Marcos Antonio ARAVENA FLORES
Universidad Autónoma de Chile (Chile)
marcos.aravena1@cloud.uautonoma.cl
ORCID: https://orcid.org/0009-0006-0262-7618

Recibido: 19/08/2024 - Aceptado: 21/10/2024

Para citar este artículo / To reference this article / Para citar este artigo:
Aravena Flores, Marcos Antonio. (2024). Los principios éticos y jurídicos aplicables a los sistemas autónomos de armas letales. *Revista de Derecho, 23*(46), Artículo e463. https://doi.org/10.47274/DERUM/46.3

Los principios éticos y jurídicos aplicables a los sistemas autónomos de armas letales

Resumen: la inteligencia artificial (IA) es una tecnología que ha dado origen a la creación de nuevas armas militares, capaces de seleccionar y atacar a sus objetivos con mayor autonomía, pudiendo prescindir del control humano. Frente al impacto de dicha tecnología en la sociedad, el presente artículo tiene como objetivo comprender los principios éticos que en la actualidad están orientando su desarrollo y uso, y a la vez, los principios legales provenientes del derecho internacional humanitario (DIH) que son susceptibles de aplicarse, lo cual permite pronosticar los limites necesarios a imponer ante los potenciales riesgos que implica esta tecnología. El texto presenta una investigación cualitativa que utilizó preferentemente una revisión de tipo documental que permitió analizar ciertos principios éticos prevalentes reconocidos por determinados Estados y de los principios legales reconocidos ampliamente por la literatura existente.

Palabras clave: Derecho internacional humanitario; Inteligencia artificial; Principios éticos; Sistemas autónomos de armas letales.

Ethical and legal limits applicable to autonomous lethal weapon systems

Abstract: artificial intelligence (AI) is a technology that has given rise to the creation of new military weapons, capable of selecting and attacking their targets with greater autonomy, being able to dispense with human control. Faced with the impact of this technology on society, this article aims to understand the ethical principles that are currently guiding its development and use, and at the same time, the legal principles from international humanitarian law (IHL) that are likely to apply, which allows predicting the necessary limits to be imposed in the face of the potential risks involved in this technology. The text presents qualitative research that used preferably a documentary type review that allowed the analysis of certain prevailing ethical principles recognized by certain States and of the legal principles widely recognized by the existing literature.

Keywords: International humanitarian law; Artificial intelligence; Ethical principles; Lethal autonomous weapon systems.

Princípios éticos e jurídicos aplicáveis aos sistemas autónomos de armas letais

Resumo: A inteligência artificial (IA) é uma tecnologia que tem dado origem à criação de novas armas militares, capazes de selecionar e atacar os seus alvos com maior autonomia, podendo dispensar o controlo humano. Perante o impacto desta tecnologia na sociedade, este artigo tem como objetivo compreender os princípios éticos que estão atualmente a orientar o seu desenvolvimento e utilização e, simultaneamente, os princípios legais do Direito Internacional Humanitário (DIH) que são susceptíveis de ser aplicados, o que nos permite prever os limites necessários a impor face aos potenciais riscos envolvidos nesta tecnologia. O texto apresenta uma investigação qualitativa que recorreu preferencialmente a uma análise de tipo documental que permitiu analisar certos princípios éticos dominantes reconhecidos por certos Estados e os princípios jurídicos amplamente reconhecidos pela literatura existente.

Palavras-chave: Direito internacional humanitário; Inteligência artificial; Princípios éticos; Sistemas de armas letais autónomas.

1. Introducción

La IA es un concepto acuñado en el área de las tecnologías de la información, que se ha discutido en casi todos los campos de las ciencias en las últimas décadas debido a los efectos disruptivos que se han manifestado en su fabricación, implementación y sus distintos usos (Saria y Celik, 2021). Si bien no existe una definición universalmente aceptada, en términos generales se ha precisado que consiste en una tecnología que incluye una variedad de técnicas computacionales y procesos asociados (Raso et al., 2018), siendo un sistema que "piensa, actúa y tiene su propia racionalidad" (Čerka et al, 2017, p.317).

Respecto de esta tecnología, es frecuente afirmar que las características de la IA están progresando con la gestión de la innovación y el despliegue de los recursos en tecnología. Sin embargo, a la luz de los efectos disruptivos que la IA ha demostrado ser capaz de causar, se ha comenzado a manifestar una convergencia global en torno a determinar los valores y principios que deben orientarla. En atención a los efectos que tiene esta tecnología y que se espera que siga teniendo en distintas áreas, se ha incrementado las advertencias sobre los riesgos y cuestionamientos que su implementación traerá consigo (Azuaje Pirela y Finol González, 2023).

Ante la mayor autonomía que han alcanzado las máquinas dotadas de IA, surgen interrogantes con respecto a su posible impacto dañino en las sociedades humanas. Esta incertidumbre se debe a la falta de claridad sobre la dirección que seguirá esta tecnología y a qué velocidad evolucionará, principalmente si se considera el alcance del aprendizaje automático (o *machine learning*) [1] y el aprendizaje profundo (o *deep learning*)[2]. Estos avances permiten que los sistemas de IA completen con éxito muchas tareas, como la detección de objetos en las imágenes, la comprensión de los idiomas y el procesamiento de señales de voz (Samek et al, 2017), el reconocimiento e identificación de las personas, y la seguridad privada o nacional. Basado en los logros tecnológicos de la IA hasta la fecha, estos desarrollos comenzaron a tener grandes impactos en la industria de la defensa a principios del siglo XXI (Saria y Celik, 2021).

Particularmente, el uso militar o de defensa de la IA ha despertado mayor preocupación a nivel mundial. Por un lado, no existe hasta el momento claridad en cuanto a los límites que deben aplicarse a ella; y, por otro, los impactos nocivos que pudiesen derivarse del desarrollo de esta tecnología, principalmente por el surgimiento de armamentos complejos y avanzados. Estos sistemas, actualmente denominados como Sistemas de Armas Autónomos Letales (SAAL), se refieren a aquellas armas que, al ser programadas con IA, cuentan con la capacidad para tomar decisiones referentes al uso de la fuerza en el campo de batalla sin necesidad de intervención humana directa, logrando así seleccionar y atacar objetivos militares de manera autónoma (López Jacoiste, 2020).

Frente al progresivo desarrollo tecnológico de estas armas autónomas, en un contexto geopolítico y multipolar complejo, con diferentes intereses entre los estados con capacidad

1 El aprendizaje automático se define como "un subconjunto de la IA que consiste en entrenar algoritmos para reconocer patrones en datos". En el ámbito de defensa o militar, el aprendizaje automático se ha utilizado para mejorar la precisión del reconocimiento de objetos, el seguimiento de objetos y la detección de obstáculos (Soori et al., 2023, p.61).

2 El aprendizaje profundo es un "subconjunto del aprendizaje automático que procesa cantidades masivas de datos utilizando redes neuronales". Por medio de este, por ejemplo, un dron puede llevar a cabo tareas difíciles como la navegación autónoma y la cartografía, permitiéndoles detectar y evitar obstáculos en tiempo real (Soori et al., 2023, p.61).

tecnológica, ha surgido una creciente incertidumbre y riesgo respecto a la regulación y uso de esta tecnología. Por ello, el presente artículo tuvo como objetivo identificar y analizar los límites aplicables al desarrollo y utilización de los SAAL en el ámbito militar, todo ello a la luz de los lineamientos éticos y legales del derecho internacional humanitario (DIH).

Metodológicamente, el artículo se construyó a partir de un enfoque cualitativo y descriptivo (de tipo documental) y, por lo tanto, se aplicó el método dogmático jurídico al recopilar la literatura más relevante que aborda pautas éticas orientadoras a la IA militar, que representan principios orientadores para determinados Estados y entidades supraestatales. El factor decisivo para la selección de las pautas éticas se fundamentó en la intención de proporcionar un mapeo del campo de la ética militar de la IA, respecto de aquellos estados que cuentan con mayor desarrollo y disponibilidad tecnológica, mediante la comparación de cada instrumento y el alcance que cada uno de ellos confiere a cada principio ético. Asimismo, el análisis de los lineamientos de carácter jurídico y/o legales recurrió al análisis de los principios destacados por la literatura en estos últimos años, a fin de proporcionar un examen realizado desde el DIH convencional, para así establecer si los SAAL son capaces de adecuarse a dichos principios.

El artículo está estructurado de la siguiente manera: en primer lugar, se explica el desarrollo y la aplicación de la IA en el ámbito militar, y cómo ha dado origen a nuevas armas con mayor grado de autonomía respecto a sus usuarios. En segundo lugar, se analizan los distintos instrumentos que contienen principios éticos reconocidos por determinados Estados, para así comprender el alcance de cada uno de ellos y su repercusión en el debate actual. Finalmente, se examinan los principios legales derivados del DIH, que resultan aplicables a este tipo de armas autónomas.

2. Sistemas autónomos de armas letales: incorporación de la inteligencia artificial en el ámbito militar

La IA en el ámbito militar está siendo la base para la creación de nuevas armas, dando origen así a los SAAL, armamentos que funcionan por medio de procesos complejos como el aprendizaje automático. Estos sistemas están convirtiendo en un fenómeno global y se espera que se vuelvan más comunes, especialmente en los países que cuentan con la tecnología disponible (Saria y Celik, 2021). Estos países gradualmente pretenden incorporar a los SAAL en el ámbito militar debido al avance de esta tecnología y al "aumento de las capacidades de los sensores, las capacidades analíticas y su integración en respuesta al ritmo cada vez mayor de las operaciones militares" (Sehrawat, 2021, p.361).

En particular, los SAAL se han conceptualizado como "armas que, una vez activadas, seleccionan y atacan objetivos cuyo enfrentamiento no está predeterminado por un ser humano (el llamado "hombre fuera del escenario del bucle")" (Bo, 2021, p.21). Estas tecnologías, por medio de "una computadora a bordo, elige[n] los objetivos y toma[n] decisiones de forma autónoma sin un ser humano en el circuito" (Sehrawat, 2021, p.316), por lo que "están preprogramadas para no necesitar un humano detrás de los controles para disparar, moverse o comunicarse cuando se enfrenta a un enemigo" (Press, 2018, p.1339).

Aunque es posible identificar características que destacan el funcionamiento de estas armas, existe la dificultad de establecer un concepto categórico, puesto que no se ha querido limitar su alcance debido al constante avance tecnológico. Bajo la misma lógica, no existe un acuerdo internacional que establezca qué tipo de armas pueden considerarse bajo esta denominación. Este término varía según la realidad de cada país que emplea tales tecnologías, siendo una realidad compleja que amerita acuerdos internacionales para lograr consensos respecto a una conceptualización de uso general[3] (Aravena Flores, 2024b). Sin embargo, según lo expresado por el Comité Internacional de la Cruz Roja (CICR), se trataría de un término amplio que comprendería "cualquier tipo de sistemas de armas, sea que operen en el aire, en la tierra o en el mar, con autonomía en sus funciones críticas, esto es, un arma que puede seleccionar y atacar objetivos sin intervención humana" (Comité Internacional de la Cruz Roja, 2015, p. 58). La importancia de este concepto radica en que, aunque los pronunciamientos del CICR no son una fuente vinculante de Derecho Internacional, forman parte del *soft law* y tienen una influencia significativa en la práctica estatal, el desarrollo normativo internacional y la interpretación del derecho vigente (Aravena Flores, 2024b).Debido al avance que han adquirido estas armas, han surgido diversos niveles que pretenden explicar la autonomía que pueden alcanzar estas armas, lo cual es posible apreciar en función de la menor o mayor intervención humana en la operación (Krishnan, 2009). En este sentido, tomando como referencia los niveles de autonomía propuestos por López-Casamayor[4], es posible advertir importantes diferencias. En primer lugar, los sistemas de armas autónomos corresponden a aquellos que, "una vez activado[s], está[n] habilitado[s] para seleccionar y atacar un objetivo sin ulterior intervención humana". En segundo lugar, los sistemas de armas supervisados son aquellos que, "una vez activado[s], identifica[n] y ataca[n] objetivos conforme a los criterios fijados en su programación, facultando al operador humano a intervenir, incluso abortando el ataque antes de que este se produzca dentro de un margen de tiempo desde la activación". Finalmente, los sistemas de armas semiautónomos "gozan de un grado de automatización limitado a, una vez activados, atacar objetivos o grupos de objetivos que han sido se seleccionados por un operador humano dentro de un área determinada" (López-Casamayor, 2019, p. 184).

En coherencia con lo anterior, se infiere que los riesgos de la utilización de estas armas provienen de la primera categoría expuesta, "en tanto que la intervención humana queda circunscrita al momento de activación del arma, dejando a su arbitrio cualquier decisión posterior relativa al empleo de fuerza letal" (López-Casamayor, 2019, p. 184). No obstante, aunque no existe un despliegue generalizado del primer nivel de autonomía, se han evidenciado las alertas y preocupaciones respecto a los últimos dos niveles de autonomía. Se han planteado cuestionamientos centrados en aclarar si una IA será responsable de los daños causados por sistemas semiautónomos y totalmente autónomos, y si un acto

3 De este modo, según Obregón Fernández y Lazcoz Moratinos (2024), en Estados Unidos (EE. UU.) no se considera como SAAL a los programas informáticos autónomos. En el caso de Francia, quedan excluidos los *softwares* que controlan los misiles, y finalmente, en Rusia, quedan excluidos los *softwares* que controlan las municiones inteligentes.

4 Los niveles de autonomía de las SAAL se han abordado desde diferentes nomenclaturas, pero en términos homologables. De esa manera, ha sido posible distinguir entre tres categorías: las armas *human-in-the-loop*, que pueden seleccionar objetivos y entregar fuerza solo con un comando humano; armas *human-on-the-loop*, capaces de seleccionar objetivos y entregar fuerza bajo la supervisión de un operador humano que puede anular las acciones de los robots; y las armas human-*out-of-the-loop*, capaces de seleccionar objetivos y entregar fuerza sin ningún aporte o interacción humana. En concreto, las armas fuera del circuito humano son la principal preocupación, puesto la toma de decisiones humanas está eliminada del proceso. Sin embargo, las armas de humanos en el circuito y humanos en el circuito también pueden generar preocupación (Sehrawat, 2021; López-Casamayor, 2019).

de matar ejecutado por un sistema basado en IA es legal o no (Szpak, 2019, p.118). Por esta razón, conforme estas armas adquieren mayor autonomía en la toma de decisiones, se ha observado un incremento en "el riesgo de erosión o difusión de la responsabilidad y la rendición de cuentas por estas decisiones" (Sehrawat, 2021, pp. 322-323).

Si bien inicialmente esta problemática podría parecer distante o difícil de visualizar, no se puede desconocer que esta tecnología actualmente está siendo plenamente utilizada y exhibida en conflictos bélicos en esta última década[5]. La tecnología para desarrollar y construir SAAL existe en la actualidad, lo cual ha suscitado debates debido al peligro que implica su desarrollo y uso, así como la dificultad para determinar los límites que deben aplicarse, un marco regulatorio efectivo y la identificación del sujeto responsable por dichos actos. En este contexto, es crucial reconocer la alta probabilidad de que un SAAL en el campo de batalla puede seleccionar, apuntar y atacar, ocasionando la muerte o hiriendo a la población civil, e incluso llevar a cabo ataques desproporcionados o no programados, cuestión que ha suscitado preocupación en diversos sectores de la población civil y grupos organizados[6].

Así, el rápido desarrollo de estas tecnologías ha sensibilizado progresivamente a la sociedad en relación con los riesgos que podrían dar lugar el desarrollo y uso de estas tecnologías por los evidentes efectos dañinos en perjuicio de bienes jurídicos protegidos, dado que puede generar riesgos para los derechos de las personas y servir como herramienta para fines maliciosos, incluso servir como un medio para eludir responsabilidades (Aravena Flores, 2024b). Como resultado, esto ha motivado la búsqueda de límites éticos, normativos y/o legales destinados a maximizar los beneficios y, al mismo tiempo, restringir el potencial negativo de este tipo de armas, en conformidad con los compromisos que han asumido los Estados y los marcos regulatorios aplicables hasta el momento.

3. Los lineamientos éticos-jurídicos aplicables a los sistemas autónomos de armas letales

En virtud del desarrollo progresivo, la versatilidad y la adaptabilidad de las tecnologías de IA, surge la necesidad de determinar cuáles deben ser los lineamientos éticos y legales que deben aplicarse. Esto ha llevado a un extenso análisis y discusión sobre las consideraciones éticas y/o legales, dado la incertidumbre regulatoria que rodea a una tecnología cuyo alcance y grado de desarrollo aún no se han definido.

Ante esta incertidumbre jurídica, el debate preliminar se ha enfocado en la unificación de ciertos principios éticos cuyo objeto es resolver escenarios no previstos expresamente por el legislador, para así orientar el desarrollo y uso de esta tecnología conforme a

5 Uno de los casos más destacados en términos de repercusión social ocurrió en Libia en el año 2019, donde se utilizaron drones STM Kargu-2 de fabricación turca. Estos drones estaban armados y controlados por IA, y realizaron ataques a fuerzas rebeldes mientras se encontraban en repliegue (Majumdar *et al* , 2021). Es importante destacar estos ataques se realizaron sin la participación ni control humano, basándose en la decisión y control de la IA, hecho que alertó al mundo.

6 Véase, por ejemplo, los argumentos expuestos por Algorithm Watch que se centra en observar y analizar los sistemas automatizados de toma de decisiones y su impacto en la sociedad. De manera similar, destaca el movimiento Stop Killer Robts, que representa a organizaciones no gubernamentales (ONG) de amplia difusión con el objetivo de garantizar el control humano en el uso de la fuerza.

las bases o fundamentos de un determinado orden institucional. Como consecuencia del reconocimiento de tales principios, han convergido distintos instrumentos y recomendaciones emanadas tanto de organismos internacionales como de Estados, quienes han pretendido delimitar el desarrollo y despliegue generalizado de esta tecnología.

3.1. Los principios éticos aplicables a los sistemas autónomos de armas letales

Según el avance tecnológico y la creación de SAAL se está haciendo frente a nuevos dilemas éticos, producto del mayor nivel de autonomía que ha alcanzado estas armas. Aunque ninguna máquina actualmente alcanza "un nivel de autonomía, cuantitativa y ontológicamente cercano al de un ser humano, pese a que existan las que son capaces de ejecutar ciertas funciones o tareas con altos niveles de independencia respecto de los controladores, usuarios o diseñadores" (Navarro, 2023, pp. 55-56).

Frente a estos dilemas éticos, el desarrollo de la IA ha sido objeto de seguimiento, mediante la consolidación y el cumplimiento de ciertos valores éticos y/o principios que han pretendido encauzar la investigación y el desarrollo de esta tecnología, toda vez que se ha reconocido que "los sistemas que interactúan con los seres humanos requieren una atención particular a las ramificaciones éticas de su comportamiento" (Anderson y Anderson, 2015, p. 324). Aunque al principio se logró mitigar las preocupaciones éticas, evitando su uso en escenarios complejos, lo cierto es que actualmente no es posible restringir completamente el despliegue de la IA. Por lo tanto, los actos ejecutados por estos sistemas deben guiarse por principios éticos explícitamente que deben ser "abstracciones declarativas completas y comprensibles que representan sucintamente este consenso de manera centralizada, extensible y auditable" (Anderson y Anderson, 2015, p. 324).

Recientemente, se ha evidenciado una preocupación pública y una prioridad mundial por delimitar esta tecnología, motivo por el cual se han consolidado ciertos textos que han uniformado criterios limitadores, denominados Directrices Éticas de Inteligencia Artificial (DEIA). Estos instrumentos han establecido una serie de deberes de conducta tanto para investigadores y desarrolladores, sin perjuicio de que algunos de sus mandatos también sean aplicables directamente a sus usuarios, con el fin de reforzar pautas sobre la rectitud ética de las acciones relacionadas con la IA.

Bajo dicha perspectiva, estas directrices se han transformado en lineamientos de carácter deontológicos, al tratarse de "imperativos categóricos que se consideran universalmente válidos e incontrovertibles" (Bedecarratz Scholz y Aravena Flores, 2023, p. 206), cuya labor se ha encausado en prevenir eventuales resultados lesivos que se podrían verificar a partir de la interacción indiscriminada de la IA, afectando así determinados bienes jurídicos protegidos por la comunidad. En el ámbito civil de esta tecnología, se reconocen principios éticos fundamentales como la transparencia, privacidad, seguridad, justicia y responsabilidad[7].

7 Para un análisis en particular, véase Bedecarratz Scholz y Aravena Flores (2023) p. 207 y siguientes, en donde se distinguen los principios éticos prevalentes a partir del análisis de determinadas y particulares directrices éticas de diversos Estados, entre ellas: *Usa Draft Guidance for regulation Artificial intelligence applications* de EE. UU., *Beijingd AI principles* del Gobierno de China, *Recommendation of the council on artificia intelligence* de la OCDE, *Directrices éticas para una IA fiable* de la Comisión Europea y la *Ethically Aligned Design* emanada del IEEE.

Sin embargo, la situación resulta ser más compleja en el plano estrictamente militar, puesto que es posible afirmar que existe menor claridad en cuanto a los límites éticos que deben aplicarse debido a su incipiente desarrollo por parte de los Estados. La preocupación en esta materia varía dependiendo del alcance tecnológico y los propósitos que se pretendan alcanzar con estas tecnologías. Esto ha llevado a que los gobiernos y entidades supraestatales acordar un marco ético uniforme o compartido para regular la IA militar (Galliott *et al*, 2022). Sin embargo, los Estados con mayor alcance tecnológico, como EE. UU. y Reino Unido, y determinadas comunidades internacionales, como la Organización del Tratado del Atlántico Norte (OTAN) y la Organización para la Cooperación y el Desarrollo Económicos (OCDE), han expuesto un ecosistema de valores éticos para IA militar, orientados a generar confianza en los usuarios finales de estas armas, los combatientes implicados en los conflictos, el público en general, los aliados y los socios internacionales.

Tabla 1. *Visión de conjunto respecto a los principios reconocidos en los textos analizados*

Texto analizado	*U.S. Department OF Defense Responsible Artificial Intelligence Strategy and Implementation Pathway*[8]	*Ambitious, Safe, Responsible: our Approach to the Delivery of Ai-Enabled Capability in Defence*	*Estrategia de inteligencia artificial para la OTAN*[9].	*Recommendation of the council on OECD Legal Instruments Artificial Intelligence*
Institución	Gobierno de Estados Unidos	Gobierno de Reino Unido	Organización del Tratado del Atlántico Norte. (OTAN/NATO)	Organización para la Cooperación y el Desarrollo Económico (OCDE/OECD)
Año de dictación	2022	2022	2021	2024
Control humano		●		
Legalidad			●	●
Responsabilidad/ rendición de cuenta	●	●	●	●
Equitativo	●		●	
Trazable/ explicable	●	●	●	●
Confiable/ seguro	●	●	●	●
Gobernable	●		●	
Mitigación de riesgos y daños		●	●	●

Fuente: Elaboración propia del autor a partir de los antecedentes disponibles en cada instrumento.

8 Este documento puede analizarse de forma complementaria o accesoria con los siguientes documentos dictados por el mismo Gobierno de EE. UU: i) *AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense* y ii) *AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense*.

9 En el transcurso de 2021, la OTAN elaboró su primera estrategia de IA. De esa manera, en el texto aprobado por los aliados miembros dieron a conocer un resumen de la estrategia, en consonancia con el objetivo declarado de la alianza sobre proporcionar ejemplo y fomentar el desarrollo y uso de la IA de manera responsable.

Revista de Derecho. Año XXIII (Diciembre 2024), Nº 46, pp. 1-25 - https://doi.org/10.47274/DERUM/46.3 | ISSN: 1510-5172 (papel) - ISSN: 2301-1610 (en línea)
Universidad de Montevideo, Uruguay

De esta forma, frente a la existencia de determinadas pautas éticas, es indispensable determinar aquellos valores éticos que se están siendo adoptados por cada Estado u organización, dado que revelan los intereses de cada parte y, al mismo tiempo, permiten comprender la dirección hacia la cual se pretende llegar con esta tecnología. Es importante subrayar que la formulación de dichos principios para los fines de defensa o militares es un área de trabajo reciente para los gobiernos y para interesados, pero ha demostrado un rápido avance en estos últimos años.

En la Tabla 1 se exponen cuatro instrumentos específicos que han abordado los límites de la inteligencia militar, especialmente en el plano de la defensa. El objetivo es identificar ciertos principios prevalentes y el ámbito de alcance que se le ha atribuido a cada uno.

En conformidad a lo evidenciado en la visión conjunta de dichos instrumentos y a pesar de la diversidad de principios que se pretende reconocer, es posible identificar que predominantemente se vislumbra la tendencia a reconocer como valores éticos indispensables en el desarrollo de IA militar los principios de responsabilidad o rendición de cuentas, principio de trazabilidad o explicabilidad y finalmente el principio de confiabilidad. A continuación, se procede a explicar y analizar cada uno de estos principios.

a) Principio de responsabilidad o rendición de cuentas

Uno de los principios que ha alcanzado mayor reconocimiento tanto para la aplicación de IA en sus usos civiles como en el ámbito militar es el principio de responsabilidad. El alcance de este principio se ha ampliado o reducido por cada emisor, atendiendo a la realidad de cada uno de los Estados en conformidad con sus políticas públicas nacionales o los fines que orientan la implementación de la IA en el ámbito militar.

Según explica el Gobierno de EE. UU., en virtud de este principio se pretende ejercer los niveles apropiados de juicio y cuidado, asegurando que las autoridades mantengan la responsabilidad del desarrollo, despliegue y uso de las capacidades de la IA. En este sentido, en el ámbito de la defensa se exige una supervisión continua en cada una de sus etapas, que incluye "normas de comprobación, controles de responsabilidad, orientación, integración de sistemas humanos y consideraciones de seguridad" (DoD Responsible AI Working Council, 2022, pp. 5-6). Se destaca que esta vigilancia va más allá de las métricas de rendimiento tradicionales y se manifiesta en todo momento, presentando orientaciones y normas desde el prototipo hasta la producción y el uso.

En virtud de este principio, se pretende evitar que la IA se aplique de manera poco ética o irresponsable, incluso en escenarios de combate. A través de este enfoque, se favorecerá que los desarrolladores y usuarios cuenten con niveles adecuados de confianza en el sistema de IA, facilitando así la rápida adopción y operacionalización de la nueva tecnología (DoD Responsible AI Working Council, 2022, p.6).

Por otro lado, según el instrumento del Ministerio de Defensa de Reino Unido, este principio se centra en establecer la responsabilidad humana por los actos o funciones que ejecute un sistema de IA, asegurando la rendición de cuentas por los resultados. Esto se logra mediante medios definidos que clarifiquen y mantengan el control humano en cuanto a los medios utilizados, la naturaleza y las limitaciones de dicho control. Este control debe estar presente en cada uno de los ciclos, desde la recopilación y filtrado

de la información presentada a los tomadores de decisiones, la automatización de procesos previamente dirigidos por humanos o procesos mediante los cuales los sistemas habilitados para IA aprenden y evolucionan, inclusive después de su implementación inicial (UK Ministry of Defense, 2022, p.9). Asimismo, se reconoce la complejidad que significa el control humano sobre esta tecnología, especialmente debido a los diversos contextos y capacidades con las que cuenta cada sistema de IA. Sin perjuicio de ello, se exige una articulación de la propiedad del riesgo desde el desarrollo hasta el desmantelamiento, incluida la redistribución en nuevos contextos (UK Ministry of Defense, 2022, p.10).

Finalmente, en cuanto a lo expuesto por las organizaciones supraestatales, OTAN destaca que, respecto de esta tecnología se debe desarrollar y utilizar conforme a los niveles apropiados de juicio y cuidado. El objetivo es establecer la responsabilidad humana y, de esa manera, garantizar la rendición de cuentas (OTAN, 2021a; OTAN, 2021b). Por otra parte, la OCDE declaró que los actores de la IA deben ser responsables del correcto funcionamiento de los sistemas de IA, considerando sus funciones y el contexto, de manera coherente con el estado de la técnica y su capacidad de actuar. Esto implica gestionar en cada una de las fases del ciclo de vista de estos sistemas los riesgos que deriven de ellos, tanto de forma individual por parte de cada uno de los responsables como a través de la cooperación entre los diferentes sujetos responsables (OCDE, 2024, p. 9).

A partir del análisis de los distintos enfoques atribuibles a este principio, es posible deducir que existe una tendencia compartida en cuanto a garantizar la responsabilidad de las personas humanas. Esto se hace con el objetivo de evitar escenarios donde nadie sea responsable o donde la responsabilidad se atribuya a la capacidad de la IA en lugar de a los humanos. Este objetivo se pretende lograr mediante la supervisión continua y el control del riesgo en cada una de las etapas, incluyendo el desarrollo del sistema de IA y el desmantelamiento de este.

Sin embargo, se aprecia que, los instrumentos aplican de forma diferenciada la expresión responsabilidad o rendición de cuentas o, en términos intercambiables, siendo formalmente distinto cada uno de ellos. La responsabilidad implica la atribución de consecuencias, ya sea jurídicas o no, por las acciones u omisiones cometidas, por lo que está ligada al concepto de responder ante una infracción de un marco normativo. Por otro lado, la rendición de cuentas implica un proceso activo de justificación y transparencia, donde el actor responsable no solo responde por sus actos, sino que debe explicar y proporcionar evidencia del cumplimiento de sus deberes. Hacer esta distinción es crucial, pues mientras la responsabilidad puede abordarse post-facto, la rendición de cuentas opera como un mecanismo preventivo que busca evitar acciones indebidas mediante el control continuo.

La relación entre ambos conceptos es complementaria, ya que la rendición de cuentas no solo materializa el principio de responsabilidad, sino que también lo refuerza al permitir la evaluación y fiscalización del comportamiento del actor responsable en tiempo real. Mientras que la responsabilidad suele activarse tras la comisión de un hecho, la rendición de cuentas se activa antes, durante y después del acto, lo que contribuye a prevenir fallos y asegura que las acciones estén alineadas con los estándares éticos y normativos. De este modo, la rendición de cuentas actúa como una herramienta

fundamental para garantizar que el principio de responsabilidad no quede reducido a la mera atribución de consecuencias, sino que también se fomente la corrección y mejora continua de las conductas que podrían generar perjuicios.

A pesar de esta distinción terminológica, a futuro estos instrumentos tendrán que velar por la implementación efectiva de estos principios en el contexto de los SAAL, con el fin de no restringir el alcance de estos límites éticos a meras declaraciones de compromisos. Según este enfoque será crucial la creación de protocolos de evaluación de riesgos y procedimientos de control ex ante y ex post. Esto implica establecer estándares rigurosos de pruebas y certificación para los SAAL, que evalúen su comportamiento en escenarios complejos y de alto riesgo. Además, es necesario diseñar procedimientos que definan las instancias de revisión en casos de fallos, atribuyendo responsabilidades de manera precisa entre fabricantes, operadores y los propios sistemas autónomos, en función de la naturaleza del fallo. De este modo, se garantiza que estos principios éticos se traduzcan en una aplicación real y eficaz, evitando que se conviertan en meras aspiraciones normativas.

b) Principio de trazabilidad o explicabilidad

En los términos expuestos por el Gobierno de EE. UU., este principio establece que las capacidades de la IA deben desarrollarse y desplegarse para que "el personal pertinente posea una comprensión adecuada de la tecnología, métodos aplicables a las capacidades de IA, incluyendo con metodologías, fuentes de datos y diseño procedimiento y documentación" (DoD Responsible AI Working Council, 2022, p. 42). De forma equivalente, para el Gobierno de Reino Unido, este principio precisa que los sistemas de IA y sus resultados "deben ser entendidos adecuadamente por las personas relevantes, con mecanismos para permitir esta comprensión como parte explícita del diseño del sistema" (UK Ministry of Defense, 2022, p. 10). En este contexto, la toma de decisiones efectiva y ética se extiende "desde la primera línea de combate hasta las operaciones administrativas, siempre está respaldada por una comprensión adecuada del contexto por parte de quienes toman las decisiones" (UK Ministry of Defense, 2022, p. 10).

En ese orden de ideas, es posible constatar que todo el personal de defensa debe tener una comprensión adecuada y específica del contexto de los sistemas habilitados para IA con los que operan y con los que trabajan. Sin embargo, el nivel de comprensión diferirá según el conocimiento requerido para actuar éticamente en un rol dado y con un sistema dado (UK Ministry of Defense, 2022, p. 10). Para cumplir con este principio, se explicita que las personas deben estar debidamente capacitadas y ser competentes para operar o comprender estas herramientas. Para ello, es necesario verificar que los sistemas habilitados para IA funcionen según lo previsto.

No obstante, debido a la naturaleza de "caja negra" de algunos sistemas de aprendizaje automático, que dificulta explicar por completo su funcionamiento, se requerirá "poder auditar los sistemas o sus resultados a un nivel que satisfaga a quienes son debidas y formalmente responsables" (UK Ministry of Defense, 2022, p. 10). Por lo tanto, los mecanismos para interpretar y comprender estos sistemas deben ser una parte crucial y explícita del diseño del sistema durante todo el ciclo de vida. Según lo planteado, esta exigencia debe extenderse igualmente a los proveedores comerciales, fuerzas aliadas y civiles.

Por su parte, la OTAN expuso fundamentos similares, sosteniendo que las aplicaciones de IA deberán ser comprensibles y transparentes, incluso mediante "el uso de metodologías, fuentes y procedimientos de revisión. Esto incluye mecanismos de verificación, evaluación y validación a nivel de la OTAN y/o nacional" (OTAN. 2021a; OTAN, 2021b). En términos similares, la OCDE declaró el compromiso con la transparencia y la divulgación, velando por la comprensión de los sistemas de IA, incluidas sus capacidades y limitaciones. Esto implica informar sobre las fuentes de datos/entradas, factores, procesos y/o lógica que llevaron a la predicción, recomendación o decisión, permitiendo así a los afectados por un sistema de IA comprender el resultado. Además, se destaca la importancia de ofrecer información que permita a quienes se ven afectados negativamente por un sistema de IA cuestionar su resultado (OCDE, 2024, p.9).

De la forma que se ha ido precisando este principio, es evidente que busca dar respuesta a ciertos desafíos técnicos inherentes a la naturaleza de los sistemas de IA, particularmente la opacidad de su funcionamiento. Esta opacidad se refiere a la situación en la cual "no se puede conocer, incluso para sus creadores, y, por lo tanto, mucho menos para sus usuarios legos, cómo funciona un sistema de IA y toma decisiones" (Giannini y Kwik, 2023, p. 11) Por ello, este principio pretende garantizar un entendimiento adecuado por las personas relevantes, con el fin de que los sistemas de IA militares sean comprensibles al encontrarse claras sus metodologías o procedimientos de funcionamiento.

c) Principio de confiabilidad

Desde el Gobierno de EE. UU se ha manifestado la necesidad de contar con una IA responsable para lograr una legítima confianza en la utilización de esta tecnología, puesto que se ha afirmado que "sin confianza los combatientes y los líderes no emplearán Al con eficacia" y que "el pueblo estadounidense no apoyará el uso continuado y la adopción de dicha tecnología" (DoD Responsible AI Working Council, 2022, p. 8). Para ello, se pretende garantizar la confianza en la IA, teniendo en cuenta las condiciones en las que se va a desplegar, entre otros factores relevantes. Este objetivo se irá cumpliendo en la medida en que se aplique un enfoque integral de gestión de riesgos que aborde los riesgos a nivel de sistema, institucionales y sociotécnicos. Esto generará una "evaluación multidimensional y contextual multidimensional y contextual del riesgo en el diseño, despliegue, desarrollo y uso de las capacidades de IA en una amplia gama de escenarios" (DoD Responsible AI Working Council, 2022, p. 8).

Otro aspecto para destacar se refiere al alcance e importancia que se le ha otorgado a este principio, toda vez que se sostiene que no se puede confiar únicamente en los avances tecnológicos por sí mismos, sino que habrá que atender a "los factores clave de la fiabilidad, (...) la capacidad de demostrar una gobernanza fiable estructura, así como la provisión de una formación y educación adecuadas de la mano de obra" (DoD Responsible AI Working Council, 2022, p. 8). Todo esto tiene como objetivo fomentar niveles adecuados de confianza, permitiendo a los trabajadores pasar de concebir la IA como una tecnología enigmática e incomprensible a comprender las capacidades y limitaciones de esta tecnología ampliamente adoptada y aceptada. Esto también aplica para los desarrolladores y usuarios, quienes confiarán en que existen medidas para aplicar los principios éticos de IA y para informar sobre posibles casos de afectación por ellos. Por tanto, este principio se entiende de tal manera que la IA "tendrá usos explícitos y

bien definidos, y la seguridad, protección y eficacia de dichas capacidades estarán sujetas a pruebas y garantías dentro de esos usos definidos a lo largo de todo su ciclo de vida" (DoD Responsible AI Working Council, 2022, p. 41).

Por otro lado, este principio ha sido reconocido por el Ministerio de Defensa del Reino Unido, considerando que los sistemas dotados de IA "deben ser demostrablemente confiables, robustos y seguros" (UK Ministry of Defense, 2022, p. 11). En este sentido, "deben cumplir con los criterios de diseño e implementación previstos y funciona sociotécnicos r como se espera, dentro de parámetros de rendimiento aceptables" (UK Ministry of Defense, 2022, p. 11). Estos criterios deben revisarse y probarse regularmente para garantizar la confiabilidad de manera continua y progresiva a medida que estas tecnologías aprenden y evolucionan con el tiempo o se implementan en nuevos contextos o funciones. Asimismo, dado el contexto operativo único de defensa y los desafíos del entorno de la información, se precisa que este principio "requiere que los sistemas habilitados para IA sean seguros y un enfoque sólido para la ciberseguridad, la protección de datos y la privacidad" (UK Ministry of Defense, 2022, p. 11).

Para la OTAN, este principio se traduce en que las aplicaciones de IA tendrán que estar condicionadas a usos explícitos y bien definidos, estando la seguridad, la protección y la solidez de dichas capacidades sujetas a pruebas y garantías dentro de esos casos de uso durante todo su ciclo de vida, incluso a través de procedimientos de certificación establecidos por la OTAN y/o nacionales. Y en el caso de la OCDE, este principio se manifiesta en la necesidad de velar por una IA robusta y segura para que, en condiciones de uso normal, uso previsible o mal uso, u otras condiciones adversas, funcione adecuadamente y no plantee riesgos irrazonables para las personas. En caso de que se le otorgue un uso indebido, existen mecanismos necesarios para que sean anulados, reparados y/o desmantelados de manera segura según sea necesario (OCDE, 2024).

Para concluir esta sección, se sostiene que, a partir del análisis de los principios enunciados, se puede afirmar que su reconocimiento significa dar un gran paso en una trayectoria mucho más larga e intrincada, producto del acelerado avance de la IA en el ámbito militar, lo que depara un largo proceso de transformación, adopción y adaptación tecnológica. Dentro de los aspectos positivos, se advierte que la construcción de principios éticos permite crear nuevas estructuras organizacionales, lo que constituye el primer paso para construir las bases de nuevos marcos regulatorios específicos para esta tecnología. Sin embargo, para integrar estos principios es indispensable brindar las orientaciones y herramientas necesarias que permitan darles operatividad a estos principios, con el fin de aplicarlos en el contexto del desarrollo y uso de la IA militar.

Es importante agregar que la construcción de estos límites éticos, reconocidos por cada Estado u organización, se alza como un límite para cada parte que los suscribe. No obstante, también alcanzan a los sectores industriales, ya sean públicos o privados, que diseñan, desarrollan o usan esta clase de IA. Esto permite tomar consciencia de que el uso ético de la IA no solo es un mero imperativo de carácter moral, sino que también debe concederse el mismo sentido para efectos de la sostenibilidad operativa de esta tecnología, a la que deben adherirse igualmente los contratistas o el sector privado en todo el proceso o curso de desarrollo y uso de esta tecnología, en función de estos principios éticos.

Igualmente, es importante destacar que el desarrollo de estas directrices éticas en el plano militar o de defensa permite crear consciencia y una mentalidad que garantice estos principios en toda la comunidad internacional. Esto podría cambiar la forma en que los distintos gobiernos se autorregulen y colaboren mutuamente en esta materia para alcanzar uniformidad en los principios rectores de esta tecnología, que podrían repercutir en el desarrollo de una regulación vinculante para los Estados.

Pese a lo enunciado anteriormente, es necesario plantear ciertas consideraciones críticas que permitan aclarar si estos límites éticos resultan ser suficientes para evitar los efectos disruptivos de estas tecnologías. Para dar respuesta a ello, si se considera la naturaleza de las directrices éticas y los documentos que contienen estos principios éticos, se puede afirmar que su cumplimiento dependerá en gran medida de la voluntad de cada suscriptor, al no ser constitutivo de un instrumento que permita exigir su cumplimiento forzado, dado su carácter deontológico. Por este motivo, su efectividad se ve reducida al libre albedrío de cada uno de los destinatarios. Por otro lado, el análisis de estas dimensiones éticas ha señalado que "la generalidad y superficialidad de las pautas éticas en muchos casos no solo impide a los actores adecuar su propia práctica a ellas, sino que alienta la devolución de la responsabilidad ética a otros" (Hagendorff, 2020, p. 112).

En concordancia con lo expuesto, se hace evidente que estas construcciones éticas no son suficientes para limitar esta tecnología militar, lo que lleva a reflexionar si se pueden complementar estos valores éticos con alguna normativa adicional que limite el uso y desarrollo de esta tecnología, siendo indispensable examinar la existencia de normas jurídicas o legales que den cobertura a dichos principios.

3.2. Los principios reconocidos por el derecho internacional humanitario aplicables a sistemas autónomos de armas letales

Como punto de partida, se debe considerar que a partir de la realidad actual de los ecosistemas regulatorios aplicables a la IA es posible sostener que, por el momento, no se cuenta particularmente con normas de carácter formales, tales como leyes, que permitan regular la investigación y desarrollo de esta tecnología (Bedecarratz Scholz y Aravena Flores, 2023, p. 205). Por este motivo, en lo que respecta a los SAAL, la cuestión principal se relaciona con la capacidad de esta tecnología para cumplir con las normas éticas de conducta adecuada en la guerra (*Ius in Bello*).

Examinando este tema de acuerdo con los principios fundamentales del DIH, entendiendo que son "directrices universales, reconocidas por las naciones civilizadas obligatorias para los Estados más allá de un vínculo convencional" (López Díaz, 2009, p. 230), se puede limitar y guiar el comportamiento de los intervinientes en un conflicto armado.

A pesar de que estas normas establecen restricciones al uso de medios y métodos de combate durante las hostilidades, la realidad es que el DIH no fue originalmente diseñado para abordar los desafíos presentados por el desarrollo de la IA en contextos bélicos (Vigevano, 2021, p. 1). No obstante, es posible analizar la problemática de los SAAL en las intersecciones del DIH, con el propósito de limitar los avances tecnológicos. En este contexto, lo que se ha discutido en los últimos años sobre estas armas autónomas es si

pueden limitarse por los principios fundamentales de esta rama del derecho y si ellas pueden cumplir con tales exigencias.

a) Principio de distinción

Uno de los principios que mayor discusión ha planteado en relación con los SAAL es el principio de distinción, consagrado en los Protocolos Adicionales a los Convenios de Ginebra, en particular del artículo 48 del Primer Protocolo Adicional a los Convenios de Ginebra relativo a la Protección de Víctimas de los Conflictos Armados Internacionales, junto a lo dispuesto en los artículos 51 y 52 de dicho texto, cuyo objeto es restringir los objetivos susceptibles de ser atacados, por lo que se requiere distinguir entre objetivos militares y aquellos que no lo son, estableciéndose así una distinción entre bienes civiles y objetivos militares, siendo estos últimos los únicos que podrán ser objeto de ataque(Chávez *et al*, 2019).

Sobre la base de tales diferenciaciones, el principio de distinción establece que "solamente los que participan en las hostilidades y los objetivos militares podrán ser objeto de ataques, no pudiendo, por ende, atacarse a la población civil, que en todo tiempo y circunstancia deberá ser respetada" (Salmón, 2012, p. 54). En consecuencia, se requiere saber con certeza que las personas a las que se ataca son un objetivo militar o si se trata de un bien civil, lo que implica tener claro que las partes en conflicto deben tomar todas las medidas necesarias para poder cumplir con este principio, con miras a poner a los civiles en un menor riesgo (Chávez *et al*, 2019).

Aplicando este principio a los SAAL, es posible advertir la incertidumbre que surge respecto de si estas armas dan cumplimiento a las exigencias de este principio. Frente a tales escenarios, la literatura ha brindado distintas respuestas en torno a si estas armas pueden o no cumplir con el principio de distinción.

Hay quienes proponen que los SAAL pudiesen cumplir con este principio en escenarios donde, al no ser identificable el objetivo como militar o combatiente, el arma autónoma, ante la duda, asuma que el objetivo es civil y, por tanto, no concrete el ataque en contra del objetivo (Bills, 2014). También se ha sugerido que el arma autónoma solo lleve a cabo el ataque cuando reciba efectivamente un ataque o disparos (Thurnher, 2012), esta afirmación se ha fundamentado en la idea de que estas armas podrían diferenciar entre objetivos civiles y militares de manera más eficiente que los seres humanos, dado que "el juicio humano puede resultar mucho menos fiable que los indicadores técnicos cuando se está en mitad de una batalla" (Schmitt y Thurnher, 2013, p. 235).

En términos similares, se ha intentado legitimar a los SAAL restringiendo su uso tan solo para responder ante algún ataque, empleándolos en operaciones militares donde resulte fácil de ubicar o seleccionar el objetivo, o programándolos para atacar armas, y no a quien las porta (Petman, 2017, p. 38). Incluso, algunos creen que los SAAL pueden "limitarse a las áreas rurales y despobladas, todo lo cual disminuiría significativamente estas preocupaciones" (Farinella, 2021, p. 510), siendo factible incluso que sean utilizados únicamente con la finalidad de combatir otros SAAL En esta misma línea argumentativa, se sostiene que este tipo de armas autónomas se restrinjan a zonas militarizada, de manera que no pongan a peligro a civiles y a la vez no estén equipadas con armas de destrucción masiva, causando daño solo a combatientes en un campo de batalla (Dombrovszki, 2021).

En contraposición a los argumentos expuestos, se ha precisado que los SAAL nunca podrán cumplir con este principio en la forma establecida por el Convenio de Ginebra y sus protocolos adicionales, dado que estas armas no son capaces de prevenir las diversas situaciones que pueden ocurrir en el contexto de conflictos armados ni lograr prever los mismos (Asaro, 2012). Por ello, estas armas aún no han alcanzado un grado de conciencia o razonamiento en términos iguales o superiores a los humanos y, como consecuencia de ello, pueden concretar actos sobre objetivos no autorizados por el DIH e incluso fallar en distinguir entre objetivos militares y personas o bienes civiles.

De forma complementaria, se ha sostenido que en un conflicto armado pueden suscitarse escenarios que no estén expresamente contemplados por la norma positiva e incluso por las normas consuetudinarias del DIH. Aún más, "han ocurrido situaciones que hasta el momento no han podido preverse por el derecho", por lo que, en el caso de los SAAL, estos "no podría razonar y actuar acorde a las distintas exigencias del DIH, ya que se trata de una programación y desarrollo anterior al combate o escenario de conflicto" (Chávez *et al*, 2019, p. 17).

A ello se debe agregar que, si bien se considera que estas armas continuarán progresando en el tiempo, persiste la incertidumbre sobre la posibilidad de que estas armas pueda llegar a distinguir correctamente entre un combatiente y un no combatiente. Además, el escenario se complejizará exponencialmente cuando los intervinientes en el conflicto no porten el uniforme o distintivo, y deban ser identificados como combatientes tan solo en función de su participación o en las hostilidades (Foy,2014).

Asimismo, se ha indicado que los algoritmos de funcionamiento de estas armas autónomas no son capaces de contextualizar y efectuar juicios humanos cualitativos para distinguir lo lícito de lo ilícito (Brehm, 2017). Por ello, Sharkey sostuvo que No existe actualmente un sistema de IA que pueda discernir entre bienes civiles y objetivos militares de manera adecuada, ya que las exigencias de este principio no se limitan a una mera diferenciación basada en uniformes, sino que implica reconocer a quienes están fuera del combate debido a heridas, enfermedad o incluso aquellos que se rinden. Por esta razón, afirmó que estas máquinas carecen no cuentan con aquellos elementos esenciales para asegurar su capacidad de acatamiento de este principio[10].

De este modo, pese a que un arma autónoma puede ejecutar un acto según su programación, se asevera que no es posible que dicho programa traduzca el principio de distinción en los términos del lenguaje de una computadora, partiendo de la base de que ni siquiera el DIH expresa una definición de sujeto civil que sea lo suficientemente específica y completa como para poder transformarla en un *software*[11]. Por ello, "el software del arma deberá abarcar toda posible situación en donde sea necesario aplicar

10 Para clarificar este argumento, Sharkey tomó como ejemplo el arma Harpy israelí, que corresponde a un vehículo aéreo no tripulado, cuyo funcionamiento depende de la detección de señales de radar. Mediante una base de datos, determina si estas señales son o no amistosas y, en caso de no serlo, conduce un ataque al objetivo seleccionado. Sin embargo, no existe forma de que dicho vehículo pueda determinar, por medio de sus radares, que el objetivo se encuentra en una base militar o en el techo de una escuela. Por esta razón, es incapaz de diferenciar si el objeto de su ataque es un objetivo militar o no. Por lo tanto, si bien se ha calificado al Harpy israelí como un arma autónoma, su funcionamiento es supervisado por una estación de control en tierra a cargo de supervisión humana. Por lo anterior, no llega a una autonomía completa y por sí sola no es capaz de dar cumplimiento con este principio (Sharkey, 2012).

11 Para efectos de comprender este argumento, se debe precisar que la conceptualización de personas "civiles" que contempla el Protocolo I de los Convenios de Ginebra y las normas de costumbre, se encuentra descrita en términos negativos, expresando que corresponde a todo aquel que no es un combatiente (Sharkey, 2012, p. 789).

distinción, e inclusive modificar sus actuaciones en base a los cambios que se puedan suscitar durante las hostilidades" (Carrera Herrera, 2021, p. 10). Sin embargo, esto resulta complejo de cumplir según los avances alcanzados por esta tecnología.

Por tanto, a partir de los argumentos expuestos, se puede concluir que en la actualidad no existe manera de que un SAAL pueda algoritmizar totalmente las exigencias de este principio, debido a que aún persisten diversas barreras tecnológicas que deben resolverse para que en el futuro puedan lograrlo (Carrera Herrera, 2021, p. 13). Entendiendo que incluso el procesamiento de estas máquinas está por detrás del razonamiento humano, especialmente en la evaluación de cambios contextuales que pueden conducir a errores en la selección y ataques a objetivos, con el potencial de poner en peligro vidas humanas, una situación que no puede ser aceptada según las demandas de este principio (Aravena Flores, 2024a).

b) Principio de precaución

El principio de precaución se encuentra consagrado en el artículo 57 del Protocolo Adicional I a los Convenios de Ginebra. Según lo dispuesto en dicho texto "las operaciones militares se realizarán con un cuidado constante de preservar a la población civil, a las personas civiles y a los bienes de carácter civil"[12]. Por ello, ante alguna violación a este principio, los combatientes deberán suspender o cancelar cualquier agresión en contra de objetivos no susceptibles de ataque"[13].

Así, las precauciones entre quienes planifican o deciden un ataque son, por un lado, hacer lo posible para verificar que los objetivos a atacar sean militares y no civiles, y, por otro lado, deben tomar todas las precauciones posibles en la elección de los medios y métodos de ataque, para evitar o disminuir cualquier daño en contra de objetos o personas civiles[14].Como se desprende de la expresión de este principio, e requiere que las partes tomen una serie de medidas cautelares al llevar a cabo un ataque, diseñadas para evitar que la población civil sufra sufrimientos innecesarios o excesivos, al punto de "abstenerse de realizar un ataque cuando sea de prever que causará incidentalmente muertos o heridos en la población civil, daños a bienes de carácter civil o ambos, que serían excesivos en relación con la ventaja militar prevista" (Salmón, 2012, p. 99).

Según lo previsto en el artículo 57 del Primer Protocolo Adicional y en relación con la realidad que presentan los SAAL, se podría argumentar que estas armas cumplirán con este principio siempre que las operaciones militares realizadas con ellas eviten causar daños o sufrimientos innecesarios y excesivos a civiles u otros bienes protegidos. Además, si se produce algún daño colateral como resultado del ataque, este no debe ser desproporcionado en comparación con la ventaja militar obtenida mediante el uso de estas armas.

Atendiendo a lo expuesto, se argumenta que un arma autónoma podría cumplir con el principio de precaución siempre que tenga la capacidad de interrumpir o detener el ataque si el objetivo parece no ser militar o está sujeto a protección especial. Esto implica que, frente a escenarios cambiantes donde el objetivo militar puede variar en su calidad

12 Protocolo I Adicional a los Convenios de Ginebra, art. 57(1).
13 Protocolo I Adicional a los Convenios de Ginebra, art. 57(2.a.iii) y (2.b).
14 Protocolo Adicional I a los Convenios de Ginebra, art. 57(a)(i)-(ii).

Revista de Derecho. Año XXIII (Diciembre 2024), Nº 46, pp. 1-25 - https://doi.org/10.47274/DERUM/46.3 | ISSN: 1510-5172 (papel) - ISSN: 2301-1610 (en línea)
Universidad de Montevideo, Uruguay

y no exista la suficiente certeza de su naturaleza, estas armas puedan abstenerse de concretar el ataque. Existiendo dicha capacidad, se ha estimado que el uso de armas autónomas respetaría el principio de precaución (Chávez *et al*, 2019). Además, se argumenta que los SAAL pueden mejorar el cumplimiento del principio de precaución, dado que pueden operar de manera conservadora al optar por disparar como una alternativa secundaria. También pueden ser programados para evaluar situaciones de forma similar a un operador humano, detectando cambios en el contexto y suspendiendo los ataques cuando resulta claro que los civiles están siendo seleccionados como objetivos (Sassóli, 2014).

En consecuencia, debido a las particulares condiciones y escenarios que se pueden verificar en un conflicto armado, las medidas necesarias dependen en gran medida del contexto y pueden cambiar de manera rápida e impredecible. Es fundamental verificar constantemente la selección del objetivo, el tipo de arma, el tiempo y el método de ataque, puesto que, si el ataque se vuelve ilegítimo, debe cancelarse o suspenderse antes de que impacte al objetivo. Por lo tanto, el arma debe ser capaz de recibir información continuamente para evaluar en todo momento la legitimidad del ataque y equilibrarla con la ventaja militar (Carrera, 2021). En ese sentido, se ha inferido que estas armas, en la mayor parte de las situaciones, no serán capaces de llevar a cabo la evaluación necesaria de forma independiente, siendo crucial contar con supervisión humana en las decisiones de ataque para garantizar que se dirijan únicamente contra objetivos militares legítimos (Boothby, 2009).

De este modo, es necesario destacar que el principio de precaución implica evaluaciones complejas que son difíciles de traducir adecuadamente a un *software* de IA (Boulanin y Verbruggen 2017). En consecuencia, se sostiene que las armas autónomas podrían cumplir con este principio y ser consideradas lícitas únicamente cuando no sea factible emplear un sistema que proteja mejor a los objetivos civiles sin comprometer la ventaja militar (Thurnher, 2016). Sin embargo, para lograr esto, un SAAL tendría que ser capaz de cancelar o suspender un ataque si el objetivo deja de ser militar, lo cual requiere un examen detallado caso por caso, algo que estas armas aún no pueden garantizar completamente (Aravena Flores, 2024a).

c) Principio de proporcionalidad

Este principio establece que un ataque no debe causar daños desproporcionados o excesivos en relación con la ventaja militar específica y directa obtenida (Carrera, 2021). Bajo esos términos, el Protocolo Adicional I determina que un ataque es considerado desproporcional "cuando sea de prever que causarán incidentalmente muertos y heridos entre la población civil, o daños a bienes de carácter civil, o ambas cosas, que serían excesivos en relación con la ventaja militar concreta y directa prevista"[15]. Esto significa que los daños a la población no deben ser desproporcionados en relación con la ventaja militar concreta, directa y prevista esperada.

En virtud de este principio, se debe cesar cualquier ataque que pueda previsiblemente causar la muerte o el daño a civiles, así como a sus pertenencias, por considerarse excesivo

15 Artículo 51.5, Protocolo Adicional I a los Convenios de Ginebra de 1949 relativo a la protección de las víctimas de los conflictos armados internacionales.

en relación con la ventaja militar pretendida. Sin embargo, es importante destacar que ello no significa que se prohíban absolutamente todo tipo de daños colaterales, siendo materia de prohibición la concreción de daños que pueden resultar excesivos (Thurnher, 2012).

Se ha planteado a favor de estas armas autónomas que podrían evaluar los daños colaterales durante un ataque, dado que las capacidades de la IA podrían ser superior a la de un soldado humano. Además, "evaluar un ataque proporcionado es una tarea computacional, donde las habilidades de las máquinas son mejores que las de los humanos" (Dombrovszki, 2021, p. 19). Para respaldar este argumento, se sostiene que los principales ejércitos ya realizan esta evaluación al respecto mediante un sistema conocido como "metodología de estimación de daños colaterales", que utiliza datos científicos y estándares objetivos (Boulanin y Verbruggen, 2017, p. 74; Schmitt, 2013, pp. 19-20).

Mediante estos métodos de estimación sistemáticos, se podrían tomar decisiones sobre la proporcionalidad requerida y programar adecuadamente los SAAL antes de un ataque, permitiéndoles incorporar en sus sistemas diferentes umbrales en función de niveles aceptables de daño colateral para objetivos militares específicos ya definidos. Por esta razón, siendo capaces de analizar estos factores y desarrolladas con tecnología avanzada, es probable que tengan más éxito en determinar estas variables en comparación con una persona natural, especialmente si se considera que los seres humanos utilizan diversas tecnologías para calcular el daño esperado (Chávez *et al*, 2019).

Además, se argumenta que, al incorporar ciertas las restricciones geográficas en el uso de SAAL, se contribuiría a cumplir con este principio en situaciones simples. Por ejemplo, se señala que las armas autónomas deberían ubicarse en áreas escasamente pobladas o a lo largo de una zona desmilitarizada, como una frontera (Boulanin y Verbruggen, 2017, pp. 74-75). Por otro lado, debido a que la evaluación que realizan las armas autónomas respecto a la situación no se ve afectada por aspectos emotivos como la ira o el miedo, ni por prejuicios raciales u otros, estas armas podrían actuar de manera más imparcial que los soldados o personas involucradas en un conflicto.

Contrariamente a los argumentos mencionados, se señala que, a pesar de los avances significativos en IA, siempre se ha atribuido exclusivamente a los seres humanos la facultad de juzgar si un ataque es proporcional o no. Esto se debe a que dicho juicio de valor se equipara a lo que haría un soldado razonable en ese lugar y bajo esas circunstancias específicas de un contexto (Thurnher, 2012). De manera similar, ciertas instituciones no gubernamentales exponen que los SAAL no podrán cumplir las exigencias de este principio, puesto que estos requieren más que un simple examen de datos cuantitativos, y un robot no puede programarse para replicar el proceso psicológico del juicio humano necesario para evaluar la proporcionalidad (Human Rights Watch, 2012).

Por tanto, no hay un método definitivo para cuantificar la importancia de la proporcionalidad, ya que no existe una norma que establezca cuántas vidas civiles pérdidas son proporcionales al ataque de una base militar (Petman, 2017). De este modo, para cumplir con este principio "deberán reaccionar a todas las situaciones cambiantes, y constantemente calcular el daño colateral y la ventaja militar" (Carrera, 2021, p. 15). Para ello, se requeriría de un sistema que cuenta con un procesamiento sofisticado y complejo, con facultades de sensibilidad y de discernimiento, y un algoritmo capaz de tomar decisiones rápidas y certeras. Esto se debe a que necesitan una evaluación

detallada basada en criterios subjetivos como la buena fe, el sentido común, la capacidad de discernimiento y la racionalidad (Sharkey, 2012).

En este sentido, actualmente no está claro si un SAAL podrá realizar evaluaciones subjetivas que involucran análisis de costo-beneficio inherentes a las decisiones de proporcionalidad (Farinella, 2021). Además, para cumplir con este principio, el arma autónoma necesitaría ser "actualizada constantemente sobre operaciones y planes militares" (Sassóli, 2014, p. 332), algo que en la actualidad es técnicamente poco factible de cumplir.

A partir de lo expuesto, es evidente que un SAAL no podría determinar un ataque adecuado a la luz del DIH, especialmente en lo concerniente al principio de proporcionalidad. Por lo tanto, se sugiere que dicho análisis requiere un "criterio humano" (Petman, 2017, p. 38), puesto que los SAAL no son capaces de capturar en tiempo real las señales contextuales necesarias para determinar la proporcionalidad entre el daño y la ventaja militar esperada. Además, no están aptos para realizar juicios de valor subjetivos que se basan en emociones y experiencias humanas (Farinella, 2021).

Conclusiones

En el contexto de este artículo de investigación, se analizó brevemente el alcance de algunos principios prevalentes sobre la IA ética en el ámbito militar. Se evidenció una convergencia en torno a tres principios éticos (responsabilidad o rendición de cuentas, trazabilidad o explicabilidad y confiabilidad), con alto grado de similitud en torno al sentido y el alcance, lo que demuestra una uniformidad sobre los valores éticos que deben delimitar el desarrollo y uso de estas armas. Pero, la uniformidad en torno a estos principios éticos plantea un desafío, que alude a la posible disparidad en su interpretación cuando se aplican en contextos culturales, ideológicos y de gobierno diversos. Aunque tales principios parecen tener consenso en su reconocimiento, su implementación puede variar según el marco político, legal y ético de cada país o región. Esta pluralidad de interpretaciones podría generar tensiones sobre nuevas regulaciones que orienten el desarrollo y uso de la IA militar.

Es crucial reconocer que el abordaje primario de esta problemática debe ser ético, instando a los gobiernos a llevar a cabo revisiones éticas de la IA militar para restringir su uso en combate hasta que no haya un consenso público lo suficientemente sólido y robusto que limite los potenciales riesgos que implica esta tecnología. Sin embargo, la naturaleza de los instrumentos que contienen tales principios carece de una suficiencia que permita delimitar la IA militar, dado que estos textos dependen de las voluntades de sus suscriptores y de los compromisos que se hayan asumido.

Como consecuencia de ello, el uso de esta tecnología debe velar por el cumplimiento de los lineamientos éticos como punto de partida, para así asegurar consecuencialmente el DIH. Desde este último plano, el uso de estas armas autónomas enfrenta varios obstáculos, puesto que, hasta la fecha y conforme al desarrollo actual de esta tecnología, las armas autónomas no son capaces de cumplir estrictamente con los principios del DIH. Por lo tanto, seguirá siendo necesario mantener un juicio humano en los actos desplegados

por esta tecnología militar, lo que hace indispensable mantener aún un control humano significativo en la decisión de atacar y seleccionar ciertos objetivos. Por ello, parece haber un consenso sobre la necesidad de que el ser humano ejerza un control suficiente sobre los SAAL. Mantener el papel fundamental del ser humano en esas decisiones y su control sobre ellas será esencial para evitar consecuencias impredecibles tanto para civiles como para combatientes, lo cual está en concordancia con los principios éticos y legales.

En resumen, frente a la realidad que presentan los SAAL, se debe reconocer la necesidad de cumplir con un estándar alto, tanto a nivel ético como legal, antes de desplegar el desarrollo y uso de este tipo de armas. Por los argumentos expuestos, los Estados deberían avanzar en la adopción de medidas legislativas a nivel nacional, con el objetivo de contar con normas legales específicas que regulen esta clase de armas autónomas, además de promover conversaciones sobre una elaboración de una norma internacional de carácter vinculante que atienda a los principios éticos prevalentes y a los principios del DIH.

Referencias bibliográficas

Anderson, M., y Anderson, S. (2015). Toward ensuring ethical behavior from autonomous systems: A case-supported. *Industrial Robot, 42*(5), 324-331. https://doi.org/10.1108/IR-12-2014-0434

Aravena Flores, M.A. (2024a). Dilemas derivados del uso de sistemas autónomos de armas letales en el derecho internacional humanitario. *Justicia (Barranquilla. En línea), 29*(45), 1-15. https://doi.org/10.17081/just.29.45.7143

Aravena Flores, M.A. (2024b). Inteligencia artificial militar: problemas de responsabilidad penal derivados del uso de sistemas autónomos de armas letales. *Revista De Derecho (Coquimbo), 31*, 1-34. https://doi.org/10.22199/issn.0718-9753-6632

Asaro, P. (2012). On banning autonomous weapon systems: Human rights, automation, and the dehumanization of lethal decision-making. *International Review of the Red Cross, 94*(886), 687-709. https://doi.org/10.1017/S1816383112000768

Azuaje Pirela, M., y Finol González, D. (2023). Aproximaciones a la noción de inteligencia artificial y otros conceptos vinculados con ella. En M. Azuaje (Ed.), *Introducción a la Ética y el Derecho de la Inteligencia Artificial* (pp. 18-34). La Ley.

Bedecarratz Scholz, F., y Aravena Flores, M.A. (2023). Principios y directrices éticas sobre inteligencia artificial. En M. Azuaje (Ed.), *Introducción a la Ética y el Derecho de la Inteligencia Artificial* (pp. 204-218). La Ley.

Bills, G. (2014). LAWS onto themselves: Controlling the development and use of lethal autonomous weapons systems. *George Washington Law Review, 83*(1). https://www.gwlr.org/wp-content/uploads/2015/03/83-Geo-Wash-L-Rev-176.pdf

Bo, M. (2021). Autonomous weapons and the responsibility gap in light of the mens rea of the war crime of attacking civilians in the ICC statute. *Journal of International Criminal Justice, 19*(2), 275-299. https://doi.org/10.1093/jicj/mqab005

Boothby, W. (2009). *Weapons and the law of armed conflict*. Oxford University Press.

Boulanin, V., y Verbruggen, M. (2017). *Mapping the development of autonomy in weapon systems*. Stockholm International Peace Research Institute. https://www.sipri.org/publications/2017/other-publications/mapping-development-autonomy-weapon-systems

Brehm, M. (2017). Defending the boundary: Constraints and requirement on the use of autonomous weapon systems under international humanitarian law and human rights law. *Geneva Academy Briefing, 9*. https://doi.org/10.2139/ssrn.2972071

Carrera Herrera, B. (2021). Responsabilidad penal internacional para el empleo de armas completamente autónomas durante conflictos armados. *Revista de Investigación Académica y Educación ISTCRE, 5*(2), 93-101. https://www.revistaacademica-istcre.edu.ec/articulo/94

Čerka, P., Grigienė, J., y Sirbikytė, G. (2017). Is it possible to grant legal personality to artificial intelligence software systems? *Computer Law & Security Review, 33*(5), 685-699. https://doi.org/10.1016/j.clsr.2017.03.022

Chávez, D.M., Cruz García, C., y Herrera Jaramillo, P. (2019). Robots asesinos: ¿Realidad o ficción? Los sistemas de armas autónomas en el marco del derecho internacional humanitario. *USFQ Law Review, 6*(1), 11-28. https://doi.org/10.18272/lr.v6i1.1405

Comité Internacional de la Cruz Roja. (2015). *XXXII Conferencia Internacional de la Cruz Roja y la Media Luna Roja: Informe* (CICR, ES 321/C/15/11).

Defense Innovation Board. (2020a). *AI principles: Recommendations on the ethical use of artificial intelligence by the Department of Defense*. https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF

Defense Innovation Board. (2020b). *AI principles: Recommendations on the ethical use of artificial intelligence by the Department of Defense*. https://media.defense.gov/2019/Oct/31/2002204459/-1/-1/0/DIB_AI_PRINCIPLES_SUPPORTING_DOCUMENT.PDF

Departamento de Defensa de Estados Unidos. (2022). *U.S. department of defense responsible artificial intelligence strategy and implementation pathway*. http://bitly.ws/MnJP

DOD Responsible AI Working Council. (2022). *U.S. Department of Defense responsible artificial intelligence strategy and implementation pathway*. https://media.defense.gov/2022/Jun/22/2003022604/-1/-1/0/Department-of-Defense-Responsible-Artificial-Intelligence-Strategy-and-Implementation-Pathway.PDF

Dombrovszki, Á. (2021). The unfounded bias against autonomous weapons systems. *Információs Társadalom, 21*(2), 13-28. https://doi.org/10.22503/inftars.XXI.2021.2.2

Farinella, F. (2021). Sistemas de armas autónomos y principios del derecho internacional humanitario. *Quaestio Iuris, 14*(2), 504-514. https://doi.org/10.12957/rqi.2021.54593

Foy, J. (2014). Autonomous weapons systems: Taking the human out of international humanitarian law. *Dalhousie Journal of Legal Studies*, *23*, 47-70. https://digitalcommons.schulichlaw.dal.ca/djls/vol23/iss1/3/

Galliott, J.C., Cappuccio, M.L., y Wyatt, A. (2022). Taming the killer robot: Toward a set of ethical principles for military artificial intelligence. *Journal of Indo-Pacific Affairs*. https://www.airuniversity.af.edu/JIPA/Display/Article/3091254/taming-the-killer-robot-toward-a-set-of-ethical-principles-for-military-artific/

Giannini, A., y Kwik, J. (2023). Negligence failures and negligence fixes: A comparative analysis of criminal regulation of AI and autonomous vehicles. *Criminal Law Forum*. https://doi.org/10.1007/s10609-023-09451-1

Human Rights Watch. (2012). Losing humanity: The case against killer robots. https://www.hrw.org/sites/default/files/reports/arms1112_ForUpload.pdf

Krishnan, A. (2009). Killer robots: Legality and ethicality of autonomous weapons. Ashgate Publishing.

López Díaz, P. (2009). Principios fundamentales del derecho internacional humanitario. Revista Marina, 3, 230-238. https://revistamarina.cl/revistas/2009/3/lopez.pdf

López Jacoiste, E. (2020). El empleo de drones armados desde la perspectiva del derecho internacional humanitario y del derecho internacional de los derechos humanos. En M. J. Cervell (Ed.), Nuevas tecnologías en el uso de la fuerza: Drones, armas autónomas y ciberespacio (pp. 67-110). Thomson Reuters Aranzad. https://www.researchgate.net/publication/348993836_El_empleo_de_drones_armados_desde_la_perspectiva_del_Derecho_internacional_humanitario_y_el_Derecho_internacional_de_los_derechos_humanos

López-Casamayor, A. (2019). Armas letales autónomas a la luz del derecho internacional humanitario: Legitimidad y responsabilidad. Cuadernos de Estrategia, 201, 177-213. https://dialnet.unirioja.es/servlet/articulo?codigo=7230262

Majumdar Roy Choudhury, L., Aoun, A., Badawy, D., De Alburquerque, L.A., Marjane, Y., y Wilkinson, A. (2021). Final report of the Panel of Experts on Libya established pursuant to Security Council resolution 1973 (2011). https://documents-dds-ny.un-.org/doc/UNDOC/GEN/N21/037/72/PDF/N2103772.pdf?OpenElement

Ministerio de Defensa del Reino Unido. (2022). Ambitious, safe, responsible: Our approach to the delivery of AI-enabled capability in defence. http://bitly.ws/MnKm

Obregón Fernández, A., y Lazcoz Moratinos, G. (2024). La supervisión humana de los sistemas de inteligencia artificial de alto riesgo. Aportaciones desde el Derecho Internacional Humanitario y el Derecho de la Unión Europea. Revista Electrónica de Estudios Internacionales, 42, 1-29. https://reei.tirant.com/reei/article/view/2483

Organización del Tratado del Atlántico Norte (OTAN). (2021a). An artificial intelligence strategy for NATO. http://bitly.ws/MnKD

Organización del Tratado del Atlántico Norte (OTAN). (2021b). Summary of the NATO artificial intelligence strategy. https://www.nato.int/cps/en/natohq/official_texts_187617.htm

Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2024). Recommendation of the Council on Artificial Intelligence. https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449

Petman, J. (2017). Autonomous weapons systems and international humanitarian law: 'Out of the loop'?. The Eric Castren Institute of International Law and Human Rights. https://um.fi/documents/35732/48132/autonomous_weapon_systems_an_international_humanitarian_law__out_of_the/c0fca818-3141-b690-0337-7cfccbed3013?t=1525645981157

Press, M. (2018). Of robots and rules: Autonomous weapon systems in the law of armed conflict. Georgetown Journal of International Law, 48, 1337-1366. https://www.law.georgetown.edu/international-law-journal/wp-content/uploads/sites/21/2018/05/48-4-Of-Robots-and-Rules.pdf

Raso, F.A., Hilligoss, H., Krishnamurthy, V., Bavitz, C., y Kim, L. (2018). Artificial intelligence & human rights: Opportunities & risks. Berkman Klein Center Research Publication, 2018(6), 1-62. https://doi.org/http://dx.doi.org/10.2139/ssrn.3259344

Salmón, E. (2012). Introducción al derecho internacional humanitario. Instituto de Democracia y Derechos Humanos de la Pontificia Universidad Católica del Perú y el Comité Internacional de la Cruz Roja.

Samek, W., Wiegand, T., y Mülle, K. R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv, 1-8. https://doi.org/10.48550/arXiv.1708.08296

Saria, O., y Celik, S. (2021). Legal evaluation of the attacks caused by artificial intelligence-based lethal weapon systems within the context of Rome Statute. Computer Law & Security Review, 42, 1-15. https://doi.org/10.1016/j.clsr.2021.105564

Sassóli, M. (2014). Autonomous weapons and international humanitarian law: Advantages, open technical questions and legal issues to be clarified. International Law Studies, 90(1). https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=1017&context=ils

Schmitt, M. (2013). Autonomous weapon systems and international humanitarian law: A reply to the critics. Harvard National Security Journal, 1-37. https://doi.org/http://dx.doi.org/10.21

Schmitt, M.N., y Thurnher, J.S. (2013). Out of the loop: Autonomous weapon systems and the law of armed conflict. Harvard National Security Journal, 14, 231-281. https://harvardnsj.org/wp-content/uploads/2013/01/Vol-4-Schmitt-Thurnher.pdf

Sehrawat, V. (2021). Autonomous weapon system and command responsibility. Florida Journal of International Law, 31(3). https://scholarship.law.ufl.edu/fjil/vol31/iss3/2

Sharkey, N.E. (2012). The evitability of autonomous robot warfare. International Review of the Red Cross, 94(886), 787-799. https://www.icrc.org/en/doc/assets/files/review/2012/irrc-886-sharkey.pdf

Soori, M., Arezoo, B., y Dastres, R. (2023). Artificial intelligence, machine learning and deep learning in advanced robotics: A review. Cognitive Robotics, 3, 54-70. https://doi.org/10.1016/j.cogr.2023.04.001

Szpak, A. (2019). Legality of use and challenges of new technologies in warfare – The use of autonomous weapons in contemporary or future wars. Cambridge University Press, 28(1), 118-131. https://doi.org/10.1017/S1062798719000310

Thurnher, J.S. (2012). No one at the controls: Legal implications of fully autonomous targeting. Joint Force Quarterly, 67(4), 77-84. https://ndupress.ndu.edu/Portals/68/Documents/jfq/jfq-67/JFQ-67_77-84_Thurnher.pdf

Thurnher, J.S. (2016). Means and methods of the future: Autonomous systems. In P. Ducheine, M. Schmitt, y F. Osinga (Eds.), Targeting: The challenges of modern warfare (pp. 177-199). T.M.C. Asser Press. https://doi.org/10.1007/978-94-6265-072-5_9

UK Ministry of Defense. (2022). Ambitious, safe, responsible: Our approach to the delivery of AI-enabled capability in defence. https://www.gov.uk/government/publications/ambitious-safe-responsible-our-approach-to-the-delivery-of-ai-enabled-capability-in-defence

UN General Assembly. (1998). Estatuto de Roma de la Corte Penal Internacional. https://www.refworld.org.es/docid/50acc1a12.html

Vigevano, M.R. (2021). Inteligencia artificial aplicable a los conflictos armados: Límites jurídicos y éticos. Arbor: Ciencia, Pensamiento y Cultura, 197(800), 1-13. https://doi.org/10.3989/arbor.2021.800002

Contribución de los autores (Taxonomía CRediT): el único autor fue responsable de la: 1. Conceptualización, 2. Curación de datos, 3. Análisis formal, 4. Adquisición de fondos, 5. Investigación, 6. Metodología, 7. Administración de proyecto, 8. Recursos, 9. Software, 10. Supervisión, 11. Validación, 12. Visualización, 13. Redacción - borrador original, 14. Redacción - revisión y edición.

Disponibilidad de datos: El conjunto de datos que apoya los resultados de este estudio no se encuentra disponible.

Editor responsable Miguel Casanova: mjcasanova@um.edu.uy