

INTELIGENCIA ARTIFICIAL APLICABLE A LOS CONFLICTOS ARMADOS: LÍMITES JURÍDICOS Y ÉTICOS

Marta R. Vigevano

Universidad de Buenos Aires
ORCID: 0000-0002-5294-6901
martavigevano@derecho.uba.ar

Cómo citar este artículo/Citation: Vigevano, Marta R. (2021). Inteligencia artificial aplicable a los conflictos armados: límites jurídicos y éticos. *Arbor*, 197(800): <https://doi.org/10.3989/arbor.2021.800002>

Recibido: 21 febrero 2021. Aceptado: 2 junio 2021.
Publicado:

RESUMEN: Las normas del Derecho Internacional Humanitario (DIH) establecen límites al uso de los medios y métodos de combate en el desarrollo de las hostilidades. Si bien en su origen el DIH no fue elaborado teniendo en cuenta los desafíos planteados por el desarrollo de la inteligencia artificial (IA) en ese contexto, es una realidad que en la actualidad la evolución de esta inteligencia, los algoritmos y su aplicación militar emergente constituyen un desafío a la luz de las normas humanitarias. Ese desafío comprende tres ámbitos fundamentales: el técnico, el legal y el ético.

Si bien se puede decir que la IA, en la etapa actual de desarrollo, permite que un programa computerizado basado en algoritmos realice ciertas tareas en un entorno complejo e incierto, a menudo con mayor precisión que los seres humanos, también debemos destacar que no hay tecnología que haga que una máquina se comporte como un ser humano, quien puede determinar si una acción es lícita o ilícita y decidir no continuar con lo programado, teniendo como objetivo primordial la protección de las víctimas.

Los Estados deben adoptar regímenes de verificación, prueba y monitoreo como parte del proceso para determinar e imponer limitaciones o prohibiciones de acuerdo a los principios esenciales de distinción y proporcionalidad que el DIH establece en la utilización de armas durante los conflictos armados internacionales o no internacionales. Tanto desde una perspectiva legal como ética el ser humano está en el centro de esta problemática, ya que la responsabilidad por el uso de la fuerza no puede transferirse a sistemas de armas o algoritmos, y sigue siendo una responsabilidad humana.

PALABRAS CLAVE: Inteligencia artificial (IA), conflictos armados, derecho internacional humanitario (DIH), selección de objetivos, ética.

ARTIFICIAL INTELLIGENCE IN ARMED CONFLICTS: LEGAL AND ETHICAL LIMITS

Copyright: © 2021 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia de uso y distribución *Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0)*.

ABSTRACT: The rules of International Humanitarian Law (IHL) set limits on the use of the means and methods of combat in the conduct of hostilities. While in its origins IHL was not developed taking into account the challenges posed by Artificial Intelligence in this context, it is a reality that today the evolution of this intelligence, the algorithms and their emerging military application constitute a challenge in the light of humanitarian standards. This challenge comprises three fundamental aspects: legal, technical and ethical.

While it can be said that AI, in the current stage of development, allows a computer programme based on algorithms to perform certain tasks in a complex and uncertain environment, often with greater accuracy than humans, we must also stress that there is no technology that makes a machine behave like a human being who can determine whether an action is lawful or unlawful and decide not to proceed with the programmed action, with the protection of victims as the primary objective. This is one of the dominant themes in doctrinal debates on the application of IHL to means and methods of combat involving AI-related techniques.

States must adopt verification, testing and monitoring systems as part of the process to determine and impose limitations or prohibitions in accordance with the essential principles of distinction and proportionality that IHL establishes in the use of weapons during international or non-international armed conflicts. Moreover, it is worth noting that from a legal as well as an ethical perspective, the human being is at the center of this issue, since the responsibility for the use of force cannot be transferred to weapons systems or algorithms, as it remains a human responsibility.

KEYWORDS: Artificial Intelligence (AI), armed conflicts, international humanitarian law, targeting, ethic.

INTRODUCCIÓN

La inteligencia artificial (IA) se ha expandido en las últimas décadas en varios ámbitos, entre ellos en el desarrollo de tecnología militar utilizable en los conflictos armados. Su aplicación crea cierta incertidumbre sobre conceptos como la atribución, el control y la responsabilidad, todos estrechamente vinculados al factor humano y al principio de que los Estados deben respetar y hacer respetar el Derecho Internacional Humanitario (DIH), así como los Derechos Humanos, ya que ambos cuerpos normativos tienen por fin la protección de la vida y la dignidad humana en tiempos de guerra y en tiempos de paz. Esta obligación es plenamente aplicable al posible proceso de desarrollo, adquisición, uso o transferencia de la tecnología militar desarrollada por la inteligencia artificial. El interés de este trabajo es determinar en primera instancia qué capacidades especiales y avances tecnológicos plantea la IA en el contexto bélico y cuáles pueden ser sus aspectos desfavorables en el marco de la aplicación de las normas jurídicas del DIH.

Con este fin presentaré en primer lugar una breve introducción a las normas del DIH que rigen las hostilidades. Esencialmente tres son las normas básicas que deben ser respetadas durante el desarrollo de las hostilidades: la distinción, la proporcionalidad y la precaución en el ataque a los efectos de evitar causar males superfluos o sufrimientos innecesarios.

En segundo lugar se planteará el concepto de IA y su evolución, caracterizada por la combinación del reconocimiento de imagen, el procesamiento de lenguaje y las redes neuronales. Seguidamente, expondré las aplicaciones de estas tecnologías a los conflictos armados, así como sus usos en otros ámbitos, como la ayuda humanitaria.

Por último, se realizará el análisis de la aplicación del DIH a esta nueva tecnología, subrayando la necesidad de que el centro de atención esté puesto en las responsabilidades y obligaciones que incumben a los seres humanos en las decisiones relativas a la utilización del uso de fuerza armada.

1. EL DIH Y EL ESTABLECIMIENTO DE MEDIOS Y MÉTODOS DE COMBATE EN LOS CONFLICTOS ARMADOS

El DIH, como sistema jurídico que regula los conflictos armados, establece la limitación de los medios y métodos de combate y la protección de las víctimas, los bienes y el medioambiente que puedan ser afectados. Se consideran *medios* las armas utilizadas y la manera de usarlas, y *métodos* la conducta general de quienes participan en el desarrollo de las hostilidades. Según el jurista suizo Jean Pictet (1966: 456, traducción de la autora) el término *derecho humanitario internacional*, «en un sentido amplio, es el conjunto de disposiciones jurídicas internacionales escritas o consuetudinarias, que aseguran el respeto de la persona humana». Este comprende el Derecho de la Haya, que establece los deberes y derechos de los beligerantes en la conducción de las operaciones de la guerra y el límite a la elección de los medios de combate¹, así como el Derecho de Ginebra, denominado el derecho humanitario propiamente dicho, que tiende a proteger a los militares fuera de combate y a las personas que no participan en las hostilidades.

Se considera 1864 la fecha de nacimiento de este derecho con la creación del primer instrumento multilateral del Derecho Internacional Humanitario: el Convenio de Ginebra de 22 de agosto de 1864 para el mejoramiento de la suerte de los militares heridos en los ejércitos en campaña (Swinarski 1990: 16)².

Las restricciones y prohibiciones al empleo de ciertas armas datan de varios siglos atrás. Como sugiere Matthias Maas (2019: 7), «el desarrollo y la proliferación de nuevas tecnologías militares permitieron una brutalidad imprevista en varias guerras sistémicas, que motivaron desarrollos clave en el derecho internacional»³. Las primeras normas aplicables a la guerra estuvieron destinadas a regular el uso de armas específicas para evitar daños y sufrimientos desproporcionados. En el siglo XIX la *Declaración de San Petersburgo de 1864 con el objeto de prohibir el uso de determinados proyectiles en tiempo de guerra*

1 Estas normas jurídicas se estipulan en la Convención de la Haya de 1899, revisadas por la Convención de la Haya de 1907.

2 Después de 1949 el DIH se concreta convencionalmente en: I. Convenio de Ginebra para aliviar la suerte que corren los heridos y los enfermos de las fuerzas armadas en campaña, 1949; II. Convenio de Ginebra para aliviar la suerte que corren los heridos, los enfermos y los náufragos de las fuerzas armadas en el mar, 1949; III. Convenio de Ginebra relativo al trato debido a los prisioneros de guerra, 1949; IV. Convenio de Ginebra relativo a la protección debida a las personas civiles en tiempo de guerra, 1949. Protocolo I adicional a los Convenios de Ginebra de 1949 relativo a la protección de las víctimas de los conflictos armados internacionales, 1977; Protocolo II adicional a los Convenios de Ginebra de 1949 relativo a la protección de las víctimas de los conflictos armados sin carácter internacional, 1977.

3 Maas afirma que la brutalidad de Guerra de los 30 años sería la causa de que en el Tratado de Westfalia de 1648 se intentara poner fin a determinadas prácticas de guerra.

establece por primera vez con un carácter general⁴ que el objetivo legítimo de la guerra era «el debilitamiento de las fuerzas militares del enemigo», para lo cual bastaba «poner fuera de combate al mayor número posible de hombres», y que ese fin quedaría superado si se emplearan armas que agravaran inútilmente los sufrimientos de los hombres puestos fuera de combate, o que provocaran una muerte inevitable, por lo que el empleo de esas armas sería contrario a las leyes de la humanidad (Declaración de San Petersburgo, 29 de noviembre/11 de diciembre de 1868, Preámbulo).

Sobre la base de los principios de distinción y de precaución en el ataque para evitar males superfluos e innecesarios⁵, el DIH prohíbe de manera convencional y consuetudinaria el empleo de armas de efectos indiscriminados, es decir, armas que no pueden dirigirse contra un objetivo militar concreto, o cuyos efectos no sea posible limitar y que, en cualquiera de los casos, pueden alcanzar tanto a objetivos militares y a personas civiles como a bienes de carácter civil. Quedan comprendidos dentro de esta prohibición los sistemas de armas que, debido a la característica propia de la tecnología empleada y el uso para el que fueron diseñadas, cabe prever que causen daños incidentales excesivos a la población civil (Melzer, 2019: 122-123).

Estos principios se aplicaban en la conducción de hostilidades aún antes de la existencia de DIH como tal, y han sido *a posteriori* el fundamento de la elaboración de una serie de tratados que regulan el empleo de armas específicas. Algunas de ellas han sido explícitamente prohibidas por las convenciones internacionales, entre otras: las balas expansivas dum-dum (Declaración prohibiendo el empleo de balas que se ensanchan de La Haya, 1899); las armas biológicas (Convención sobre la prohibición del desarrollo, la producción y el almacenamiento de armas bacteriológicas (biológicas) y tóxicas y sobre su

destrucción, 1972); las armas químicas (Declaración de La Haya de 1899 relativa a los gases asfixiantes, el Protocolo de Ginebra, 1925, Convención sobre armas químicas, 1993); las armas láser cegadoras (Convención sobre ciertas armas convencionales, Protocolo IV, 1995); las minas antipersona (Convención sobre la prohibición de las minas antipersona, 1997); las municiones de racimo (Convención sobre municiones de racimo, 2008); las armas nucleares (Tratado sobre la prohibición de armas nucleares, 2017). Las disposiciones contenidas en estos acuerdos internacionales mencionados están inspiradas en el necesario equilibrio que debe existir entre la necesidad militar y las consideraciones de humanidad⁶.

Los Estados no pueden optar por cualquier medio o método de combate⁷, ya sea porque estén prohibidos expresamente o porque su capacidad destructiva pueda causar muy graves sufrimientos humanos y producir «un daño mayor al que sería inevitable para alcanzar objetivos militares legítimos» (Corte Internacional de Justicia, *Legalidad de la amenaza o el empleo de armas nucleares* (opinión consultiva), 1996, párr. 97)⁸. Además, se considera que la prohibición de empleo de armas que causen males superfluos o sufrimientos innecesarios y cuyos efectos sean indiscriminados son consideradas normas de carácter consuetudinario y como tales son oponibles a todos los Estados (Henckaerts y Doswald-Beck, 2007: 265-279).

En todo caso, y de conformidad con el artículo 36 del Protocolo Adicional I, relativo a la Protección de las Víctimas de los Conflictos Armados Internacionales, los Estados tienen la obligación, cuando desarrollen, estudien o adopten un arma nueva, de considerar que el empleo de la misma, ya sea en ciertas condiciones o en cualquier circunstancia, no esté prohibido por las normas del DIH, ya sean convencionales o consuetudinarias (véase Gutiérrez Posse, 2003: 34).

4 La Declaración de San Petersburgo no tenía carácter convencional obligatorio, pero en su elaboración participaron la mayoría de los Estados con mayor poderío armamentístico de esa época, lo cual puso de relieve que existía entre ellos un consenso importante sobre estas cuestiones.

5 El principio de distinción implica que las partes beligerantes deben hacer distinción en todo momento entre la población civil y los combatientes, y entre los bienes de carácter civil y los objetivos militares, y deben dirigir sus operaciones únicamente contra los objetivos militares (ver Protocolo I, Art. 48). En virtud del principio de precaución para evitar males superfluos o innecesarios, enunciado en el artículo 57, del Protocolo I, el DIH restringe o prohíbe el uso de ciertas armas por tener efectos que se consideran excesivamente crueles sin importar las circunstancias.

6 Los textos de las disposiciones relativas a las balas dum-dum, biológicas, químicas, láser cegador, minas antipersona, municiones en racimo y armas nucleares pueden encontrarse, respectivamente, en las pp. 99ss, 135ss, 95ss, 212ss, 285ss y 127ss de Schindler, Dietrich y Toman (1981).

7 Artículo 35- "1. En todo conflicto armado, el derecho de las Partes en conflicto a elegir los métodos o medios de hacer la guerra no es ilimitado" (Protocolo Adicional I a los Cuatro Convenios de Ginebra de 1949, 1977).

8 Esta Opinión Consultiva emanada de la CIJ puso en relieve el carácter de destrucción masiva de las armas nucleares, antes de la adopción del actual Tratado que las prohíbe, avalando desde lo jurisprudencial los principios de distinción y de evitar males innecesarios y superfluos.

Esto implica una restricción a los importantes avances tecnológicos de las últimas décadas, incluida la IA, ya que toda nueva tecnología de guerra debe utilizarse respetando las normas vigentes del DIH. Del artículo 36 del Protocolo Adicional I, 1977 —que establece que cada Estado Parte debe verificar si el empleo de los medios o métodos de combate que estudie, desarrolle, adquiera o adopte son compatibles con las normas de derecho internacional— se desprende, efectivamente, que la obligación de examinar la licitud se aplica a *todas* las nuevas armas. El Protocolo I, sin embargo, no establece cómo debe determinarse la licitud de las armas, los medios y los métodos de guerra, por consiguiente, son los Estados Parte los que deben tomar las medidas administrativas, legislativas y reglamentarias que permitan el cumplimiento de los principios esenciales de distinción y proporcionalidad. El Comité Internacional de la Cruz Roja (CICR) ha elaborado una guía donde se desarrolla el ámbito material de aplicación y los aspectos funcionales del examen que los Estados deben seguir para establecer o mejorar los procedimientos para determinar la licitud de las nuevas armas, de conformidad con el artículo 36 del Protocolo adicional I a los Convenios de Ginebra de 1949.

Este examen incorpora cuestiones de procedimiento que comprenden: determinar la autoridad nacional responsable de la prueba, las instancias institucionales que deben participar en el proceso, los mecanismos relativos a la toma de decisiones y el registro de exámenes. El mecanismo que se adopte, según esta guía, se debe basar en un enfoque imparcial y multidisciplinario de los exámenes jurídicos de las nuevas armas y que los Estados procedan a intercambiar información sobre los procedimientos. La guía debe servir para determinar la licitud de las armas en el marco de las prohibiciones y las restricciones de carácter convencional y consuetudinario, así como las normas generales de DIH para poder determinar sus efectos en la población civil, en los combatientes, en la salud y en el medio ambiente. (CICR/ICRC, 2006). La IA aplicable al desarrollo de las hostilidades no puede sustraerse de este proceso de examen jurídico y técnico.

2. ¿QUÉ SE ENTIENDE POR INTELIGENCIA ARTIFICIAL?

La IA es una disciplina científica que pone en práctica métodos destinados a la construcción de maquinarias inteligentes capaces de actuar con una previsión adecuada en un entorno complejo, es decir, utilizando modelos entrenados a partir de datos. La Comisión Europea considera que:

«el término «inteligencia artificial» (IA) se aplica a los sistemas que manifiestan un comportamiento inteligente, pues son capaces de analizar su entorno y pasar a la acción—con cierto grado de autonomía— con el fin de alcanzar objetivos específicos. Los sistemas basados en la IA pueden consistir simplemente en un programa informático (p. ej. asistentes de voz, programas de análisis de imágenes, motores de búsqueda, sistemas de reconocimiento facial y de voz), pero la IA también puede estar incorporada en dispositivos de *hardware* (p. ej. robots avanzados, automóviles autónomos, drones o aplicaciones del internet de las cosas)» (COM/2018/237 final, 2018: 1).

Para la gran mayoría de las aplicaciones actuales, la IA consiste en algoritmos⁹ que forman la base del *software* de reconocimiento de patrones. Cuando se combina con la potencia informática de alto rendimiento, los científicos de datos (*data scientist*) pueden buscar y encontrar fuentes de información en colecciones masivas de datos (*big data*). Las redes neuronales¹⁰ potencian la capacidad de los algoritmos para identificar y organizar patrones en los datos, entrenándolos de esta manera para asociar patrones específicos con los resultados deseados (Ledesma Orozco, 2011: 187-188). Los algoritmos realizan comparaciones incrementales, a los efectos de reducir la cantidad de coincidencias en ciclos sucesivos. Múltiples capas de redes neuronales, conocidas como redes neuronales de aprendizaje profundo (*deep learning*), son las que hacen posibles los enfoques actuales de «aprendizaje automático», «aprendizaje supervisado» y «aprendizaje por refuerzo» (Schmidhuber, 2015: 85-117)¹¹. En el ámbito del aprendizaje automático o *machine learning* (ML), los algoritmos gestionan la información del entorno de tal manera que las computadoras aprendan a tomar decisiones sin la necesidad de ser programadas explíci-

9 Las definiciones del término son variadas. Tal vez la más intuitiva y comprehensiva es la que propone Monasterio: «Un código *software* que procesa un conjunto limitado de instrucciones» (Monasterio, 2017: 186).

10 Las redes neuronales permiten buscar la combinación de parámetros que mejor se ajusta a un determinado problema. Una red neuronal artificial es un método de computación inspirado en modelos biológicos que trata de imitar la operación básica de las neuronas cerebrales.

11 Es de destacar que dado que estamos analizando la utilización de la IA en el ámbito de las situaciones de violencia armada, es necesaria una supervisión continua por parte de los Estados de los componentes informáticos adoptados en el empleo de medios y métodos de combate que impliquen técnicas o herramientas relacionadas con la IA.

tamente. Con los algoritmos de *aprendizaje supervisado* se genera un modelo predictivo, basado en datos de entrada y salida; éstos se organizan en base a un conjunto de datos previamente etiquetado y clasificado, es decir, es preciso partir de un conjunto de muestra, del cual ya se sabe a qué grupo, valor o categoría pertenecen los ejemplos. Con este grupo de datos, llamados *datos de entrenamiento*, se realiza el ajuste al modelo inicial planteado. Es así como el algoritmo va *aprendiendo* a clasificar las muestras de entrada comparando el resultado del modelo y la etiqueta real de la muestra, realizando las compensaciones respectivas al modelo de acuerdo a cada error en la estimación del resultado.

Los algoritmos de *aprendizaje no supervisado* trabajan de forma muy similar a los supervisados, con la diferencia de que estos sólo ajustan su modelo predictivo tomando en cuenta los datos de entrada, sin importar los de salida. Un ejemplo de sistema con aprendizaje no supervisado lo constituyen los sistemas topográficos. Estos sistemas se caracterizan por tener como entrada una fotografía aérea y como salida la elaboración de un mapa (Cáceres Tello, 2007: 1-6).

Los algoritmos de *aprendizaje por refuerzo* definen modelos y funciones enfocados en maximizar una mejora de respuesta usando un proceso de retroalimentación. Este algoritmo es el más cercano a la psicología conductista de los humanos, ya que es un modelo acción-recompensa, que busca que el algoritmo se ajuste a acciones a tomar que puedan obtener una mejor recompensa (Simeone, 2018: 648-650)¹².

La IA trabaja en diferentes campos como la *representación del conocimiento*, cuyo objetivo es definir mecanismos para representar el conocimien-

to por medio de símbolos y codificar el pensamiento humano para ser utilizado computacionalmente; la búsqueda heurística (cualquier técnica que mejore la búsqueda del proceso de solución de problemas); el procesamiento de lenguaje natural (campo de la IA que se ocupa de investigar la manera de comunicar las máquinas con las personas mediante el uso de las diversas lenguas naturales) y el ya mencionado aprendizaje automático.

Como ya se ha mencionado, uno de los aspectos que la IA plantea es la innovación en los métodos de aprendizaje de las máquinas (*machine learning*) a través de métodos de análisis de datos que utilizan algoritmos para construir un programa informático para mejorar su rendimiento predictivo¹³. Una de las principales necesidades es proporcionar a la máquina autónoma un contexto sobre el entorno y un objetivo. Para que un vehículo autónomo o un robot inteligente pueda actuar en un ámbito concreto es necesario integrar tecnologías de IA con otras muchas que proceden de otras disciplinas. Así, para obtener información del entorno es necesario disponer de sensores adecuados, de lo contrario sus algoritmos por muy potentes que sean no podrán interpretarlo para tomar decisiones; pero, por otra parte, si no dispone de fuentes de energía tampoco podrá tener un carácter autónomo. Esta sinergia de todas las tecnologías es la que permite que realice su función en un ámbito de aplicación específico (Serrano, 2019: 42-43).

Los trabajos desarrollados en inteligencia artificial han permitido evolutivamente categorizar sus productos bajo cuatro enfoques: sistemas que piensan racionalmente, comprende los sistemas expertos¹⁴;

12 Los agentes de aprendizaje profundo se han utilizado para tomar el lugar de un jugador humano en juegos como ajedrez, shogi o go de gran exigencia para el intelecto humano. El AlphaZero es una versión modificada de AlphaGo Zero que puede jugar shogi, ajedrez y go. El agente modificado comienza con solo información básica de las reglas del juego, y también se entrena completamente a través del auto aprendizaje.

13 Dentro del campo de *machine learning* hay un derivado que es el *deep learning* que se basa en la idea de aprendizaje exclusivamente desde el entrenamiento. Es así que, en lugar de que el programador incorpore al sistema una lista de normas para resolver el problema, se le da directamente un modelo para que pueda, mediante ejemplos, realizar la evaluación de imagen, texto, video y audio. Por ejemplo en el caso de las operaciones bancarias los ordenadores y el *software* analizan las transacciones, hora de las mismas, importe y destinatarios, de esta forma pueden detectar cuando algún movimiento extraño está fuera de la estadística de un cliente.

14 También se puede mencionar como un campo de la IA, a partir de la década del 60, el desarrollo de sistemas expertos, tal como el Dendral, cuyo propósito era interpretar la estructura molecular, que se divide en tres partes funcionales: plan, generación y prueba, desarrollado en la Universidad de Stanford en la década de 1960 durante casi diez años. Planteaba la posibilidad del uso de ordenadores para modelar el pensamiento científico que luego fue seguido por otros sistemas en diversos ámbitos como la medicina, la economía, las finanzas, los sistemas de defensa y seguridad. De acuerdo con Badaró, Ibáñez y Agüero, «un sistema experto es un sistema que emplea conocimiento humano capturado en una computadora para resolver problemas que normalmente requieren de expertos humanos. Los sistemas bien diseñados imitan el proceso de razonamiento que los expertos utilizan para resolver problemas específicos. Dichos sistemas pueden ser utilizados por no-expertos para mejorar sus habilidades en la resolución de problemas. Los sistemas expertos también pueden ser utilizados como asistentes. Además, estos sistemas pueden funcionar mejor que cualquier humano experto individualmente, tomando decisiones en una específica y acotada área de pericia, denominada como dominio» (Badaró, Ibáñez y Agüero, 2013: 351).

sistemas que piensan como humanos, comprende las redes neuronales; sistemas que actúan racionalmente, comprende los agentes inteligentes; sistemas que actúan como humanos, comprende la robótica (Russell y Norvig, 2009: cap. 1). Dentro de cada uno de estos enfoques pueden encontrarse aplicaciones de uso bélico, pero sin duda el último de ellos, la robótica, es de particular interés a la hora de establecer las implicaciones de la IA en los conflictos armados.

Cabe distinguir entre robots inteligentes y robots industriales. Como señala Adriana Margarita Porcelli (2021: 10-11), una de las definiciones más completas del término *robot industrial* es aquella que da la International Organization for Standardization (ISO) en la revisión de 2012 de la 8373, según la cual el robot industrial es «un manipulador multiusos, reprogramable y controlado automáticamente, programable en tres o más ejes, que pueden estar fijos o móviles para uso en aplicaciones de automatización industrial»¹⁵. Esta definición es aceptada por la International Federation of Robotics (IFR, 2017: 1). El robot inteligente está provisto, necesariamente, de las tecnologías de información e IA que hemos descrito, y comprende desde los drones y los coches autónomos, a los robots *software*, como los *chatbots*.

Por *robots autónomos letales*, que son los que plantean conflictos en el ámbito del DIH, se entiende «sistemas de armas robóticas que, una vez activados, pueden seleccionar y atacar objetivos sin necesidad de intervención adicional de un operador humano (Heyns, 2013: 8). Lo definitorio de un arma letal autónoma es que el robot puede optar sin intervención humana por seleccionar o no el objetivo y decidir sobre el empleo o no de fuerza letal, «es decir, el objeto a atacar está determinado por el procesamiento de los sensores, no por humanos» (Heyns, 2013: 8). El CICR ha definido los sistemas de armas autónomos como «cualquier sistema de armas con autonomía en sus funciones críticas. Es decir, un sistema de armas que puede seleccionar (esto es, buscar o detectar, identificar, rastrear, seleccionar) y atacar (es decir, usar la fuerza contra, neutralizar, dañar o destruir) objetivos sin la intervención humana» (CICR/ICRC, 2016: 1, traducción de la autora). Human Right Watch (HRW) denomina al sistema que puede seleccionar y atacar objetivos sin intervención humana *Human-out-of-the-Loop Weapon*, y comprende en este concepto los sistemas de armas completamente autónomas, las

armas letales autónomas y los robots asesinos (HRW y Harvard Law School's International Human Rights Clinic (IHRC), 2012).

El Instituto Internacional de Estudios para la Paz de Estocolmo en un Informe elaborado sobre sistemas armamentísticos autónomos letales (LAWS, por sus siglas en inglés) (SIPRI, 2017: 8) desarrolla el concepto de estas armas bajo diversas categorías: la primera consiste en una definición fundada sobre la naturaleza de la relación de mando y control que se establece en el sistema hombre-máquina; por lo tanto, esta perspectiva incluye aquellas armas que, una vez activadas, pueden seleccionar y atacar objetivos sin participación humana. La segunda categoría abarca las definiciones basadas en parámetros de capacidad, con lo que se contempla a estas máquinas como «armas capaces de comprender un nivel superior de intención y dirección» y a partir de esa comprensión emprender la acción adecuada para lograr el estado deseado sin depender de supervisión humana. La tercera categoría de definiciones son aquellas establecidas según criterios jurídicos, «haciendo hincapié en la naturaleza de las tareas realizadas en forma autónoma», es decir sustituyendo al ser humano en funciones críticas como el uso de la fuerza, selección, neutralización, daño o destrucción de objetivos (traducción de la autora).

Es interesante, en este contexto, la clasificación que hace el Ministerio de Defensa de Estados Unidos de las armas autónomas de acuerdo al grado de intervención humana, y por tanto dentro de la primera categoría descrita, según la cual existen tres tipos: a) aquellas armas cuyos sistemas requieren un comando humano para seleccionar y atacar objetivos (las denomina semiautónomas); b) el arma cuyo sistema *per se* selecciona y ataca a los objetivos, pero lo hace bajo la supervisión de un operador humano (las denomina armas autónomas supervisados por humanos) y c) el sistema de arma que puede seleccionar y atacar objetivos sin ningún tipo de control humano (Department of Defense, Directive nº 3000.09, 2012 (Glossary Part II, pp. 13-14).

En todo caso, estos sistemas han encontrado resistencia por parte de grupos civiles. En noviembre de 2019, la *Campaign to Stop Killer Robots* elaboró un documento sobre los elementos clave para un tratado sobre armas totalmente autónomas enfatizando el control humano y rechazando cualquier tipo de arma

15 <https://www.iso.org/obp/ui/#iso:std:iso:8373>. La traducción es nuestra.

totalmente autónoma¹⁶. En este trabajo sostenemos que no puede concebirse la puesta en marcha de un sistema de medios de combate que prescindiera del continuo control humano. Estas consideraciones no están exentas de un análisis desde el ámbito de la protección de los Derechos Humanos, como se referirá más adelante.

3. LA IA Y LAS NUEVAS TECNOLOGÍAS APLICABLES EN LOS CONFLICTOS ARMADOS

Tal como se ha expresado, la IA se ha desarrollado en los últimos años gracias a importantes avances tecnológicos como la disponibilidad de grandes cantidades de datos, el incremento de la potencia de procesamiento de los ordenadores y la innovación en los métodos de aprendizaje automático.

La IA tiene aplicación en varios aspectos de vital importancia en la conducción de hostilidades. En este contexto, se utiliza en tareas de reconocimiento de objetivos militares, vigilancia, comunicación, logística, en la manipulación de la información y el desarrollo de nuevas armas y los métodos de combate. Sin embargo, esta misma tecnología también encuentra su uso humanitario en la protección de las víctimas. En el ámbito terrestre, por ejemplo, se utilizan artefactos autónomos en el seguimiento de vehículos o, para evitar obstáculos, coordinando sistemas no tripulados con vehículos tripulados. En el ámbito marítimo esta tecnología posibilita que los navíos de superficie no tripulados, ligeros y pequeños, estén equipados con instrumentos de barrido de minas marinas. En el ámbito aéreo, son utilizados con fines humanitarios los sistemas autónomos de aeronaves de vigilancia o reconocimiento. En cuanto a las redes de información, el caso del acceso a una gran cantidad de datos y conocimiento en el ciberespacio permite a individuos y grupos un acceso inmediato a recursos estratégicos.

El uso militar de la IA plantea, sin embargo, varias preguntas esenciales con relación al método de análisis de datos que utilizan los algoritmos: ¿de dónde provienen esos datos?, ¿qué sesgos podrían tener? y ¿cómo esos sesgos podrían afectar la eficacia del modelo? Además en el contexto de un conflicto armado se hace necesario hacer juicios de percepción de manera rápida sobre grandes cantidades de contenido

visual, en las que los datos que se utilizan para preparar el algoritmo pueden discrepar a veces de los datos reales, o puede incluso haber una falta de ellos.

Por otra parte, pueden existir situaciones en las que se requiere hacer juicios complejos y en las que resulta difícil codificar nociones de DIH en los algoritmos. En ocasiones se hace complicado caracterizar el concepto de conflicto armado, al igual que la noción de población civil, la de combatiente y prisionero de guerra, la de un bien civil utilizado para un fin militar, o la participación de la población civil en las hostilidades; y esta complejidad o ambigüedad pueden implicar riesgos al generar identificaciones erróneas. Si los datos sesgados ingresan inicialmente en el sistema, no solo se traducirán en recomendaciones futuras, también podrán expandirse y ampliar su alcance a medida que esas recomendaciones sesgadas se retroalimentan en el sistema¹⁷. El proceso de transformación de normas jurídicas en códigos implica inevitablemente elecciones particulares en cuanto a la forma de interpretar la ley, que pueden verse afectadas por circunstancias extrajurídicas de los desarrolladores de los programas (Perel y Elkin-Korel, 2016: 518). Es decir, las normas jurídicas han sido establecidas por seres humanos, deben ser interpretadas y aplicadas por ellos y no por máquinas.

Reafirmamos por ello un concepto ya expresado, toda nueva tecnología de uso militar aplicable a los conflictos armados debe utilizarse de conformidad con las normas existentes del DIH teniendo en cuenta los principios de distinción y proporcionalidad. Este es un requisito mínimo. Estos principios son transversales a la aplicación de todos los medios y métodos de combates. Es cierto que las características únicas de las nuevas tecnologías de guerra, las circunstancias de su uso, y sus consecuencias humanitarias pueden plantear dudas sobre si las normas existentes son suficientes o necesitan ser aclaradas o complementadas. Frente a estos cuestionamientos sugiero que las aplicaciones militares de las tecnologías nuevas y emergentes no son inevitables y podemos (y debemos) mantener ante ellas un criterio prudencial. Se trata, además, de decisiones tomadas por los Estados, que deben cumplir con las normas existentes y tener en cuenta las posibles consecuencias para los civiles y para los combatientes que ya no participan en las

16 *Campaign to Stop Killer Robots* (Campaña de Prohibición de Robots Asesinos) creada en 2012, es una alianza de Organizaciones No Gubernamentales (ONG) que trabaja para prohibir las armas completamente autónomas y mantener un control humano sobre el uso de la fuerza. Ver: <https://www.stopkillerrobots.org/about/?lang=es>

17 Un ejemplo interesante de dichos sesgos se encuentra en esta noticia del *New York Times* <https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html>

hostilidades, así como consideraciones más amplias de «humanidad» y «conciencia pública»¹⁸. La posibilidad de una autonomía completa de estas máquinas plantea serias preocupaciones bajo los principios de la humanidad y los dictados de la conciencia pública. El CICR subraya que los principios de la humanidad requieren compasión y la capacidad de proteger (CICR/ICRC, 1996), rasgos emocionales difíciles de imbuir en una máquina.

Hay aspectos de gran importancia en el uso de la IA, como la identificación, la fiabilidad y la previsibilidad, que hoy día aún no hay certeza de que puedan ser cumplidos por una máquina en situaciones de uso de fuerza armada. Durante el funcionamiento de un arma que selecciona objetivos sin intervención humana, la identificación plantea un nivel de supervisión, intervención y capacidad de desactivación humana. El nivel de previsibilidad estima el funcionamiento y las consecuencias del uso de este tipo de armamento. Por último, la fiabilidad se establece en relación a la probabilidad de fallas. Teniendo en cuenta estas variables, ¿puede la IA ayudar a proporcionar una imagen más completa y precisa de las ventajas que aporta en casos de ataques militares o de daños esperados a civiles u objetos civiles?

La respuesta a esta pregunta no es inmediata. Utilizar la IA para armas autónomas ofensivas requiere un análisis cuidadoso. Es preciso tener en cuenta desde perspectivas jurídicas y éticas los siguientes aspectos: a) las armas semiautónomas deben permitir a los mandos militares ejercer de manera plena un control efectivo a través del juicio humano; b) la responsabilidad de los mandos militares o subordinados no puede excluirse en el uso de armas semiautónomas; c) las empresas que se dedican al desarrollo de la IA en la elaboración de armas deben tener un control y revisión gubernamental estricto, y todos los programas de IA diseñados con fines hostiles deben ser supervisados por investigadores capacitados que confirmen su adecuación a las normativas (Houston, 2018).

Según Michel Schmitt y Jeffrey Thurnher, deben distinguirse las armas completamente autónomas

de aquellas que son semiautónomas, estas últimas frecuentes en las guerras contemporáneas (Schmitt y Thurnher, 2013). Pero para el jurista Marco Sassòli esa clasificación es relativa, ya que hoy en día ningún sistema de armas tiene las características de un arma completamente autónoma; por lo tanto, inevitablemente, los seres humanos siempre estarán involucrados en su funcionamiento (Sassòli, 2014: 309). La premisa de este último autor se basa en que las armas no se crean solas sino que son producto de la intervención humana. Por ello la IA aplicada a la elaboración de armas no debe ser utilizada de la misma manera que, en algunos casos, se utiliza cierta inteligencia humana: para eludir las normas o para decidir —desde una perspectiva puramente utilitaria— que el incumplimiento de las normas de DIH puede facilitar la consecución del objetivo principal de vencer al enemigo. Esta situación puede darse tanto en la elaboración de un arma semiautónoma como en un arma completamente autónoma, por lo que ambas representan el mismo riesgo.

En la etapa actual de desarrollo de la tecnología de IA, y a la luz de las normas del DIH, no se puede admitir que haya sistemas tomadores de decisiones totalmente autónomos en el campo de batalla. En un conflicto armado, muchas de estas decisiones son esenciales, lo que implica que su adopción puede traer aparejadas muertes o lesiones graves, daños o destrucción de la propiedad, o puede restringir las libertades individuales. Mantener el papel fundamental del ser humano en esas decisiones y el control sobre ellas será esencial para evitar consecuencias impredecibles tanto para civiles como para combatientes.

No sólo en el ámbito del DIH sino también en el sistema universal de protección de los Derechos Humanos se respalda indirectamente, en virtud del cumplimiento del art. 6 del Pacto de Derechos civiles y Políticos, la limitación de estas nuevas tecnologías¹⁹. Así queda expresado en el párrafo 12 de la *Observación general núm. 36 sobre el artículo 6 del Pacto Internacional de Derechos Civiles y Políticos, relativo al derecho a la vida*, presentada por el Comité de Derechos Humanos en 2017:

18 Explícitamente reconocidos en la Cláusula de Martens, que dice: «Mientras que se forma un Código más completo de las leyes de la guerra, las Altas Partes Contratantes juzgan oportuno declarar que, en los casos no comprendidos en las disposiciones reglamentarias adoptadas por ellas, las poblaciones y los beligerantes permanecen bajo la garantía y el régimen de los principios del Derecho de Gentes preconizados por los usos establecidos entre las naciones civilizadas, por las leyes de la humanidad y por las exigencias de la conciencia pública», en el Preámbulo del (II) Convenio de La Haya de 1899 relativo a las leyes y costumbres de la guerra terrestre. Esta cláusula debe su nombre al delegado de Rusia en la Conferencia de la Paz de La Haya de 1899.

19 El artículo 6 reconoce y protege el derecho a la vida de todos los seres humanos. Se trata del derecho supremo respecto del cual no se autoriza suspensión alguna, ni siquiera en situaciones de conflicto armado y otras emergencias públicas.

Los Estados partes que usan las armas existentes y estudian, desarrollan, adquieren o adoptan nuevas armas y nuevos medios o métodos de guerra deben tener siempre en cuenta su impacto sobre el derecho a la vida. Por ejemplo, el desarrollo, con vistas a su uso en operaciones militares, de nuevos robots autónomos letales que carecen de discernimiento y compasión humanos plantea difíciles cuestiones jurídicas y éticas en relación con el derecho a la vida, incluidas cuestiones relacionadas con la responsabilidad jurídica por su utilización. [El Comité considera, por tanto, que estos sistemas de armas no deberían [desarrollarse y] ponerse en funcionamiento, ni en tiempos de guerra ni en tiempos de paz, al menos no antes del establecimiento de un marco normativo que garantice su utilización conforme al artículo 6 y a otras normas pertinentes del derecho internacional].²⁰

Como se ha mencionado, tal vez la aplicación más amplia y de mayor alcance sea el uso de la IA y el aprendizaje automático para la toma de decisiones, que permite la recopilación y el análisis generalizados de fuentes de datos para identificar personas u objetos, evaluar patrones de vida o comportamiento, hacer recomendaciones para la estrategia u operaciones militares, o hacer predicciones sobre acciones o situaciones futuras. Uno de los aspectos que podrían considerarse más influyentes en el uso bélico de la IA son los mecanismos para generar información falsa como método de combate, pues esta información es cada vez más difícil de distinguir de la información real. El uso de estos sistemas por las partes en conflicto para amplificar los antiguos métodos de propaganda y manipular la opinión e influir en las decisiones podría tener implicaciones significativas sobre el terreno²¹.

Estos sistemas de apoyo a la toma de decisiones son una ampliación de las herramientas de inteligencia, vigilancia y reconocimiento utilizadas por la IA. El uso posible de estos sistemas es extremadamente amplio: desde para decidir estrategias militares —como a quién/es o qué atacar, a quién detener y durante cuánto tiempo— hasta los intentos de predecir o anticiparse al contendiente. Es importante relacionar estas aplicaciones con el principio de proporcionalidad, el principio de buena fe, la previsibilidad y la fiabilidad. Estos principios, que no pueden ser comprendidos por una máquina, deben de estar presentes en la toma de

decisiones, ya que en relación a ellos se establece la responsabilidad internacional imputable a un Estado por la violación de una norma jurídica exigible y la comisión de un hecho internacionalmente ilícito, así como con la responsabilidad penal individual ante la realización de crímenes de guerra. Dentro del ámbito penal internacional un elemento esencial es la existencia de la culpa y la pena, términos que no pueden aplicarse en el ámbito jurídico a un sistema de armas autónomas letales. El Consejo de Derechos Humanos de las Naciones Unidas, en el Informe del relator especial Christof Heyns, establece: «Los robots no tienen capacidad de discernimiento moral, por lo que si causan pérdidas de vida no se les puede exigir ningún tipo de responsabilidad, como sería normalmente el caso si las decisiones las hubieran tomado seres humanos. ¿En quién recae entonces la responsabilidad?» (Heyns, 2013: 16, párrafo 76). En el mismo informe se indica: «La responsabilidad legal puede recaer, entre otros, en los programadores informáticos, los fabricantes o vendedores de equipo, el mando militar que ordena el empleo, los subordinados que despliegan esos sistemas y los dirigentes políticos» (Heyns, 2013: 16, parágrafo 77).

Cabe preguntarse, para ahondar en el problema de la responsabilidad de un mando respecto a un arma autónoma letal, si ésta podría ser considerada un subordinado. Ante un subordinado humano el superior incurre en responsabilidad cuando no impidió el crimen o no aplicó las sanciones correspondientes a quien lo cometió. El superior jerárquico militar es responsable por su acción cuando ese sistema de armas letales incurre en un crimen de guerra. ¿Podría el superior militar conocer cabalmente los alcances técnicos y consecuencias de esa arma letal de tal manera que pudiese impedir su acción? A la luz de las normas sobre responsabilidad frente a hechos ilícitos podría, sí, imputarse esta responsabilidad al Estado, pero en cuanto a la responsabilidad penal estas violaciones quedarían impunes. Por lo tanto podría considerarse que si es imposible exigir responsabilidades de índole penal por las consecuencias del uso de un arma, su uso debe ser considerado ilícito.

Los sistemas de apoyo a la toma de decisiones basados en la IA y el aprendizaje automático deben contemplar que sean los seres humanos quienes tomen, en última instancia, decisiones a la hora de llevar a cabo las hostilidades; ya que si bien en determinadas

20 Accesible en: https://www.ohchr.org/Documents/HRBodies/CCPR/GCArticle6/GCArticle6_SP.pdf

21 Para el CICR, existe la preocupación de que los civiles puedan, como resultado de la desinformación digital o la información errónea, ser objeto de detención o malos tratos, discriminación o denegación de acceso a los servicios esenciales, o ataques contra su persona o propiedad.

circunstancias estos sistemas de apoyo pueden aportar mayor precisión y minimizar los riesgos para los civiles, nada garantiza que, teniendo en cuenta las posibles limitaciones de la tecnología como la imprevisibilidad y la fiabilidad, no tengan como efecto violaciones del DIH. Tal sería el caso de un sistema de IA utilizado para iniciar directamente un ataque, en lugar de producir un análisis que permita a los responsables humanos la toma de decisiones, ya que ese ataque a un objetivo militar —que es una acción lícita— no pueda ser abortado en el supuesto de que hubiera civiles que se encontraran incidentalmente en ese lugar (Davison, 2017: 5-18).

Otro aspecto a destacar en el análisis del uso de la IA en la toma de decisiones automatizadas desde lo jurídico y lo ético es el relativo a la detención de personas en el contexto de un conflicto armado, ya que pueden plantear problemas de inexactitud y parcialidad. Como ha expresado el CICR en más de una oportunidad, estos tipos de herramientas de IA podrían conducir a una creciente personalización de la guerra, con sistemas digitales que reúnen información personal identificable procedente de múltiples fuentes, bases de datos, comunicaciones, datos biométricos, informes sociales para formar una caracterización generada algorítmicamente sobre una persona, su estatus y su capacidad de ser blanco de ataques, o para predecir sus acciones futuras (CICR/ICRC, 2019: 5). El DIH contempla normas específicas respecto de las decisiones de las detenciones durante los conflictos armados que tienen como principal finalidad evitar los malos tratos, garantizar el acceso a la justicia y respetar todas y cada una de las categorías de víctimas protegidas por el DIH²². Si los sistemas inteligentes se entrenan con datos predominantemente sesgados hacia ciertas características, por ejemplo: de sexo, raza, etnia, corremos un grave riesgo de contravenir estas normas.

Mantener el papel fundamental del ser humano en esas decisiones y el control sobre ellas será esencial para evitar consecuencias impredecibles para los civiles y los combatientes.

4. LA INCIDENCIA DE LA ÉTICA EN LA IA

La IA, como ya señalamos, puede ser útil para la introducción de formas innovadoras de capturar y explotar la información digitalizada con la finalidad de asistencia

humanitaria digital. Algunas organizaciones humanitarias —por ejemplo, el Equipo Argentino de Antropología Forense (EAAF)²³— están estudiando proyectos pilotos centrados en áreas como la biometría para identificar a personas desaparecidas. La identificación biométrica se utiliza para verificar la identidad de una persona midiendo, de manera digital, determinadas características físicas y comparándolas con aquellas de la misma persona que se encuentran en una base de datos. También se emplean drones para llevar ayuda a lugares remotos y para determinar daños causados en poblaciones civiles. El CICR, por su parte, ha desarrollado «consolas de escaneo del entorno que utilizan IA y *machine learning* para capturar y analizar grandes volúmenes de datos con el fin de informar y apoyar su trabajo humanitario en contextos operativos específicos, incluyendo el uso de análisis predictivos para ayudar a determinar las necesidades humanitarias» (CICR/ICRC, 2019: 6, traducción del autor), tales como provisión de alimentos, agua, asistencia médica, refugio. Se puede por tanto, utilizar la IA como una inteligencia beneficiosa, destacando el elemento humano y de esta forma dotarla de una aceptación ética.

En efecto, como tantos gobiernos han declarado, el desarrollo de la IA debe estar de acuerdo con valores compatibles con la dignidad humana, los derechos y las libertades. Así, el Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial de la Comisión Europea subrayó la importancia de «la agencia y la supervisión humanas» de la IA. Por su parte, el Departamento de Defensa de Estados Unidos en su Estrategia de IA de 2018 para la Defensa hace referencia a la adopción reflexiva, responsable y centrada en el ser humano de la IA (Department of Defense USA, 2018: 17). En Francia, el Ministerio de Defensa se ha comprometido a utilizar la IA de acuerdo con tres principios rectores: el cumplimiento del derecho internacional, el mantenimiento de un control humano suficiente y la garantía de la responsabilidad de mando permanente (Ministère des Armées, France 2019). Esto significa que los Estados en sus normativas internas contemplan el principal aspecto que determina el DIH en cuanto al uso de la IA: el control humano.

En opinión del CICR, preservar el control humano sobre las tareas y el juicio humano en las decisiones que puedan tener graves consecuencias para la vida de las personas será también esencial para preservar una medida de humanidad en la guerra. El CICR ha

22 Ver nota 1 donde se mencionan los tratados que contemplan esas normas de protección.

23 En concreto, el EAAF ha utilizado algoritmos para la búsqueda e identificación de personas desaparecidas, creando redes complejas y métodos estadísticos que han contribuido a la investigación de la identificación de soldados argentinos enterrados en las Islas Malvinas durante el conflicto armado entre el Reino Unido, Irlanda del Norte y la República Argentina.

subrayado la necesidad de conservar la capacidad de decisión humana sobre el uso de la fuerza en los conflictos armados, una opinión que se deriva de consideraciones éticas más amplias sobre la humanidad, la responsabilidad moral, la dignidad humana y los dictados de la conciencia pública (CICR/ICRC, 2018: 15).

Preservar el control y el juicio humanos será un componente esencial para garantizar la observancia de las normas y mitigar las preocupaciones éticas que plantean ciertas aplicaciones de la IA y el aprendizaje automático.

Los responsables de utilizar determinados medios o métodos de combate que involucren técnicas o herramientas relacionadas con la IA deberán evaluar diversas situaciones, teniendo en consideración las normas del DIH, entre ellas:

- Proteger a las posibles víctimas del uso de la fuerza armada tomando todas las precauciones en el ataque, para asegurar el respeto a la condición civil.
- Atribuir al Estado la responsabilidad penal de los individuos, ya sean miembros de las fuerzas armadas o de grupos armados, que empleen armas que impliquen la aplicación de IA, como ya se ha mencionado en este artículo.
- Evaluar las funciones de todos los componentes de los sistemas de armas que involucren técnicas de IA, antes y después de su empleo mediante una supervisión continua.
- Considerar la viabilidad de umbrales de desactivación automática y/o revisión adicional con relación a los medios o métodos que impliquen técnicas de IA.
- Analizar la seguridad en el uso continuado de medios y métodos de combates vinculados a la IA.
- Contemplar la imposibilidad de probables improvisaciones en el empleo de medios o métodos relacionados con la IA.
- Conocer los componentes computacionales adoptados, es decir la configuración de los modelos y sus componentes respecto al empleo de armas, medios y métodos de guerra que impliquen técnicas o herramientas relacionadas con la IA;
- Evaluar los sesgos que puedan surgir en relación con los componentes informáticos adoptados para el empleo de medios o métodos de guerra que impliquen técnicas o herramientas relacionadas con la IA.

Es necesario que se mantenga un control humano en la aplicación militar de las tecnologías de IA, que se cumplan las normas del DIH ya existentes que establecen reglas legales y éticas sobre el empleo de la fuerza letal y que aquellas técnicas que violen sus disposiciones sean dejadas de lado o en última instancia prohibidas, tal como ya ha sucedido con otros medios y métodos de combate que no respetaron los esenciales principios humanitarios de distinción y proporcionalidad, por ejemplo las minas antipersonas o las armas nucleares²⁴.

CONCLUSIÓN

Sin lugar a duda el desarrollo de la tecnología ha aportado grandes avances al mundo, no debemos negarnos a ella, pero es necesario tener claro que se debe mantener el control humano y el juicio humano en aquellas decisiones que puedan tener graves consecuencias para las personas. Y una situación de conflicto armado es justamente donde es más necesario conservar la decisión humana sobre el uso de la fuerza. El avance de la tecnología es irreversible y el dilema que se presenta no se resuelve con retrotraerse, sino reflexionando cómo construir y cómo avanzar éticamente para obtener el mayor bien posible para el ser humano y evitar sufrimientos innecesarios.

Los sistemas de aprendizaje automático construyen sus propias reglas, siguen en ciertos casos aprendiendo y pueden cambiar el modelo después de desplegar su tarea específica. Pueden, por lo tanto, al incrementar su autonomía apartarse de las normas que regulan el uso de la violencia armada.

En realidad, hablar de un sistema autónomo no sólo remite a la toma de decisiones de manera independiente, esa autonomía implica también tener la facultad de determinarse y reconocerse en el cumplimiento de las normas jurídicas, y si estas se incumplen, asumir responsabilidades concretas: esto hace a la dignidad humana. Durante el desarrollo de un conflicto armado, no puede ponerse en manos de sistemas inteligentes decisiones que puedan agravar aún más la condición de víctimas de las personas afectadas.

Pero debemos reconocer que si se hace referencia a una ética de los sistemas inteligentes no podemos dejar de poner especial atención en el hecho de cómo orientar el uso humano de estos sistemas en forma ética.

24 Convención de 1997 sobre la prohibición de las minas antipersona y su destrucción y el Tratado sobre la prohibición de armas nucleares de 2017.

REFERENCIAS

- Badaró, Sebastián; Ibáñez, Javier y Agüero, Martín Jorge (2013). Sistemas Expertos: Fundamentos, Metodologías y Aplicaciones. *Ciencia y Tecnología*, 13: 349-364. Disponible en: https://www.palermo.edu/ingenieria/pdf2014/13/CyT_13_24.pdf
- Cáceres Tello, Jesús (2007). Reconocimiento de patrones y el aprendizaje no supervisado. Trabajo presentado en CISCI 2006. Disponible en línea: https://www.researchgate.net/profile/Jesus_Tello/publication/228857048_Reconocimiento_de_patrones_y_el_aprendizaje_no_supervisado/links/0c960517e7e677b522000000.pdf
- CICR/ICRC (1996). The Fundamental Principles of the Red Cross and Red Crescent. Disponible en línea en: https://www.icrc.org/en/doc/assets/files/other/icrc_002_0513.pdf
- CICR/ICRC (2006). *Guía para el examen jurídico de las armas los medios y los métodos de guerra nuevos. Medidas para aplicar el artículo 36 del Protocolo adicional I de 1977*. Disponible en línea: https://www.icrc.org/spa/assets/files/other/icrc_003_0902.pdf
- CICR/ICRC (2016). Views of the International Committee of the Red Cross (ICRC) on autonomous weapon system. Artículo presentado en Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS) 11-15 April 2016, Ginebra. Disponible en: <https://www.icrc.org/en/download/file/21606/ccw-autonomous-weapons-icrc-april-2016.pdf>
- CICR/ICRC (2018). *ICRC Strategy 2019-2022*. Disponible en línea en: <https://www.icrc.org/en/publication/4354-icrc-strategy-2019-2022>.
- CICR/ICRC (2019). Artificial intelligence and machine learning in armed conflict: A human-centre approach. Geneva: International Committee of the Red Cross. Disponible en línea en: https://www.icrc.org/en/download/file/96992/ai_and_machine_learning_in_armed_conflict-icrc.pdf
- Comisión Europea (2018). Inteligencia artificial para Europa. Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. COM/2018/237 final. Disponible en línea en: <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=COM:2018:237:FIN>
- Davison, Neil (2017). Autonomous weapon systems under international humanitarian law. En *Perspectives on Lethal Autonomous Weapon Systems*, United Nations Office for Disarmament Affairs (UNODA) Occasional Papers No. 30, pp. 5-18. UN Publications. Disponible en: <https://www.un.org/disarmament/wp-content/uploads/2017/11/op30.pdf>
- Department of Defense, USA (2012). Directive 3000.09. *Autonomy in Weapon Systems*. Disponible en: <https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf>
- Department of Defense, USA (2019). *Summary of the 2018 Department Of Defense Artificial Intelligence Strategy*. Disponible en: <https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF>
- EAAF (Equipo argentino de antropología forense). Nuevos algoritmos para la búsqueda e identificación de personas desaparecidas. Disponible en línea en: <https://eAAF.org/desarrollos-cientificos-e-informaticos/>
- European Commission AI HLEG (2019). *Ethics Guidelines for Trustworthy AI*, High-Level Expert Group on Artificial Intelligence, 8 April 2019. Disponible en español en: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60423
- Gutiérrez Posse, Hortensia (2003). Interacción entre tratado y costumbre. Aporte de los Protocolos Adicionales al desarrollo del Derecho Internacional Humanitario. *Revista Lecciones y Ensayos*, 78: 29-47. Disponible en: <http://www.derecho.uba.ar/publicaciones/lye/revistas/78/lecciones-y-ensayos-78-paginas-29-48.pdf>
- Henckaerts, Jean-Marie y Doswald-Beck, Louise (2007). *El derecho internacional consuetudinario*. (vol. 1: Normas). Buenos Aires: Comité Internacional de la Cruz Roja. Disponible en: https://www.icrc.org/es/doc/assets/files/other/icrc_003_pcustom.pdf
- Heyns, Christof (2013). *Informe del Relator Especial sobre las ejecuciones extrajudiciales, sumarias o arbitrarias, Christof Heyns*. Consejo de Derechos Humanos de Naciones Unidas. Asamblea General A/HRC/23/47. Disponible en: https://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_sp.pdf
- Human Rights Watch (HRW) y Harvard Law School's International Human Rights Clinic (IHRC) (2012). *Losing Humanity: The Case against Killer Robots*. Disponible en línea en: <https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots>
- Human Rights Watch. (2020). *World report 2020: Ante la amenaza de los robots asesinos, crecen las demandas para que los humanos continúen en control del uso de la fuerza*. Disponible en línea en: <https://www.hrw.org/es/world-report/2020/countrychapters/337416>
- Huston, Pat (2018). *Artificial Intelligence at the Frontiers of International Law concerning Armed Conflict* at Harvard Law School (workshop) <https://blogs.icrc.org/law-and-policy/2019/expert-views-frontiers-artificial-intelligence-conflict/>
- IFR (2017). The Impact of Robots on Productivity, Employment and Jobs. A positioning paper. International Federation of Robotics. April 2017. Disponible en: <https://ifr.org/papers/download>
- Ledesma Orozco, Sergio (2011). Redes neuronales: consideraciones prácticas. En Grigori Sidorov (ed.). *Métodos modernos de Inteligencia Artificial*, pp. 187-228. Ciudad de México: Sociedad mexicana de inteligencia artificial.
- León Serrano, Gonzalo (2019). Situación y perspectivas de las tecnologías y aplicaciones de inteligencia artificial. *La inteligencia artificial aplicada a la defensa*. Documentos de Seguridad y Defensa, 79: 39-68. Instituto Español de Estudios Estratégicos, Ministerio de Defensa. Disponible en: https://publicaciones.defensa.gob.es/media/downloadable/files/links/d/s/dsd_79_la_inteligencia_artificial_aplicada_a_la_defensa.pdf
- Maas, Matthius M. (2019). International Law Does Not Compute: Artificial Intelligence and the Development, Displacement or Destruction of the Global Legal Order.

- Melbourne Journal of International War*, 20: 1-29. Disponible en: https://law.unimelb.edu.au/__data/assets/pdf_file/0005/3144308/Maas.pdf
- Melzer, Nils (2019). *Derecho Internacional Humanitario. Una Introducción Integral*. Ginebra: Comité Internacional de la Cruz Roja.
- Ministère des Armées, France Florence Parly souhaite une Intelligence artificielle performante, robuste et maîtrisée, 10 avril 2019. Disponible en línea en: <https://www.defense.gouv.fr/fre/actualites/articles/florence-parly-souhaite-une-intelligence-artificielle-performante-robuste-et-maitrisee>
- Monasterio, Aníbal (2017). Ética algorítmica: Implicaciones éticas de una sociedad cada vez más gobernada por algoritmos. *Dilemata. Revista internacional de éticas aplicadas*, 24: 185-217. Disponible en: <https://www.dilemata.net/revista/index.php/dilemata/article/view/412000107>.
- Perel, Maayan y Elkin-Korel, Nivan (2016). Accountability in Algorithmic Copyright Enforcement. *Stanford Technology Law Review*, 19: 473-533.
- Pictet, Jean (1966). The Principles of International Humanitarian Law. *International Review of the Red Cross*, 66: 455-469. Disponible en: <https://international-review.icrc.org/sites/default/files/S0020860400011451a.pdf>
- Porcelli, Adriana (2021). La inteligencia artificial aplicada a la robótica en los conflictos armados. Debates sobre los sistemas de armas letales autónomas y la (in)suficiencia de los estándares del derecho internacional humanitario. *Revista Estudios Socio-Jurídicos*, 23 (1). Disponible en <https://doi.org/10.12804/revistas.urosario.edu.co/sociojuridicos/a.9269>
- Russell, Stuart y Norvig, Peter (2009). *Artificial Intelligence: A Modern Approach* (3rd Edition). New Jersey: Pearson Prentice Hall.
- Sassóli, Marcos (2014). Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified. *International Law Studies*, 90: 308-340. Disponible en: <https://digital-commons.usnwc.edu/ils/vol90/iss1/1>
- Schindler, Dietrich y Toman, Jiri (eds.) (1981). *The Laws of Armed Conflicts. A Collection of Conventions, Resolutions and Other Documents*, (2nd revised and complete ed.). Alphen aan den Rijn: Sijthoff and Noordhoff; Geneva: Henry Dunant Institute.
- Schmidhuber, Jürgen (2015). Deep Learning in Neural Networks: An Overview. *Neural Networks*, 61: 85-117.
- Schmitt, Michael N. y Thurnher, Jeffrey S. (2013). "Out of the loop": autonomous weapon systems and the law of armed conflict. *Harvard National Security Journal*, 4: 231-281. Disponible en <https://harvardnsj.org/wp-content/uploads/sites/13/2013/01/Vol-4-Schmitt-Thurnher.pdf>
- Simeone, Osvaldo (2018). A Very Brief Introduction to Machine Learning With Applications to Communication Systems. *IEEE Transactions on Cognitive Communications and Networking*, (4).
- SIPRI. Mapping the Development of Autonomy in Weapon System (2017). https://www.sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_systems_1117_1.pdf
- Swinarsky, Christophe (1990). *Principales nociones e institutos del Derecho Internacional Humanitario como sistema internacional de protección de la persona humana*. Cátedra Jean Pictet (Fundada por el Comité Internacional de la Cruz Roja), San José Costa Rica. Instituto Interamericano de Derechos Humanos.