas.

Desde esta perspectiva, los conceptos de riesgo, amenaza, peligro y daño adquieren una especial utilidad para el análisis jurídico de la inteligencia artificial militar. Para efectos del presente escrito, el riesgo algorítmico[1] puede entenderse como la posibilidad abstracta de que un sistema de inteligencia artificial produzca decisiones, recomendaciones o acciones capaces de afectar bienes jurídicos protegidos como consecuencia de errores de diseño, sesgos de entrenamiento, deficiencias en los datos utilizados o limitaciones inherentes al funcionamiento del propio sistema.

En esta fase inicial, la afectación todavía no se ha materializado. Sin embargo, la mera existencia de una probabilidad significativa de daño resulta suficiente para justificar

[1] La definición propuesta se construye a partir de la teoría de la sociedad del riesgo desarrollada por Beck (1998), de los enfoques regulatorios basados en riesgo adoptados por el Reglamento Europeo de Inteligencia Artificial (AI Act) y de la literatura contemporánea sobre riesgos asociados a sistemas algorítmicos.

mecanismos de supervisión y control. El derecho contemporáneo ha abandonado progresivamente la lógica reactiva que exigía la producción efectiva del perjuicio para activar consecuencias jurídicas. En su lugar, se ha consolidado una visión preventiva orientada a intervenir antes de que el daño ocurra.

La amenaza aparece cuando el riesgo deja de ser una hipótesis abstracta y se convierte en una posibilidad concreta de afectación. En el ámbito militar, ello podría ocurrir cuando un sistema de inteligencia artificial presenta fallos identificados que comprometen la correcta selección de objetivos, generan errores sistemáticos en la identificación de combatientes o incrementan significativamente la probabilidad de afectación a la población civil. En tales circunstancias, el riesgo adquiere una dimensión tangible que exige respuestas inmediatas por parte de las autoridades competentes.

El peligro constituye un grado de concreción aún más avanzado dentro de esta secuencia conceptual. Se configura cuando la posibilidad de afectación alcanza niveles de proximidad que hacen razonablemente previsible la producción del daño. En operaciones militares apoyadas por inteligencia artificial, el peligro podría manifestarse durante la ejecución de ataques en los cuales la información procesada por el sistema resulta insuficiente, desactualizada o incompatible con las condiciones reales del entorno operacional.

Finalmente, el daño representa la materialización efectiva de la afectación. En contextos de conflicto armado, dicho daño puede traducirse en pérdidas de vidas humanas, lesiones a personas protegidas, destrucción de bienes civiles, ataques indiscriminados o vulneraciones de derechos fundamentales. En estos casos, las consecuencias jurídicas ya no se limitan al ámbito preventivo, sino que involucran cuestiones relativas a responsabilidad estatal, responsabilidad individual y reparación de las víctimas.

La importancia de esta secuencia radica en que permite comprender que el riesgo algorítmico no surge exclusivamente en el momento en que se ejecuta una operación militar. Por el contrario, puede originarse mucho antes, durante las fases de diseño, programación, entrenamiento, validación y despliegue de los sistemas de inteligencia artificial. Esta circunstancia posee profundas implicaciones jurídicas, pues desplaza parcialmente la atención desde la respuesta frente al daño consumado hacia la identificación temprana de riesgos previsibles.

En consecuencia, la regulación de los sistemas de inteligencia artificial de alto riesgo no puede limitarse a evaluar los resultados producidos por la tecnología. Resulta igualmente necesario examinar las condiciones bajo las cuales dicha tecnología fue concebida, desarrollada y autorizada para su utilización. Desde esta perspectiva, la respuesta anticipatoria jurídica a la inteligencia artificial exige una aproximación preventiva basada en la gestión integral del riesgo, especialmente cuando los sistemas son empleados en escenarios de conflicto armado donde los márgenes de error pueden traducirse en consecuencias irreparables para la vida, la integridad y la dignidad humana.

2.2. La sociedad del riesgo y los riesgos manufacturados: de Ulrich Beck a la inteligencia artificial militar

La creciente incorporación de sistemas de inteligencia artificial en ámbitos tradicionalmente reservados a la decisión humana constituye una manifestación paradigmática de las transformaciones que caracterizan a las sociedades contemporáneas. Lejos de tratarse únicamente de un avance tecnológico, la inteligencia artificial representa una expresión de los nuevos riesgos derivados del propio desarrollo científico y técnico, fenómeno ampliamente estudiado por la teoría de la sociedad del riesgo.

Ulrich Beck sostiene que la modernidad produjo una transformación sustancial en la naturaleza de los riesgos. Mientras las

sociedades preindustriales se enfrentaban principalmente a amenazas derivadas de fenómenos naturales, las sociedades contemporáneas se caracterizan por la producción de riesgos que son consecuencia del propio desarrollo científico, industrial y tecnológico. En este sentido, la preocupación jurídica ya no se limita a la reparación de daños consumados, sino que se desplaza progresivamente hacia la identificación y prevención de riesgos futuros generados por la propia actividad humana (Beck, 1998).

Los riesgos manufacturados se caracterizan por su complejidad, dificultad de predicción y potencial capacidad de producir consecuencias de gran magnitud. A diferencia de los peligros naturales, cuya fuente suele encontrarse fuera del control humano, los riesgos manufacturados son el resultado directo de procesos de innovación desarrollados por la propia sociedad. En consecuencia, la discusión jurídica deja de centrarse exclusivamente en la reacción frente al daño consumado para trasladarse hacia la prevención, la evaluación y la gestión de riesgos futuros.

Anthony Giddens complementa esta perspectiva al señalar que las sociedades contemporáneas se encuentran obligadas a adoptar decisiones permanentes en contextos marcados por la incertidumbre y por riesgos derivados de la propia modernidad. En este escenario, la gestión del riesgo se convierte en una función esencial de las instituciones públicas y privadas, llamadas a enfrentar amenazas cuya magnitud y consecuencias no siempre pueden ser determinadas con absoluta certeza (Giddens, 2000).

La inteligencia artificial constituye uno de los ejemplos más representativos de este fenómeno. Su capacidad para procesar grandes volúmenes de información, identificar patrones complejos y apoyar procesos de toma de decisiones ha transformado sectores como la salud, las finanzas, el transporte, la seguridad y la defensa. Sin embargo, la misma tecnología que promete aumentar la eficiencia también introduce riesgos derivados de la opacidad algorítmica, la dependencia tecnológica, los sesgos de entrenamiento y la progresiva autonomía de determinados sistemas.

En el ámbito militar, esta situación adquiere una importancia inusitada. La utilización de inteligencia artificial en sistemas de vigilancia, reconocimiento, adquisición de objetivos, procesamiento de inteligencia y apoyo a la toma de decisiones operacionales incorpora nuevos niveles de complejidad jurídica. La promesa tecnológica consiste en reducir errores humanos, aumentar la precisión y optimizar el empleo de la fuerza. No obstante, la propia tecnología genera incertidumbres adicionales relacionadas con la confiabilidad de los datos, la transparencia de los algoritmos y la posibilidad de resultados no previstos por sus diseñadores.

Precisamente aquí surge una de las principales preocupaciones jurídicas. El riesgo asociado a la inteligencia artificial militar no se limita al funcionamiento final del sistema. Su origen puede encontrarse en etapas previas de diseño, programación, entrenamiento, validación y despliegue.

Uno de los principales desafíos de los sistemas de inteligencia artificial aplicados a operaciones militares radica en que los algoritmos son diseñados y entrenados por equipos técnicos que, en muchos casos, carecen de experiencia directa en la conducción de hostilidades, el derecho operacional y la aplicación práctica de los principios del Derecho Internacional Humanitario en escenarios reales de combate.

Con fundamento en ello, surge la siguiente pregunta: ¿Puede un programador civil anticipar adecuadamente las variables jurídicas, éticas y operacionales que enfrenta un comandante militar en una situación real de combate?

Si el algoritmo es entrenado con criterios inadecuados o incompletos, el riesgo no surge únicamente durante la ejecución de la operación militar, sino desde la propia fase de diseño y programación. Dicho esto, en los

sistemas de inteligencia artificial aplicados a conflictos armados, el riesgo jurídico no nace exclusivamente del empleo operacional de la tecnología, sino que puede originarse en etapas previas de diseño, programación, entrenamiento y validación algorítmica.

Un algoritmo construido sobre datos insuficientes, sesgados o desactualizados puede producir decisiones técnicamente coherentes desde una perspectiva matemática, pero incompatibles con las exigencias jurídicas derivadas del Derecho Internacional Humanitario y de los Derechos Humanos.

Esta circunstancia adquiere particular relevancia cuando se advierte que quienes diseñan los sistemas no necesariamente participan de manera directa en las operaciones militares ni poseen experiencia operacional equivalente a la de los comandantes responsables de la conducción de hostilidades. La distancia existente entre el entorno controlado de desarrollo tecnológico y la complejidad del campo de batalla introduce nuevas variables de riesgo que no pueden ser ignoradas por el derecho.

Desde esta perspectiva, los sistemas de inteligencia artificial utilizados en conflictos armados constituyen riesgos manufacturados por excelencia. No se trata de amenazas naturales ni inevitables, sino de riesgos generados por decisiones humanas susceptibles de evaluación, regulación y control. En consecuencia, la respuesta jurídica no puede limitarse a reaccionar frente a daños ya producidos. Resulta necesario desarrollar mecanismos preventivos capaces de identificar y mitigar los riesgos antes de su materialización, especialmente cuando se encuentran comprometidos bienes jurídicos tan relevantes como la vida, la integridad personal y la protección de la población civil.

La teoría de la sociedad del riesgo ofrece así un marco conceptual particularmente útil para comprender los desafíos regulatorios planteados por la inteligencia artificial militar. Bajo esta óptica, la cuestión jurídica fundamental no consiste en determinar si la tecnología puede eliminar completamente la incertidumbre inherente a la guerra, sino en establecer cuáles son los riesgos previsibles derivados de su utilización y quién debe asumir la responsabilidad de prevenirlos, controlarlos y mitigarlos.

La inteligencia artificial constituye el intento más sofisticado de la historia por reducir la incertidumbre del combate; sin embargo, la acumulación masiva de datos no elimina la niebla de guerra, sino que desplaza parte de la incertidumbre hacia el diseño y funcionamiento de los algoritmos.

2.3. La niebla de guerra, la fricción y la ilusión de la certeza tecnológica

Uno de los mayores desafíos para el análisis jurídico de la inteligencia artificial aplicada a los conflictos armados consiste en comprender que la guerra continúa siendo una actividad marcada por la incertidumbre. A pesar de los extraordinarios avances tecnológicos desarrollados durante las últimas décadas, la promesa de eliminar completamente la incertidumbre operacional ha acompañado históricamente a numerosos procesos de innovación militar sin llegar a materializarse plenamente.

Al respecto, las reflexiones de Carl von Clausewitz sobre la incertidumbre, la información incompleta y la fricción en las operaciones militares constituyen uno de los antecedentes teóricos más importantes de lo que posteriormente la doctrina militar denominaría “niebla de guerra”. Para el autor, la conducción de las hostilidades se desarrolla en un entorno caracterizado por información imperfecta, percepciones incompletas y acontecimientos imprevistos que dificultan la ejecución de los planes inicialmente concebidos (Clausewitz, 2005).

La relevancia contemporánea de estas categorías resulta evidente frente a la creciente incorporación de sistemas de inteligencia artificial en escenarios militares. Los algoritmos prometen reducir los niveles de incertidumbre mediante el procesamiento masivo de datos, la identificación de patrones

y la generación de predicciones cada vez más sofisticadas. Sin embargo, la posibilidad de procesar grandes volúmenes de información no implica necesariamente la eliminación de las condiciones estructurales de incertidumbre que caracterizan a los conflictos armados.

Esta tensión entre tecnología e incertidumbre puede observarse en las reflexiones formuladas por Robert McNamara tras la guerra de Vietnam[2], toda vez que, resultan especialmente ilustrativas para comprender los límites de los sistemas de decisión apoyados en grandes volúmenes de información. La confianza en métricas, indicadores y modelos cuantitativos fue concebida como un mecanismo para reducir la incertidumbre inherente al conflicto; sin embargo, la experiencia demostró que la disponibilidad de datos no garantiza necesariamente una comprensión adecuada de la realidad operacional ni elimina los factores humanos, políticos y estratégicos que influyen en la conducción de las hostilidades (McNamara, 1995).

La experiencia vietnamita puso de manifiesto que la complejidad de la guerra excede los límites de los modelos puramente cuantitativos. Esta reflexión conserva plena vigencia frente a los actuales sistemas de inteligencia artificial, cuya capacidad para procesar enormes volúmenes de información no elimina necesariamente la incertidumbre inherente a los conflictos armados ni sustituye el juicio humano requerido para adoptar decisiones compatibles con el Derecho Internacional Humanitario.

La inteligencia artificial contemporánea representa, probablemente, el intento más sofisticado desarrollado hasta ahora para reducir la incertidumbre inherente a las operaciones militares. Sistemas de vigilancia automatizada, reconocimiento de objetivos, análisis predictivo, apoyo a la toma de decisiones y plataformas de armas asistidas por algoritmos buscan precisamente incrementar los niveles de información disponibles para los comandantes y reducir los márgenes de error en la conducción de hostilidades.

Sin embargo, la capacidad tecnológica para procesar datos no elimina necesariamente la incertidumbre jurídica ni operacional. Los algoritmos operan sobre información previamente recopilada, clasificada y procesada. En consecuencia, la calidad de sus resultados depende directamente de la calidad de los datos utilizados durante su entrenamiento, de los criterios incorporados por sus programadores y de las condiciones bajo las cuales son desplegados.

Lo anterior se evidencia, como se mencionó líneas atrás, en que quienes diseñan los algoritmos no necesariamente poseen experiencia directa en la conducción de operaciones militares ni participan en los escenarios donde finalmente serán utilizados los sistemas. El programador desarrolla modelos matemáticos en entornos controlados, mientras que el comandante militar adopta decisiones en contextos caracterizados por la presión temporal, la incertidumbre táctica, las exigencias jurídicas derivadas del Derecho Internacional Humanitario y del DDHH.

La distancia entre el diseño algorítmico y la realidad operacional permite sostener que la inteligencia artificial no elimina la niebla de guerra. Más bien, transforma parte de la incertidumbre tradicional en nuevas formas de riesgo asociadas al funcionamiento de los propios sistemas tecnológicos. La pregunta jurídica deja entonces de ser únicamente cómo reducir la incertidumbre del combate y

[2] La experiencia de Robert McNamara como Secretario de Defensa de los Estados Unidos durante la guerra de Vietnam constituye uno de los ejemplos más citados sobre las limitaciones de los modelos de decisión excesivamente dependientes de indicadores cuantitativos y sistemas de medición. Véase: McNamara, Robert S., *In Retrospect: The Tragedy and Lessons of Vietnam*, New York, Times Books, 1995.

pasa a incluir la identificación de los riesgos derivados de confiar decisiones críticas a sistemas cuya lógica interna puede resultar opaca incluso para sus operadores.

Desde esta perspectiva, la inteligencia artificial militar no debe ser entendida como una herramienta capaz de sustituir completamente el juicio humano, sino como una tecnología que exige mecanismos reforzados de supervisión, control y responsabilidad. Precisamente porque la incertidumbre continúa existiendo, aunque bajo formas diferentes, el derecho conserva un papel esencial en la regulación de los riesgos asociados a la utilización de sistemas de inteligencia artificial de alto riesgo en conflictos armados.

3. Los sistemas de inteligencia artificial de alto riesgo y sus implicaciones jurídicas

La inteligencia artificial se ha convertido en una de las tecnologías más influyentes del siglo XXI. Su capacidad para procesar grandes volúmenes de información, identificar patrones complejos, generar predicciones y apoyar procesos de toma de decisiones ha transformado múltiples sectores de la actividad humana. No obstante, el impacto potencial de estas tecnologías no es uniforme. Algunos sistemas poseen una capacidad significativamente mayor para afectar derechos fundamentales, la seguridad pública o bienes jurídicos de especial protección, razón por la cual diversos instrumentos regulatorios han adoptado modelos diferenciados basados en el nivel de riesgo asociado a su utilización.

La aproximación regulatoria más relevante en esta materia se encuentra en el Reglamento de Inteligencia Artificial de la Unión Europea (AI Act), el cual establece un modelo escalonado de gestión del riesgo. Este instrumento parte de la premisa según la cual el nivel de intervención jurídica debe ser proporcional a los riesgos que cada sistema puede generar. En consecuencia, distingue entre sistemas de riesgo inaceptable, alto riesgo, riesgo limitado y riesgo mínimo.

Los sistemas de inteligencia artificial de alto riesgo ocupan una posición particularmente relevante dentro de esta clasificación. Se trata de tecnologías cuya utilización puede afectar significativamente la vida, la integridad personal, la seguridad, los derechos fundamentales o el funcionamiento de infraestructuras críticas. Por esta razón, se encuentran sometidos a exigencias reforzadas de transparencia, supervisión humana, calidad de datos, trazabilidad y gestión del riesgo.

Desde una perspectiva funcional, los sistemas de alto riesgo suelen compartir varias características. En primer lugar, participan en procesos de toma de decisiones con consecuencias relevantes para las personas. En segundo lugar, operan en entornos donde los errores pueden producir daños de difícil reparación. Finalmente, su funcionamiento depende de modelos algorítmicos cuya complejidad dificulta, en ocasiones, la comprensión integral de los procesos que conducen a determinados resultados.

Estas preocupaciones adquieren una dimensión aún mayor cuando la inteligencia artificial es incorporada a contextos militares. La utilización de algoritmos en actividades de vigilancia, reconocimiento, identificación de objetivos, análisis predictivo, procesamiento de inteligencia o apoyo a decisiones operacionales introduce riesgos que trascienden las preocupaciones tradicionales asociadas a otros sectores. En estos escenarios, la relevancia jurídica no deriva exclusivamente de la capacidad destructiva de las armas empleadas, sino también de la posibilidad de que los sistemas tecnológicos participen en procesos decisorios relacionados con el uso de la fuerza. El problema ya no consiste únicamente en determinar si un arma produce daños, sino en establecer cómo se generan las decisiones que conducen a su empleo y cuáles son los mecanismos disponibles para prevenir errores con consecuencias humanitarias.

Una de las principales dificultades radica en que el riesgo algorítmico no se limita a la fase operacional. Tradicionalmente, el análisis jurídico de los conflictos armados se ha concentrado en la conducta desplegada durante las hostilidades. Sin embargo, en los sistemas de inteligencia artificial una parte significativa del riesgo puede originarse mucho antes de que se produzca cualquier acción militar.

Esto por cuanto, el diseño del algoritmo, la selección de datos de entrenamiento, la definición de variables relevantes, los procesos de validación y las condiciones de despliegue constituyen etapas capaces de influir decisivamente en los resultados posteriores. En consecuencia, los errores, sesgos o limitaciones incorporados durante estas fases iniciales pueden proyectarse sobre las decisiones adoptadas en escenarios reales de conflicto. Esta circunstancia plantea una cuestión jurídica fundamental que podemos dilucidar con la siguiente afirmación: si el riesgo puede identificarse desde etapas previas al empleo operacional del sistema, entonces los deberes de prevención también deben surgir desde ese momento.

Bajo esta lógica, la gestión jurídica del riesgo algorítmico no puede limitarse a evaluar daños consumados, sino que debe incorporar mecanismos orientados a identificar, controlar y mitigar riesgos previsibles desde las fases iniciales de desarrollo tecnológico.

Desde esta perspectiva, los sistemas de inteligencia artificial militar constituyen una manifestación paradigmática de los riesgos manufacturados propios de las sociedades tecnológicas contemporáneas. Precisamente por ello, su utilización exige marcos regulatorios capaces de equilibrar los beneficios potenciales de la innovación con la protección efectiva de la vida, la dignidad humana y los principios fundamentales del Derecho Internacional Humanitario.

3.1. El riesgo algorítmico y la posición de garante: hacia una responsabilidad preventiva en los conflictos armados

La discusión jurídica sobre inteligencia artificial suele concentrarse en la determinación de responsabilidades una vez producido el daño. Sin embargo, esta aproximación resulta insuficiente frente a tecnologías cuya capacidad de generar afectaciones significativas puede identificarse incluso antes de su utilización efectiva. Desde la perspectiva de la teoría jurídica del riesgo, la cuestión central no consiste únicamente en determinar quién responde por las consecuencias de una decisión algorítmica errónea, sino en establecer quién tiene el deber jurídico de prevenir la materialización de riesgos previsibles.

Esta reflexión es esencial en los conflictos armados contemporáneos. La creciente incorporación de sistemas de inteligencia artificial en procesos de vigilancia, identificación de objetivos, análisis predictivo y apoyo a decisiones operacionales ha generado escenarios en los cuales los riesgos potenciales pueden detectarse desde etapas tempranas del desarrollo tecnológico. En consecuencia, la identificación de dichos riesgos plantea la necesidad de examinar los deberes jurídicos que recaen sobre los distintos actores involucrados en el diseño, desarrollo, validación y utilización de estas tecnologías.

En la teoría general del derecho, la posición de garante constituye una categoría fundamental para comprender los deberes positivos de protección. Tradicionalmente, esta figura ha sido utilizada para explicar aquellas situaciones en las que determinados sujetos adquieren obligaciones especiales de prevención respecto de bienes jurídicos cuya protección les ha sido confiada. En estos casos, la responsabilidad jurídica no surge exclusivamente por la realización de una conducta lesiva, sino también por la omisión de actuaciones necesarias para impedir la producción de daños previsibles.

La lógica subyacente a la posición de garante resulta particularmente útil para el análisis de los sistemas de inteligencia artificial de alto riesgo. Si el desarrollo tecnológico permite

identificar riesgos razonablemente previsibles, quienes participan en la creación, autorización o utilización de tales sistemas no pueden asumir una posición de indiferencia frente a sus posibles consecuencias. La previsibilidad del riesgo genera correlativamente deberes de evaluación, supervisión y control.

Desde esta perspectiva, el Estado ocupa una posición central. En virtud de sus obligaciones constitucionales e internacionales de protección, las autoridades públicas se encuentran llamadas a adoptar medidas destinadas a identificar, administrar y reducir aquellos riesgos tecnológicos capaces de afectar derechos fundamentales. Esta obligación adquiere una intensidad reforzada cuando la tecnología es utilizada en contextos donde pueden verse comprometidos bienes jurídicos como la vida, la integridad personal o la protección de la población civil.

No obstante, la posición de garante no recae exclusivamente sobre el Estado. La complejidad de los sistemas de inteligencia artificial distribuye funciones y responsabilidades entre diversos actores. Los desarrolladores participan en la construcción de los algoritmos; los fabricantes intervienen en la producción y validación de los sistemas; las autoridades militares autorizan su despliegue operacional; y los operadores ejecutan las decisiones dentro de escenarios concretos de conflicto. Cada uno de estos actores controla determinados factores de riesgo y, en consecuencia, asume deberes correlativos de prevención.

Particular atención merece la situación de quienes diseñan los algoritmos. A diferencia de los comandantes militares, los programadores desarrollan su actividad en entornos controlados y alejados de las condiciones reales del combate. Aunque poseen conocimientos especializados sobre arquitectura de sistemas y procesamiento de datos, no necesariamente cuentan con experiencia operacional directa ni con formación suficiente en Derecho Internacional Humanitario. Esta circunstancia no implica una transferencia automática de responsabilidad, pero sí evidencia la existencia de una brecha entre el diseño tecnológico y la realidad operacional que debe ser tenida en cuenta durante la evaluación jurídica del riesgo.

La relevancia de esta observación radica en que una parte importante del riesgo algorítmico puede originarse precisamente durante las fases de programación y entrenamiento. Los datos utilizados para construir el sistema, los criterios empleados para clasificar objetivos, las variables consideradas relevantes y los mecanismos de aprendizaje implementados pueden influir decisivamente en los resultados posteriores. De esta forma, determinadas decisiones adoptadas durante el desarrollo tecnológico pueden proyectar sus efectos mucho tiempo después, cuando el sistema sea utilizado en escenarios reales de conflicto armado.

La existencia de riesgos previsibles exige, por tanto, mecanismos preventivos de control. La supervisión humana significativa, las auditorías algorítmicas, las evaluaciones de impacto, la trazabilidad de las decisiones automatizadas y los procedimientos de validación previa constituyen instrumentos orientados precisamente a reducir la probabilidad de que riesgos identificables evolucionen hacia amenazas concretas, peligros inminentes o daños consumados.

Esta aproximación preventiva resulta coherente con la evolución contemporánea del derecho. La protección efectiva de los derechos fundamentales no puede depender exclusivamente de mecanismos reparadores posteriores a la producción del daño. En escenarios caracterizados por la magnitud potencial de las consecuencias, la función principal del derecho consiste precisamente en evitar que los riesgos previsibles lleguen a materializarse.

Por ello, el análisis jurídico de la inteligencia artificial militar no debe limitarse a la atribución de responsabilidad una vez ocurrido el daño. Resulta igualmente necesario determinar quiénes tienen el deber

de identificar, administrar y mitigar los riesgos asociados al funcionamiento de estos sistemas. En este sentido, la posición de garante ofrece un marco conceptual idóneo para comprender que la regulación y el control jurídico de la inteligencia artificial en conflictos armados exige responsabilidades preventivas distribuidas entre todos aquellos actores que participan en la generación y control del riesgo algorítmico.

4. El conflicto armado como escenario de riesgo reforzado

La valoración jurídica de los riesgos asociados a los sistemas de inteligencia artificial exige considerar el contexto en el cual dichas tecnologías son empleadas. Aunque los sistemas de alto riesgo pueden generar afectaciones relevantes en diversos ámbitos de la actividad humana, su utilización en escenarios de conflicto armado plantea desafíos particularmente complejos debido a la naturaleza de los bienes jurídicos comprometidos y a las consecuencias potenciales derivadas de errores operacionales.

La guerra constituye, por definición, una actividad caracterizada por elevados niveles de incertidumbre, violencia y riesgo. Sin embargo, la existencia de hostilidades no implica la suspensión del derecho ni la desaparición de los mecanismos jurídicos de protección. Por el contrario, el Derecho Internacional Humanitario surge precisamente como un conjunto de normas destinadas a limitar los efectos de los conflictos armados y reducir los sufrimientos innecesarios derivados de la confrontación.

La identificación de un conflicto armado constituye, por tanto, una cuestión jurídica fundamental. De ella depende la aplicación del régimen especial previsto por los Convenios de Ginebra de 1949, sus Protocolos Adicionales y las normas consuetudinarias del Derecho Internacional Humanitario.

Tradicionalmente, la doctrina ha distinguido entre conflictos armados internacionales y conflictos armados no internacionales. Los primeros encuentran su fundamento normativo principal en el artículo 2 común a los cuatro Convenios de Ginebra de 1949, el cual establece la aplicación del régimen convencional en caso de guerra declarada o de cualquier otro conflicto armado surgido entre dos o más Estados. Los segundos se encuentran regulados inicialmente por el artículo 3 común a los Convenios de Ginebra y posteriormente desarrollados por el Protocolo Adicional II de 1977.

No obstante, una de las definiciones más influyentes en la materia fue formulada por el Tribunal Penal Internacional para la ex Yugoslavia en la decisión sobre jurisdicción del caso Fiscal contra Duško Tadić (1995). En dicha providencia, el Tribunal sostuvo que existe conflicto armado siempre que se recurra a la fuerza armada entre Estados o cuando exista violencia armada prolongada entre autoridades gubernamentales y grupos armados organizados, o entre tales grupos dentro de un Estado.

La importancia de esta definición radica en que desplaza el análisis desde categorías formales hacia elementos materiales relacionados con la intensidad de la violencia y el grado de organización de los actores involucrados. Este criterio ha sido posteriormente acogido y desarrollado por diversos tribunales internacionales, organismos internacionales y por el propio Comité Internacional de la Cruz Roja.

El Comité Internacional de la Cruz Roja ha destacado que la finalidad esencial del Derecho Internacional Humanitario consiste en proteger a las personas que no participan o han dejado de participar en las hostilidades y limitar los medios y métodos de guerra empleados por las partes en conflicto. Esta finalidad es vital frente a la creciente incorporación de tecnologías capaces de intervenir en procesos de selección de objetivos, análisis de inteligencia y apoyo a decisiones operacionales.

Desde la perspectiva de la teoría del riesgo, los conflictos armados pueden ser entendidos como escenarios de riesgo reforzado. La razón

es evidente: los errores operacionales no producen únicamente consecuencias patrimoniales o administrativas, sino que pueden traducirse directamente en pérdidas de vidas humanas, lesiones graves, desplazamientos forzados o afectaciones masivas a la población civil.

De esta manera, la utilización de sistemas de inteligencia artificial en estos contextos exige estándares de evaluación particularmente rigurosos. La existencia de riesgos previsibles adquiere mayor nivel de preponderancia cuando las decisiones apoyadas por algoritmos pueden influir, directa o indirectamente, en el empleo de la fuerza armada o en la determinación de objetivos militares. Esta situación resulta especialmente significativa si se considera que los conflictos armados contemporáneos se desarrollan en entornos operacionales altamente dinámicos. La información disponible puede modificarse constantemente, los actores involucrados cambian de posición, los objetivos evolucionan y las condiciones tácticas varían en cuestión de minutos. En tales circunstancias, la confiabilidad de los sistemas tecnológicos y la capacidad de supervisión humana adquieren una importancia decisiva.

Precisamente por ello, la incorporación de inteligencia artificial en escenarios de conflicto armado no puede analizarse únicamente desde una perspectiva tecnológica. Resulta indispensable examinar su compatibilidad con las obligaciones jurídicas derivadas del Derecho Internacional Humanitario y con los principios que regulan la conducción de las hostilidades. La evaluación de riesgos debe extenderse no solo a las capacidades técnicas del sistema, sino también a las consecuencias jurídicas y humanitarias potencialmente asociadas a su utilización.

Desde esta perspectiva, el conflicto armado constituye el entorno en el cual el riesgo algorítmico alcanza su máxima expresión jurídica. La combinación entre incertidumbre operacional, empleo de la fuerza y potencial afectación de personas protegidas convierte a estos escenarios en espacios donde la prevención, la mitigación y el control de riesgos adquieren una importancia particularmente intensa. Precisamente por esta razón, el análisis de la inteligencia artificial militar debe continuar a partir de los principios fundamentales que estructuran el Derecho Internacional Humanitario y que orientan la conducción legítima de las hostilidades.

5. Inteligencia artificial militar y principios fundamentales del derecho internacional humanitario

La creciente incorporación de sistemas de inteligencia artificial en operaciones militares ha reabierto el debate sobre la compatibilidad de las nuevas tecnologías con los principios fundamentales del Derecho Internacional Humanitario. Aunque la innovación tecnológica ha acompañado históricamente la evolución de los conflictos armados, la utilización de algoritmos capaces de participar en procesos de identificación de objetivos, análisis de inteligencia y apoyo a decisiones operacionales plantea interrogantes jurídicos que trascienden las discusiones tradicionales sobre medios y métodos de guerra.

La cuestión central no consiste únicamente en determinar si la inteligencia artificial puede incrementar la eficacia operacional o reducir determinados errores humanos. El verdadero problema jurídico radica en establecer si estas tecnologías son capaces de operar dentro de los límites impuestos por el Derecho Internacional Humanitario y si los riesgos derivados de su utilización resultan compatibles con las obligaciones internacionales asumidas por los Estados.

En este contexto, el análisis debe partir de los principios que históricamente han regulado la conducción de las hostilidades: distinción, proporcionalidad, precaución, humanidad y necesidad militar. Estos principios constituyen límites jurídicos destinados a reducir los efectos de la violencia armada y a proteger a quienes no participan directamente en las hostilidades.

5.1. El principio de distinción frente a la toma de decisiones algorítmicas

El principio de distinción constituye la piedra angular del Derecho Internacional Humanitario. Su formulación convencional más clara se encuentra en el artículo 48 del Protocolo Adicional I de 1977, conforme al cual las partes en conflicto deben distinguir en todo momento entre población civil y combatientes, así como entre bienes de carácter civil y objetivos militares, dirigiendo sus operaciones únicamente contra estos últimos. Este principio ha sido igualmente reconocido como una norma consuetudinaria del Derecho Internacional Humanitario aplicable tanto a los conflictos armados internacionales como a los no internacionales (Henckaerts y Doswald-Beck, 2005)

Así, encontramos que la dificultad de los sistemas algorítmicos no consiste en identificar objetos, sino en identificar jurídicamente objetivos militares conforme a las exigencias del principio de distinción.

La aplicación de este principio exige valoraciones complejas que frecuentemente dependen del contexto específico en que se desarrolla la operación. La identificación de un objetivo militar no siempre constituye una cuestión puramente técnica. En numerosas ocasiones requiere interpretar comportamientos, analizar información contradictoria, valorar circunstancias cambiantes y comprender elementos contextuales imposibles de reducir completamente a parámetros matemáticos.

Los sistemas de inteligencia artificial pueden contribuir significativamente a la identificación de patrones y al procesamiento de grandes volúmenes de información. Sin embargo, la capacidad de reconocer correlaciones estadísticas no equivale necesariamente a la capacidad de comprender contextos jurídicos y operacionales complejos. De esta manera, el riesgo de clasificación errónea continúa existiendo incluso cuando las decisiones se encuentran apoyadas por tecnologías avanzadas.

Precisamente por ello, la identificación de objetivos mediante inteligencia artificial debe permanecer sometida a mecanismos efectivos de supervisión humana que permitan corregir errores, contextualizar información y garantizar el respeto por las exigencias derivadas del principio de distinción.

5.2. El principio de proporcionalidad y los límites de la predicción algorítmica

El principio de proporcionalidad constituye uno de los mecanismos fundamentales de protección de la población civil durante los conflictos armados. Conforme al artículo 51.5.b del Protocolo Adicional I de 1977, se encuentran prohibidos aquellos ataques respecto de los cuales sea previsible que los daños incidentales a civiles o bienes de carácter civil resulten excesivos en relación con la ventaja militar concreta y directa prevista. Este principio exige una valoración prospectiva que permita ponderar la ventaja militar esperada frente a los posibles efectos humanitarios derivados de la operación, constituyendo una de las decisiones más complejas en la conducción de las hostilidades.

La inteligencia artificial puede estimar probabilidades de daño, pero la determinación de cuándo dichos daños resultan jurídicamente excesivos continúa siendo una valoración que pertenece al ámbito del juicio humano.

A diferencia de lo que ocurre con ciertos procesos técnicos, la proporcionalidad exige valoraciones jurídicas, operacionales y éticas que difícilmente pueden ser reducidas a cálculos automáticos. La determinación de la ventaja militar esperada y la evaluación de los daños incidentales previsibles implican juicios prospectivos sujetos a incertidumbre.

Los sistemas de inteligencia artificial pueden proporcionar estimaciones y escenarios probabilísticos que apoyen la toma de decisiones. Sin embargo, la decisión final sobre la proporcionalidad de un ataque continúa exigiendo apreciaciones humanas relacionadas con factores contextuales,

estratégicos y jurídicos que exceden la lógica puramente algorítmica.

Desde la perspectiva del riesgo, la confianza excesiva en sistemas automatizados puede generar una falsa percepción de certeza que conduzca a subestimar variables no incorporadas en los modelos utilizados.

5.3. El principio de precaución y la gestión del riesgo algorítmico

El principio de precaución ocupa una posición central dentro del Derecho Internacional Humanitario, particularmente en relación con la planificación y ejecución de operaciones militares. Su fundamento normativo se encuentra en el artículo 57 del Protocolo Adicional I de 1977, disposición que exige adoptar todas las precauciones factibles para verificar la naturaleza militar de los objetivos y para evitar o reducir al mínimo los daños incidentales a la población civil y a los bienes de carácter civil.

La lógica que inspira este principio es esencialmente preventiva. A diferencia de los mecanismos de responsabilidad posteriores al daño, las obligaciones de precaución buscan intervenir antes de que las consecuencias lesivas lleguen a materializarse. En este sentido, la identificación temprana de riesgos constituye uno de sus componentes fundamentales. Desde esta perspectiva, la gestión del riesgo algorítmico puede entenderse como una manifestación contemporánea del principio de precaución.

Si los riesgos asociados a los sistemas de inteligencia artificial pueden ser identificados durante las fases de diseño, entrenamiento, validación o despliegue operacional, surge el deber jurídico de adoptar medidas razonables orientadas a su mitigación. La supervisión humana, las auditorías técnicas, la verificación de datos y la evaluación permanente de los sistemas constituyen mecanismos compatibles con las exigencias preventivas que fundamentan el Derecho Internacional Humanitario.

Este principio exige adoptar todas las precauciones factibles para evitar o reducir al mínimo los daños incidentales a la población civil y a los bienes de carácter civil. Su lógica es esencialmente preventiva y se encuentra estrechamente vinculada con la teoría jurídica del riesgo desarrollada en los apartados anteriores.

La identificación de riesgos previsibles asociados a sistemas de inteligencia artificial activa obligaciones de evaluación, supervisión y control. Cuanto mayor sea la capacidad potencial de una tecnología para producir consecuencias graves, mayor deberá ser el nivel de diligencia exigido a quienes participan en su desarrollo y utilización.

Desde esta perspectiva, la gestión del riesgo algorítmico constituye una manifestación contemporánea del principio de precaución. La realización de pruebas, auditorías, validaciones operacionales y mecanismos de supervisión humana no representa únicamente una buena práctica tecnológica, sino una exigencia jurídica derivada de los deberes de prevención propios del Derecho Internacional Humanitario.

5.4. Los blancos de oportunidad, la gobernabilidad de la decisión y la irreversibilidad operacional

Uno de los desafíos más complejos asociados a la utilización de sistemas de inteligencia artificial en operaciones militares surge cuando la tecnología participa en procesos dinámicos de identificación, priorización o modificación de objetivos durante el desarrollo de una operación. A diferencia de los escenarios estáticos de planificación, los conflictos armados contemporáneos se caracterizan por cambios permanentes en la información disponible, en la ubicación de los actores y en la valoración operacional de los objetivos militares.

En este contexto, puede ocurrir que una operación haya sido inicialmente autorizada contra un objetivo previamente validado conforme a los principios del Derecho Internacional Humanitario y que, durante su ejecución, los sistemas de vigilancia, reconocimiento o procesamiento algorítmico

identifiquen un objetivo distinto que aparente ofrecer una ventaja militar superior. Desde una perspectiva operacional, esta situación suele asociarse con los denominados blancos de oportunidad, entendidos como objetivos identificados durante el desarrollo de una operación que no habían sido contemplados inicialmente, pero que adquieren relevancia militar inmediata.

La incorporación de sistemas de inteligencia artificial a estos procesos plantea interrogantes jurídicos particularmente complejos. Si el algoritmo identifica un nuevo objetivo y recomienda modificar la decisión inicialmente adoptada, surge la necesidad de determinar quién conserva la capacidad efectiva de gobernar dicha decisión. La cuestión no se limita a establecer quién autorizó el empleo inicial de la fuerza, sino quién controla la modificación posterior de la decisión y bajo qué criterios jurídicos se valida dicha alteración.

La dificultad aumenta cuando la información utilizada por el sistema resulta incompleta, desactualizada o incorrecta. Ningún sistema de inteligencia artificial puede producir decisiones confiables a partir de inteligencia operacional estructuralmente defectuosa. Si la información de entrada contiene errores de identificación, clasificación o contextualización, el sistema puede amplificar dichos errores y proyectarlos sobre decisiones capaces de afectar directamente a la población civil o a bienes protegidos por el Derecho Internacional Humanitario.

Ahora bien, desde la ya pluricitada perspectiva de la teoría jurídica del riesgo, estas situaciones permiten observar con claridad la transición progresiva entre riesgo, amenaza, peligro y daño. Inicialmente puede existir un riesgo asociado a la posibilidad abstracta de error en la identificación del objetivo. Posteriormente, cuando el sistema recomienda una modificación concreta de la decisión operacional, el riesgo puede transformarse en una amenaza identificable.

Si la operación continúa avanzando y la posibilidad de intervención correctiva disminuye, la situación evoluciona hacia un peligro real e inminente. Finalmente, cuando la decisión se ejecuta y produce una afectación efectiva sobre personas o bienes protegidos, se materializa el daño.

Esta evolución alcanza su máxima expresión cuando la operación alcanza un punto de irreversibilidad material. Mientras exista capacidad efectiva de intervención humana, la decisión continúa siendo susceptible de corrección, validación o cancelación. Sin embargo, determinadas armas y sistemas operacionales pueden alcanzar fases en las cuales la posibilidad de modificar o revertir la decisión desaparece total o parcialmente. Una munición ya disparada, una ráfaga ejecutada o determinados misiles que han superado fases específicas de vuelo pueden escapar al control efectivo de quienes inicialmente autorizaron su empleo.

Entonces, ¿Qué ocurre cuando la información cambia después del lanzamiento del arma y ya no existe una capacidad real de corregir el resultado?

Casos como los ataques contra los puentes de Lužane y Grdelica[3] durante la campaña aérea de la OTAN en Yugoslavia evidencian la complejidad de las decisiones adoptadas en

[3] Un ejemplo frecuentemente citado en la literatura sobre conducción de hostilidades corresponde a los ataques realizados por la OTAN contra los puentes de Grdelica y Lužane durante la campaña aérea en la República Federal de Yugoslavia en 1999. En ambos casos, los puentes eran considerados objetivos militares; sin embargo, la presencia de medios de transporte civiles al momento de la ejecución produjo numerosas víctimas entre la población civil, generando posteriores debates sobre la aplicación de los principios de proporcionalidad y precaución. Véase: Amnesty International, NATO/Federal Republic of Yugoslavia: "Collateral Damage" or Unlawful Killings? Violations of the Laws of War by NATO During Operation Allied Force, Londres, Amnesty International Publications, 2000.

entornos operacionales caracterizados por información incompleta y cambios constantes en la situación táctica. En ambos supuestos, los puentes constituían objetivos militares desde la perspectiva de su utilización estratégica; sin embargo, la presencia inesperada de medios de transporte civiles durante la ejecución de los ataques produjo consecuencias humanitarias significativas.

Estos episodios ilustran cómo una decisión inicialmente compatible con los criterios de selección de objetivos puede evolucionar rápidamente hacia escenarios en los cuales la capacidad de corrección resulta limitada o inexistente; dicho de otra manera, cómo un objetivo militar legítimo puede verse afectado por cambios imprevistos en el entorno operacional una vez iniciada la secuencia de ataque. Precisamente por ello, la evaluación previa del riesgo, la calidad de la inteligencia disponible y la preservación de mecanismos efectivos de supervisión humana adquieren una importancia decisiva en la prevención de daños a la población civil.

En estos escenarios, la discusión jurídica deja de centrarse exclusivamente en la decisión misma y se traslada a la suficiencia de las medidas preventivas adoptadas antes de alcanzar el punto de no retorno. Cuanto menor sea la capacidad de corregir una decisión una vez iniciada la operación, mayor deberá ser el nivel de diligencia exigido durante las etapas previas de planificación, validación y supervisión.

Desde esta perspectiva, el riesgo algorítmico alcanza su máxima expresión cuando la decisión operacional entra en una fase de irreversibilidad en la que ya no existen mecanismos efectivos para corregir los efectos de un error de identificación, clasificación o selección del objetivo. En tales circunstancias, la discusión jurídica no puede limitarse a la atribución posterior de responsabilidades, sino que debe examinar si los riesgos previsibles fueron adecuadamente identificados, evaluados y mitigados antes de que la operación alcanzara el umbral a partir del cual sus consecuencias dejaron de ser gobernables.

Dicho lo anterior, encontramos que el principal desafío jurídico de la inteligencia artificial militar no consiste únicamente en determinar quién toma la decisión inicial, sino en establecer hasta qué punto dicha decisión continúa siendo gobernable una vez activado el sistema de armas. Precisamente por ello, la preservación del control humano significativo y la evaluación rigurosa de los riesgos previsibles constituyen exigencias indispensables para garantizar la compatibilidad de estas tecnologías con los principios fundamentales del Derecho Internacional Humanitario.

5.5. El artículo 36 del Protocolo Adicional I y la evaluación preventiva de nuevas tecnologías

El artículo 36 del Protocolo Adicional I de 1977 establece que los Estados tienen la obligación de determinar si el empleo de nuevas armas, medios o métodos de guerra estaría prohibido por el Derecho Internacional aplicable. Su importancia respecto a la inteligencia artificial militar radica en que traslada el análisis jurídico a una etapa previa al empleo efectivo de la tecnología. Antes de su utilización operacional, los Estados deben evaluar los riesgos asociados al sistema y verificar su compatibilidad con las normas aplicables.

Desde la perspectiva desarrollada en este artículo, el artículo 36 representa una manifestación concreta de la lógica preventiva que inspira tanto la teoría jurídica del riesgo como los principios fundamentales del Derecho Internacional Humanitario. Su finalidad consiste precisamente en evitar que riesgos previsibles se transformen en daños incompatibles con las exigencias de humanidad que orientan la conducción de las hostilidades. Así las cosas, la referencia a las exigencias de humanidad resulta coherente con la Cláusula de Martens, incorporada al artículo 1.2 del Protocolo Adicional I de 1977, la cual recuerda que incluso frente a desarrollos tecnológicos no previstos expresamente por el derecho convencional, la

conducción de las hostilidades permanece sometida a los principios de humanidad y a los dictados de la conciencia pública.

6. Derechos Humanos, Dignidad Humana y regulación jurídica del riesgo algorítmico en los conflictos armados

La regulación jurídica de los sistemas de inteligencia artificial utilizados en conflictos armados no puede agotarse en las disposiciones del Derecho Internacional Humanitario. Aunque este régimen constituye el marco normativo especializado para la conducción de las hostilidades, la protección de la persona humana continúa encontrando fundamento en los instrumentos internacionales de derechos humanos, cuya aplicación no desaparece por la sola existencia de un conflicto armado.

Referente a ello, la Corte Internacional de Justicia ha sostenido de manera reiterada que la protección derivada del Derecho Internacional de los Derechos Humanos continúa siendo aplicable durante los conflictos armados. En su Opinión Consultiva sobre la Licitud de la Amenaza o el Empleo de Armas Nucleares de 1996[4], la Corte afirmó que la protección del derecho a la vida no cesa en tiempo de guerra, aunque la determinación de una privación arbitraria de la vida debe apreciarse a la luz del Derecho Internacional Humanitario como régimen jurídico especial.

Posteriormente, en la Opinión Consultiva sobre las Consecuencias Jurídicas de la Construcción de un Muro en el Territorio Palestino Ocupado de 2004[5], reiteró que los instrumentos internacionales de derechos humanos continúan aplicándose en situaciones de conflicto armado junto con las normas del Derecho Internacional Humanitario. Esta aproximación ha permitido consolidar una visión complementaria en la que ambos regímenes concurren en la protección de la vida, la integridad personal y la dignidad humana.

Desde esta perspectiva, la incorporación de sistemas de inteligencia artificial en operaciones militares plantea desafíos que trascienden la legalidad de los medios y métodos de guerra. La cuestión central consiste en determinar cómo garantizar que las decisiones apoyadas por algoritmos continúen respetando los derechos fundamentales de las personas potencialmente afectadas por su utilización.

La dignidad humana ocupa una posición central dentro de este análisis. Como fundamento axiológico de los sistemas contemporáneos de protección de los derechos humanos, la dignidad exige que toda persona sea considerada como un fin en sí misma y no como un simple objeto de decisiones automatizadas. Principio sine qua non cuando las tecnologías participan en procesos relacionados con la identificación de objetivos, la evaluación de amenazas o el empleo de la fuerza.

La creciente sofisticación de los algoritmos no elimina la necesidad de preservar espacios de valoración humana. Las decisiones susceptibles de producir consecuencias irreversibles sobre la vida o la integridad personal no pueden ser analizadas exclusivamente desde criterios de eficiencia tecnológica. Por el contrario, deben permanecer sometidas a controles capaces de incorporar consideraciones jurídicas, éticas y

[4]Corte Internacional de Justicia, Opinión Consultiva sobre la Licitud de la Amenaza o el Empleo de Armas Nucleares, 8 de julio de 1996, Recueil (ICJ Reports), 1996, p. 240, párr. 25.

[5] Corte Internacional de Justicia, Opinión Consultiva sobre las Consecuencias Jurídicas de la Construcción de un Muro en el Territorio Palestino Ocupado, 9 de julio de 2004, Recueil (ICJ Reports), 2004, p. 178, párrs. 105-106.

humanitarias que exceden la lógica matemática de los sistemas automatizados.

El derecho a la vida constituye otro de los ejes fundamentales de esta discusión. Reconocido en múltiples instrumentos internacionales, este derecho impone a los Estados no solo obligaciones negativas de abstención, sino también deberes positivos de protección. Tales deberes adquieren una importancia particular cuando las autoridades conocen o pueden conocer la existencia de riesgos susceptibles de afectar bienes jurídicos fundamentales.

La teoría del riesgo desarrollada en los apartados anteriores permite comprender esta obligación desde una perspectiva preventiva. Cuando un riesgo resulta identificable y razonablemente previsible, surge el deber jurídico de adoptar medidas destinadas a reducir la probabilidad de que dicho riesgo se materialice en daño. Esta lógica resulta plenamente aplicable a los sistemas de inteligencia artificial de alto riesgo utilizados en contextos militares.

La igualdad y la prohibición de discriminación también plantean desafíos significativos frente a la utilización de sistemas de inteligencia artificial. Diversos estudios han advertido que los algoritmos pueden reproducir e incluso amplificar sesgos presentes en los datos utilizados para su entrenamiento. En este sentido, Buolamwini y Gebru[6] (2018) demostraron que varios sistemas comerciales de reconocimiento facial presentaban diferencias significativas de precisión según el género y el color de piel de las personas evaluadas, evidenciando cómo las deficiencias en los conjuntos de datos pueden traducirse en resultados discriminatorios. De manera similar, O'Neil[7] (2016) ha señalado que los sistemas algorítmicos aparentemente neutrales pueden perpetuar desigualdades preexistentes cuando operan sobre datos sesgados o bajo criterios insuficientemente transparentes.

Por ello, si los sesgos ya han sido demostrados en sistemas civiles de reconocimiento facial, la posibilidad de errores discriminatorios adquiere una dimensión aún más sensible cuando tecnologías similares participan en procesos de identificación, vigilancia o selección de objetivos en contextos de conflicto armado.

La posibilidad de que errores sistemáticos afecten de manera desproporcionada a determinados grupos poblacionales exige mecanismos rigurosos de supervisión y validación. Desde una perspectiva de derechos humanos, la neutralidad tecnológica no puede presumirse. Corresponde a quienes desarrollan, implementan y utilizan estos sistemas demostrar que los riesgos asociados han sido adecuadamente identificados y mitigados.

En este contexto cobra importancia la construcción de mecanismos jurídicos e institucionales orientados a la gestión del riesgo algorítmico. Dichos mecanismos no se limitan a la existencia de normas jurídicas formales, sino que comprenden estructuras de supervisión, evaluación y control destinadas a prevenir daños previsibles.

El control del riesgo algorítmico se encuentra estrechamente vinculado con la posición de garante asumida por los Estados. El conocimiento de riesgos previsibles genera obligaciones de actuación. De esta manera, la

[6] Buolamwini y Gebru demostraron que diversos sistemas comerciales de reconocimiento facial presentaban diferencias significativas de precisión según el género y el color de piel de las personas evaluadas, evidenciando la capacidad de los algoritmos para reproducir sesgos presentes en los datos de entrenamiento. Véase: Joy Buolamwini y Timnit Gebru, «Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification», Proceedings of Machine Learning Research, vol. 81, 201.

[7] Cathy O'Neil, Armas de destrucción matemática: cómo el Big Data aumenta la desigualdad y amenaza la democracia, Nueva York, Crown Publishing Group, 2016.

ausencia de regulación específica no puede interpretarse como una autorización para ignorar riesgos identificables. Por el contrario, cuanto mayor sea la capacidad potencial de una tecnología para afectar derechos fundamentales, mayor deberá ser el nivel de diligencia exigido a las autoridades responsables de su control.

Desde esta perspectiva, la inteligencia artificial militar plantea un desafío que trasciende el debate tecnológico. Se trata, en esencia, de una cuestión relacionada con la capacidad del derecho para preservar la prioridad de la persona humana frente a sistemas cada vez más complejos de toma de decisiones. La protección efectiva de la dignidad humana exige que la innovación tecnológica permanezca subordinada a principios jurídicos capaces de garantizar que el progreso científico no se traduzca en nuevas formas de vulnerabilidad.

Por ello, la regulación de los sistemas de inteligencia artificial de alto riesgo empleados en conflictos armados debe concebirse como una obligación derivada no solo del Derecho Internacional Humanitario, sino también de los compromisos asumidos por los Estados en materia de derechos humanos. La prevención del daño, la reducción de riesgos previsibles y la protección de la dignidad humana constituyen objetivos comunes que justifican la adopción de mecanismos regulatorios orientados a una gobernanza responsable de la inteligencia artificial en escenarios de conflicto.

Finalmente, la protección de la dignidad humana frente a los riesgos asociados a la inteligencia artificial no constituye únicamente una preocupación futura ni una cuestión dependiente de eventuales desarrollos normativos internacionales. La existencia de riesgos identificables y amenazas previsibles para los derechos fundamentales exige respuestas jurídicas orientadas a la prevención, especialmente cuando se trata de tecnologías capaces de incidir sobre la vida, la integridad personal, la libertad y otros bienes jurídicos esenciales.

Así las cosas, los derechos humanos ofrecen un marco normativo que permite abordar los desafíos planteados por la inteligencia artificial antes de que los riesgos se materialicen en daños efectivos. La condición primordial del ser humano exige que la innovación tecnológica permanezca sometida a límites jurídicos compatibles con la dignidad humana y con las obligaciones de protección que corresponden a los Estados y demás actores involucrados en el desarrollo y utilización de estas tecnologías.

7. La obligación estatal de regular los sistemas de inteligencia artificial de alto riesgo

Precisamente porque existen riesgos identificables y amenazas previsibles para los derechos fundamentales, surge la necesidad de examinar la obligación estatal de regular los sistemas de inteligencia artificial de alto riesgo.

La velocidad con la que evolucionan los sistemas de inteligencia artificial contrasta con la lentitud que caracteriza tradicionalmente los procesos de regulación jurídica internacional. Mientras los desarrollos tecnológicos avanzan de manera acelerada, la construcción de consensos normativos globales suele requerir largos periodos de negociación y adaptación institucional. Esta circunstancia plantea una cuestión fundamental: ¿deben los Estados esperar la consolidación de un régimen internacional específico para regular los riesgos asociados a la inteligencia artificial militar o, por el contrario, poseen obligaciones jurídicas inmediatas derivadas de los principios generales de protección y prevención?

La respuesta propuesta en este trabajo parte de una premisa sencilla. La ausencia de una regulación internacional exhaustiva no implica la inexistencia de deberes jurídicos. Por el contrario, cuando los riesgos son identificables y potencialmente capaces de afectar bienes jurídicos fundamentales, los Estados conservan la obligación de adoptar medidas razonables destinadas a prevenir,

controlar y mitigar sus posibles consecuencias.

Esta obligación encuentra fundamento en diversos principios jurídicos. En primer lugar, el principio de prevención exige actuar frente a riesgos previsibles antes de que estos se materialicen en daños efectivos. A diferencia de los modelos clásicos centrados exclusivamente en la reparación posterior, las tendencias contemporáneas del derecho han reconocido que la protección efectiva de los derechos fundamentales requiere mecanismos de actuación anticipada.

En segundo lugar, el principio de precaución adquiere preeminencia especial frente a tecnologías emergentes cuyas consecuencias completas aún no son plenamente conocidas. Cuando existe incertidumbre científica sobre la magnitud de los riesgos potenciales, la ausencia de certeza absoluta no puede ser utilizada como argumento para justificar la inacción estatal. Por el contrario, dicha incertidumbre refuerza la necesidad de adoptar mecanismos preventivos de control.

La seguridad no implica la eliminación absoluta de todos los riesgos ni la supresión total de las pérdidas potenciales. La propia dinámica de los conflictos armados demuestra que el riesgo forma parte inherente de las operaciones militares. Sin embargo, reconocer la existencia de riesgos no equivale a aceptar pasivamente sus consecuencias. La función del derecho consiste precisamente en identificar, administrar y reducir aquellos riesgos que puedan afectar bienes jurídicos especialmente protegidos.

La lógica preventiva que inspira la regulación de los sistemas de inteligencia artificial resulta coherente con la propia concepción contemporánea de la seguridad. En el ordenamiento jurídico colombiano, diversas disposiciones parten del reconocimiento de riesgos y amenazas susceptibles de afectar bienes jurídicamente protegidos, circunstancia que justifica la adopción de medidas anticipadas de protección. En el ámbito militar, esta aproximación puede advertirse en la definición de seguridad prevista por la Ley 1862 de 2017, la cual presupone la existencia de factores capaces de afectar personas, capacidades, operaciones o instalaciones, imponiendo la necesidad de desarrollar mecanismos orientados a prevenir o reducir tales afectaciones antes de que se materialicen en daños efectivos.

Desde la inteligencia artificial militar, esta obligación adquiere una importancia particular. Los sistemas de alto riesgo no constituyen fenómenos inevitables ni amenazas naturales. Se trata de tecnologías diseñadas, desarrolladas, entrenadas y desplegadas mediante decisiones humanas. En consecuencia, los riesgos asociados a su utilización también son susceptibles de evaluación, supervisión y control.

La ausencia de un instrumento internacional específico que regule integralmente la inteligencia artificial en contextos de conflicto armado no debe interpretarse como un obstáculo para la adopción de medidas jurídicas de prevención. Por el contrario, la identificación de riesgos previsibles asociados al empleo de estas tecnologías refuerza la necesidad de que los Estados ejerzan de manera activa sus competencias regulatorias internas. La protección de los derechos humanos no puede depender exclusivamente de la futura evolución del derecho internacional convencional. Los Estados conservan amplios márgenes de actuación para desarrollar marcos normativos orientados a la identificación, evaluación y mitigación de riesgos tecnológicos capaces de afectar bienes jurídicos fundamentales.

La potestad de configuración legislativa, en el caso colombiano, permite adoptar mecanismos de supervisión, control, transparencia y responsabilidad antes de que los riesgos identificados se materialicen en daños efectivos. Esta aproximación resulta coherente con los principios de prevención, debida diligencia y protección efectiva de los derechos humanos, particularmente cuando se trata de tecnologías cuyo potencial lesivo

puede alcanzar dimensiones significativas en escenarios de conflicto armado.

En consecuencia, la regulación jurídica de la inteligencia artificial no debe concebirse únicamente como una tarea futura del derecho internacional, sino también como una responsabilidad inmediata de los ordenamientos jurídicos nacionales frente a riesgos y amenazas cuya existencia ya resulta identificable en el contexto contemporáneo. Entre las medidas que podrían ser adoptadas destacan la exigencia de evaluaciones previas de impacto, la realización de auditorías técnicas independientes, la obligación de documentar los procesos de entrenamiento algorítmico, la implementación de sistemas de trazabilidad de decisiones automatizadas y el establecimiento de mecanismos permanentes de supervisión humana significativa.

La creciente autonomía tecnológica no debe conducir a la desaparición de la responsabilidad humana. Por el contrario, cuanto mayor sea la capacidad de una tecnología para producir consecuencias letales o afectar derechos fundamentales, mayor deberá ser la intervención humana en los procesos de autorización, supervisión y control.

Es decir, la preservación del control humano es vital respecto a decisiones relacionadas con el uso de la fuerza. Asimismo, los Estados deberían desarrollar procedimientos especializados de revisión jurídica inspirados en la lógica contenida en el artículo 36 del Protocolo Adicional I de 1977. Tales procedimientos permitirían evaluar de manera anticipada la compatibilidad de los sistemas de inteligencia artificial con las obligaciones derivadas del Derecho Internacional Humanitario y de los Derechos Humanos antes de su incorporación a operaciones reales.

El control jurídico del riesgo algorítmico exige igualmente mecanismos institucionales capaces de distribuir responsabilidades entre los distintos actores involucrados en el ciclo de vida de estas tecnologías. Los desarrolladores, fabricantes, autoridades militares, organismos reguladores y operadores participan en diferentes niveles de generación y control del riesgo. Por ello, la prevención efectiva requiere un modelo de responsabilidades compartidas orientado a evitar vacíos de protección. Adicionalmente, dicho control jurídico de los sistemas de inteligencia artificial de alto riesgo debe orientarse hacia la construcción de modelos regulatorios preventivos capaces de equilibrar la innovación tecnológica con la protección efectiva de la persona humana. Solo de esta manera será posible garantizar que los avances tecnológicos contribuyan a reducir los efectos de la violencia armada sin convertirse en nuevas fuentes de riesgo para quienes el derecho internacional pretende proteger.

En definitiva, la regulación de la inteligencia artificial militar no constituye una opción política discrecional, sino una consecuencia lógica de los deberes de protección que recaen sobre los Estados. La identificación de riesgos previsibles genera obligaciones correlativas de prevención y mitigación. Esperar la materialización del daño para actuar resultaría incompatible con los principios de precaución, prevención y protección de la dignidad humana que inspiran el orden jurídico contemporáneo.

8. Conclusiones

La expansión de los sistemas de inteligencia artificial en escenarios de conflicto armado constituye uno de los desafíos jurídicos más relevantes de la contemporaneidad. Lejos de representar únicamente una innovación tecnológica, estas herramientas plantean interrogantes fundamentales sobre la capacidad del derecho para anticipar, gestionar y controlar riesgos asociados a decisiones capaces de afectar la vida, la integridad personal y la protección de la población civil.

Los sistemas de inteligencia artificial utilizados en contextos de conflicto armado constituyen tecnologías de alto riesgo cuyas consecuencias jurídicas no pueden esperar a

la consolidación de un régimen internacional específico, sino que exigen respuestas regulatorias estatales inmediatas basadas en los principios preventivos derivados del Derecho Internacional Humanitario, los Derechos Humanos y la gestión jurídica del riesgo.

El análisis desarrollado permitió evidenciar que la teoría jurídica del riesgo ofrece un marco conceptual especialmente adecuado para abordar los desafíos planteados por la inteligencia artificial militar. A diferencia de los enfoques centrados exclusivamente en la responsabilidad posterior al daño, la lógica del riesgo desplaza la atención hacia la identificación temprana de amenazas previsibles y hacia la adopción de medidas orientadas a prevenir su materialización. Desde esta perspectiva, la regulación jurídica no debe esperar a la producción efectiva del perjuicio para intervenir, sino actuar desde las fases iniciales de diseño, programación, entrenamiento, validación y despliegue de los sistemas tecnológicos.

La investigación permitió igualmente establecer que los sistemas de inteligencia artificial utilizados en contextos militares constituyen una manifestación contemporánea de los riesgos manufacturados descritos por Ulrich Beck. Al tratarse de tecnologías creadas por decisiones humanas, los riesgos asociados a su funcionamiento no pueden ser considerados inevitables ni ajenos al control jurídico. Por el contrario, su origen tecnológico refuerza la necesidad de desarrollar mecanismos institucionales orientados a su evaluación, supervisión y mitigación.

Asimismo, el estudio de la niebla de guerra y de la fricción desarrolladas por Carl von Clausewitz, complementadas por las reflexiones derivadas de la experiencia de Robert McNamara durante la guerra de Vietnam, permitió advertir que la acumulación masiva de información no elimina necesariamente la incertidumbre inherente a los conflictos armados. La inteligencia artificial constituye, sin duda, el intento más sofisticado desarrollado hasta ahora para reducir dicha incertidumbre; sin embargo, la tecnología no suprime la complejidad humana, jurídica y operacional de la guerra. Más bien, transforma parte de la incertidumbre tradicional en nuevas formas de riesgo asociadas al funcionamiento de los propios algoritmos.

Desde la perspectiva del Derecho Internacional Humanitario, se constató que los principios de distinción, proporcionalidad y precaución continúan siendo plenamente aplicables frente a los sistemas de inteligencia artificial de alto riesgo. La incorporación de nuevas tecnologías no modifica la obligación de proteger a la población civil ni autoriza la delegación irrestricta de decisiones relacionadas con el uso de la fuerza. Por el contrario, cuanto mayor sea la capacidad de una tecnología para influir en la conducción de las hostilidades, mayor deberá ser el nivel de supervisión jurídica y humana exigido para su utilización.

La investigación también permitió demostrar que la protección de la persona humana en contextos de conflicto armado no depende exclusivamente del Derecho Internacional Humanitario. Los Derechos Humanos continúan desempeñando una función complementaria esencial, particularmente en lo relacionado con la dignidad humana, el derecho a la vida, la igualdad y las obligaciones positivas de protección que recaen sobre los Estados. En consecuencia, la regulación jurídica de la inteligencia artificial militar debe construirse a partir de una interpretación armónica de ambos regímenes jurídicos.

Uno de los hallazgos más relevantes del presente trabajo consiste en reconocer que la identificación de riesgos previsibles activa deberes jurídicos de prevención derivados de la posición de garante asumida por los Estados y por los distintos actores que participan en el ciclo de vida de estas tecnologías. Los desarrolladores, fabricantes, autoridades militares y organismos reguladores intervienen en diferentes etapas de generación y control del riesgo algorítmico,

razón por la cual la protección efectiva exige modelos de responsabilidad preventiva y no únicamente esquemas de atribución posterior del daño.

Finalmente, se concluye que la ausencia de un régimen internacional específico y universal sobre inteligencia artificial militar no exonera a los Estados de sus obligaciones jurídicas. La potestad de configuración legislativa, los principios de prevención y precaución, así como los compromisos asumidos en materia de Derecho Internacional Humanitario y Derechos Humanos, proporcionan fundamentos suficientes para adoptar medidas regulatorias orientadas a la identificación, administración y mitigación de riesgos asociados a estas tecnologías.

En consecuencia, la cuestión jurídica fundamental no consiste en determinar si la inteligencia artificial puede eliminar la niebla de guerra, sino en establecer cómo dirigir responsablemente los riesgos que ella misma genera. La verdadera innovación normativa del siglo XXI no radicará únicamente en el desarrollo de tecnologías cada vez más sofisticadas, sino en la capacidad de los sistemas jurídicos para garantizar que dicho progreso permanezca subordinado a la protección de la dignidad humana y a los principios fundamentales que limitan el ejercicio de la fuerza en los conflictos armados.

Referencias

Amnesty International. NATO/Federal (2000). Republic of Yugoslavia: "Collateral Damage" or Unlawful Killings? Violations of the Laws of War by NATO During Operation Allied Force. Londres: Amnesty International Publications,.

BECK, Ulrich. (1998). La sociedad del riesgo: hacia una nueva modernidad. Barcelona: Paidós, 1998.

BUOLAMWINI, Joy & GEBRU, Timnit. (2018). "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification". Proceedings of Machine Learning Research, vol. 81,.

CLAUSEWITZ, Carl von. (2005). De la guerra. Madrid: La Esfera de los Libros,.

Convenios de Ginebra de 12 de agosto de 1949.

Comité Internacional de la Cruz Roja. Derecho Internacional Humanitario Consuetudinario. Volumen I: Normas. Buenos Aires: CICR, 2007.

Comité Internacional de la Cruz Roja. Guía para el examen jurídico de nuevas armas, medios y métodos de guerra. Ginebra: CICR.

Corte Constitucional de Colombia. Sentencia T-399 de 2024. M.P. Gloria Stella Ortiz Delgado.

Corte Constitucional de Colombia. Sentencia T-224 de 2014. M.P. Luis Ernesto Vargas Silva.

Corte Constitucional de Colombia. Sentencia SU – 1184 de 2001, M.P. Eduardo Montealegre Lynett.

Corte Internacional de Justicia. Opinión Consultiva sobre la Licitud de la Amenaza o el Empleo de Armas Nucleares, 8 de julio de 1996.

Corte Internacional de Justicia. Opinión Consultiva sobre las Consecuencias Jurídicas de la Construcción de un Muro en el Territorio Palestino Ocupado, 9 de julio de 2004.

Corte Suprema de Justicia, Sala de Casación Penal. Radicado 25.536 del 27 de julio de 2006. M.P. Yesid Ramírez Bastidas.

DINSTEIN, Yoram. (2016). Conducta de las hostilidades conforme al Derecho Internacional de los Conflictos Armados. Comité Internacional de la Cruz Roja.

FERNANDEZ RODRIGUEZ, Juan. Carlos. & LIZ RIVAS, Lenny. (2023). El terrorismo: aspectos psicosociales en el proceso de radicalización. **"**Repercusiones de la radicalización yihadista en la seguridad Europea, Mediterránea y Latinoamericana", Thomson Reuters Aranzadi. Pp. 270-287. https://doi.org/10.5281/zenodo.14540821

GANGI GUILLEN, Giuseppe. Kodjack. VII. (2025). United States' strategic shift and

International Humanitarian Law: Implications for the Russia-Ukraine war. *Revista Científica General José María Córdova*, *23*(49), 299–315. https://doi.org/10.21830/19006586.1467

GANGI GUILLÉN, Giuseppe. Kodjack. VII. (2023). Dinámicas migratorias en la frontera colombo venezolana y su relación con la criminalidad transnacional. *Revista Científica General José María Córdova,* 21(44), 907-924. https://doi.org/10.21830/19006586.984

GANGI GUILLÉN, Giuseppe. Kodjack. VII. & Delgado Morán, Juan José. (2025). Derechos Humanos y Terrorismo. El terrorismo en Europa: la salvaguarda de la seguridad y la protección de las víctimas. Ed. Sindéresis. ISBN: 979-13-87929-25-1

GANGI GUILLÉN, Giuseppe. Kodjack. VII. (2026). Innovación en la enseñanza del Derecho Internacional Humanitario: simulaciones para asesores jurídicos operacionales. "Revolución en las aulas: cómo la innovación está reescribiendo la universidad". Ed. Colex. https://doi.org/10.69592/979-13-7011-516-6-CAP-6

GANGI GUILLÉN, Giuseppe. Kodjack. VII. (2023). An exploration of socio-cultural and liguistic issues for a sustainable migration in the global north. Aranzadi

GANGI GUILLÉN, Giuseppe. Kodjack. (2025). Derechos humanos y derecho penal en la era de la inteligencia artificial: retos y propuestas. *Cuadernos de RES PUBLICA en derecho y criminología*, https://doi.org/10.46661/respublica.11635

GIDDENS, Anthony. (2000). Un mundo desbocado: los efectos de la globalización en nuestras vidas. Madrid: Taurus,.

LIZ RIVAS, Lenny. (2018). Algunas bases neurológicas sobre la violencia y la agresión;"Conflictos y diplomacia, desarrollo y paz, globalización y medio ambiente". Thomson Reuters/Aranzadi, pp. 943-955. https://doi.org/10.5281/zenodo.14559664

Ley 599 de 2000. Código Penal Colombiano.

Ley 1862 de 2017. Código Penal Militar.

MARTINO, L. (2023). La guerra nel XXI secolo: la dimensione cyber e il conflitto russo-ucraino. In: *La guerra tiepida: Il conflitto ucraino e il futuro dei rapporti tra Russia e Occidente*. Rome: Luiss University Press.

MCNAMARA, Robert S. (1995). In Retrospect: The Tragedy and Lessons of Vietnam. Nueva York: Times Books,.

MAZURIER, Pablo. A., & PAYÁ SANTOS, Claudio. Augusto. (2018). *Amenazas híbridas: teoría de la hibridez y nuevo orden internacional.* Thomson Reuters Aranzadi.

O'NEIL, Cathy. (2016). Armas de destrucción matemática: cómo el Big Data aumenta la desigualdad y amenaza la democracia. Nueva York: Crown Publishing Group,.

Organización de las Naciones Unidas. Pacto Internacional de Derechos Civiles y Políticos, 1966.

PAYÁ SANTOS, Claudio. Augusto; RODRÍGUEZ GONZÁLEZ, Víctor; DOMÍNGUEZ PINEDA, Neidy. Z; DIZ CASAL, Javier; FERNÁNDEZ RODRÍGUEZ, Juan. Carlos. & DELGADO MORÁN, Juan. José. (2025). Role of the Human Factor in the Cybersecurity Ecosystem. *Journal of Information Systems Engineering and Management*,10(4). https://doi.org/10.52783/jisem.v10i4.8983

Protocolo Adicional I a los Convenios de Ginebra de 1949 relativo a la protección de las víctimas de los conflictos armados internacionales, 1977.

SASSÒLI, Marco. (2019). Derecho Internacional Humanitario: reglas, controversias y soluciones a los problemas planteados por la guerra. Ginebra: Academia de Derecho Internacional Humanitario y Derechos Humanos.

Unión Europea. Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (AI Act), 13 de junio de 2024.