SISTEMAS DE ARMAS AUTÓNOMOS LETALES Y RESPONSABILIDAD JURÍDICA INTERNACIONAL ALGUNAS REFLEXIONES SOBRE LA GUERRA EN UCRANIA

Antonio Pedro Marín Martínez
Doctor, Facultad de Derecho. Universidad de León

Resumen

Los Sistemas de Armas Autónomos Letales (SAAL), que utilizan tecnologías asociadas con la inteligencia artificial (IA) y la robótica, están cada vez más presentes en el campo operacional, como lo demuestra la actual guerra en Ucrania. Desgraciadamente, en muchas instancias, el progreso científico no siempre ha venido acompañado de una adecuación del derecho internacional al nuevo paradigma del ciberespacio, especialmente con relación al derecho internacional humanitario (DIH). Además, el rápido desarrollo de la investigación tecnológica expandirá las capacidades operativas de dichos sistemas sin la intervención humana, estresando aún más una adecuada implementación del derecho en el ámbito militar del ciberespacio y más concretamente con relación a la ciberguerra.

Palabras clave: Sistemas de Armas Autónomos Letales, Derecho internacional humanitario, Ciberguerra, Guerra de Ucrania.

Abstract

Lethal Autonomous Weapons Systems (LAWS) using technologies linked with Artificial Intelligence (AI) and robotics are increasingly pres-

ent in the operational environment, as we can see in the current war in Ukraine. Unfortunately, in many cases, scientific progress has not been accompanied by an adequate adaptation of International Law in relation to the new paradigm of cyberspace, especially in relation to International Humanitarian Law (IHL). Similarly, the rapid development of scientific research will increase the operational capabilities of those systems without human intervention, which further emphasizes the need for an adequate implementation of the Law in the military cyberspace domain and more specifically in relation with cyberwar.

Key words: Lethal Autonomous Weapons Systems, International Humanitarian Law, Cyberwar, Ukrainian war.

Sumario

1. INTRODUCCIÓN

En 2016 las Naciones Unidas (NU), a través de la «Convención sobre Prohibiciones o Restricciones del Empleo de Ciertas Armas Convencionales que puedan considerarse excesivamente nocivas o de efectos indiscriminados» (CCW), estableció un «Grupo General de Expertos» (GGE) sobre los SAAL reuniéndose a partir de entonces con regularidad. En el 2019, dicho grupo de expertos desarrolló una serie de principios rectores que, entre otros, establecía que el derecho internacional humanitario (DIH) era de completa aplicación a todos los sistemas de armas, incluidos aquellos con diversos grados de autonomía[1], premisa que se corroboró en

[1] Tomando como referencia el Informe del GGE sobre los SAAL de 2019, los Principios Rectores establecen que: «el derecho internacional, en particular la Carta de las Naciones Unidas y el derecho internacional humanitario, así como las perspectivas éticas pertinentes, debían guiar permanentemente la labor del Grupo» y que «El derecho internacional humanitario sigue aplicándose plenamente a todos los sistemas de armas, incluido el posible desarrollo y uso de Sistemas de Armas Autónomos Letales». Dichos principios fueron avalados por el CCW en noviembre de 2019 (NU 2019a).

2022[2]. Paralelamente, en el ámbito académico, gubernamental y militar, se han venido desarrollando estudios y propuestas teóricas y prácticas para incluir ciertos elementos éticos, así como del derecho consuetudinario, en el desarrollo de dichos sistemas, especialmente a través de la utilización de algoritmos computacionales que se sirven de la IA y la robótica. Elementos que estamentos internacionales, como las Naciones Unidas (NU) y los Estados, han identificado como básicos para los SAAL y su impacto sobre el DIH: distinción, proporcionalidad, responsabilidad, humanidad y necesidad militar, especialmente con relación a la «adquisición de objetivos» (*targeting*) e «intervención sobre objetivos» (*engagement*). No obstante, el desafío práctico pendiente es pasar de los planteamientos teóricos y filosóficos actuales a otros prácticos en el marco temporal actual, explorando los pasos necesarios para poder aplicar el derecho internacional a los SAAL sin una pérdida normativa.

2. LOS SISTEMAS DE ARMAS AUTÓNOMOS LETALES (SAAL)

El principal concepto que se debe retener al definir dichos sistemas de armas es el elemento de «autonomía». No obstante, en la actualidad no existe un consenso internacional sobre la definición de dicho término. El principal escollo proviene de que el propio concepto es relativo y, por lo tanto, los Estados hacen una distinción entre automatización, autonomía e independencia. La «automatización» se entiende como que existe el conocimiento de unas respuestas preprogramadas y predecibles de una tarea en cualquier situación. En el caso de los sistemas de armas y aunque sus algoritmos sean robustos siempre existirá un grado de comportamiento probabilístico de aleatoriedad, aunque los subsistemas individuales sean determinísticos. Asumir que la autonomía siempre producirá sistemas estables y seguros no sería válido para sistemas complejos.

En cuanto a la «autonomía» se debería entender como la capacidad para desarrollar una acción de una forma autosuficiente y autogobernable. Se incluye la libertad de una planificación *propia* en las tareas y subtareas, concepto ampliamente conocido en el ámbito informático. Dichas tareas tendrían diferentes grados de complejidad y de interacción con los humanos o con otros sistemas. Por lo tanto, el concepto de «autonomía» no

[2] Tomando como referencia el informe del GGE sobre los SAAL de 2022, cualquier acto ilegal de un Estado cometido en el uso de los SAAL implicaría la responsabilidad internacional de dicho Estado de acuerdo con el derecho internacional y que incluye el DIH (NU 2022a).

podría ser una característica simple de «encendido/apagado» y, en consecuencia, en vez de «sistemas autónomos» el concepto se debería expresar como «sistemas que tienen funciones o características autónomas», siendo difícil definir el grado de autonomía ya que cada sistema sería diferente. En cuanto a los SAAL, el enfoque se debería poner, por tanto, en el ciclo de selección de objetivos y las condiciones de autorización del uso de la fuerza letal, especialmente el retardo entre la orden y la ejecución que comprendería la dinámica de la tarea y la ventana de tiempo existente para la autorización.

En contraste con una operación autónoma, la «independencia» verdadera significaría que el sistema sería capaz de definir y decidir los objetivos finales de su funcionamiento, de la misma forma que realizan los humanos. Por lo tanto, la selección de objetivos estaría subordinada a la propia motivación del sistema, para lo que se requeriría una IA que hubiese evolucionado más allá del punto de la Singularidad[3], momento en el cual la inteligencia de las máquinas sobrepasaría la mente humana, capacidad que vendría dada por mejoras recursivas de la tecnología (NU 2018a: 2-3; 2019b: 4).

Independientemente de la conceptualización de la «autonomía» cualquier sistema de armas, incluidos los autónomos letales (SAAL), tendrán como objetivo principal la destrucción del adversario, concepto que subyace en la concepción de la guerra como advertía el teórico C. V. Clausewitz. Una premisa que se concretaría en la actualidad con el desarrollo de artefactos integrados con algoritmos de IA, con la capacidad de una notable autonomía en la toma de decisiones de dichos sistemas. En un entorno de creciente volatilidad, incertidumbre, complejidad y ambigüedad de inestabilidad geopolítica, lo que entraría en juego sería el impacto que tendría la utilización de los SAAL sobre la decisión más crítica de su función: la liberación de la fuerza, producto de una combinación entre el mundo físico y el mundo digital, lo que se denominaría como una «Realidad Mixta»[4], conduciendo a un nuevo nivel de despersonalización de la guerra, según argumentan los investigadores C. Hayns y A. Westhues (Heyns 2016: 4; Westhues 2020: 12).

---

[3] El término «Singularidad» proviene de J. V. Neumann, según indica S. Ulam en el obituario que escribió a su muerte, donde recuerda una conversación entre ambos que se: «centró en el progreso acelerado de la tecnología y los cambios en los modos de vida humana, que da la sensación de estar acercándose a alguna forma de "singularidad" esencial en la historia de la raza humana, a partir de la cual los asuntos humanos, como los conocemos, no continuarían» (Ulam ,1958: 5).

[4] J. Young *et al.* definen la «Realidad Mixta» como: la combinación del mundo real y el virtual para producir nuevos entornos y visualizaciones, donde los objetos físicos y los digitales coexisten e interaccionan en tiempo real (Young *et al.,* 2011: 2).

En dicho contexto, como argumentan los investigadores I. Bode y H. Huelss, lo decisivo sería qué tipo de control se ejerciese sobre el uso de la fuerza y no tanto dónde se llevase a cabo. Un tipo de control que sería una combinación de normas, como el derecho internacional de los conflictos armados (DICA) con otras surgidas de la práctica (desarrollo, entrenamiento, prueba y despliegue) del uso de los SAAL, englobados en una serie de principios fundamentales: la necesidad militar, la humanidad, la distinción, la proporcionalidad y el principio de precaución. Elementos que no reemplazan al propio derecho pero que lo sustentan (Bode y Huelss 2018: 395-397; Young *et al.* 2011: 2, 10).

2.1. El armamento indiscriminado y sus desafíos

Es importante destacar que el DICA establece las normas que reflejan el equilibrio entre la necesidad militar y la humanidad. Un equilibrio que según el Comité Internacional de la Cruz Roja (CICR) se refleja en el principio de distinción, que requiere que las fuerzas armadas distingan claramente entre objetivos militares y personas o bienes civiles[5]. No obstante, en un entorno de «Realidad Mixta», no queda claro como dicho principio funcionaría en el mundo cibernético, incluidos los SAAL, pues las nuevas tecnologías crean una nebulosa entre lo militar y lo civil. Los investigadores R. Geib y H. Lahmann argumentan que, en el mundo cibernético, cualquier artefacto podría ser un objeto de uso dual, pudiendo ser utilizado como objetivo militar legítimo con amplias repercusiones hacia la población civil. Así, cualquier infraestructura cibernética (computadoras, redes y cables) o incluso el propio ciberespacio podrían ser considerados como objetivos militares (Geib y Lahmann 2012: 382-383).

En todo caso, el principio de distinción estaría intrínsecamente ligado a la regla sobre «ataques indiscriminados», de acuerdo con lo establecido en el art. 51.4 del PAI sobre protección de la población civil que, en la práctica de los Estados, se ha convertido en una norma de derecho internacional consuetudinario aplicable a los conflictos armados internacionales.

[5] Artículos 48, 51(2) y 52(2) del Protocolo Adicional I (PAI) de los Convenios de Ginebra de 1949, así como el Protocolo II de la Convención sobre Prohibiciones o Restricciones del Empleo de Ciertas Armas Convencionales que puedan considerarse excesivamente nocivas o de efectos indiscriminados (CCW) de 1980, art. 3(2) (*ibid*., párr. 157); Protocolo II enmendado de la CCW de 1996, art. 3(7) (*ibid*., párr. 157); Protocolo III de la CCW de 1980, art. 2(1) (*ibid*., párr. 158); preámbulo de la Convención de Ottawa de 1997 (*ibid*., párr. 3) (CICR 1977; CICR 2007).

Además, de acuerdo con el art. 85(3)(b) del PAI, un ataque indiscriminado se consideraría como una violación grave del Protocolo (CICR 2005: 248; Geib y Lahmann 2012: 382-383).

En cuanto a los SAAL, el desafío es establecer que se entiende por principio de «distinción» y la regla sobre «armamento indiscriminado», para lo que sería necesario establecer, en primer lugar, que se considera como sistema «autónomo» o «semiautónomo» en un sistema de armas. Para el CICR, un SAAL sería aquel sistema de armas que tuviese autonomía en sus funciones críticas de: selección de objetivos (búsqueda o detección, identificación, seguimiento, selección) y de ataque (uso de la fuerza, neutralización, daño o destrucción) a un objetivo sin la intervención humana. Para el investigador H. Roff, sería un sistema de armas que estaría formado por cuatro funciones básicas: accionamiento del sistema (*triggering*), selección de objetivos (*targeting*), navegación (*navigation*) y movilidad (*mobility*), y se consideraría semiautónomo si tuviese entre uno y tres de dichos elementos automatizados informáticamente. Por lo tanto, un SAAL podría ser tanto un sistema de defensa antimisiles, un dron «suicida», un programa malicioso (*malware*), pero también un ciberataque sofisticado durmiente durante un periodo de tiempo determinado que se activaría ante un evento concreto (CICR 2016: 8, 57, 60; Roff, 2015b).

Específicamente en el ámbito de la tecnología militar, la «autonomía» sería la capacidad de un sistema de armas de determinar por sí mismo y sin ningún tipo de interferencia humana cuando y contra quién se utilizaría una fuerza letal. En dicho contexto, el sistema debería poseer los siguientes elementos: la capacidad de analizar todos los posibles resultados y sugerir la mejor estrategia posible; hacer que los robots inteligentes se coordinasen entre sí para una acción común; y tener la capacidad analítica de mostrar un discernimiento moral igual que el de los seres humanos. El investigador J. F. Carson puntualiza que, en la actualidad, la IA solo sería capaz de cumplir las dos primeras premisas, como en el caso de los sistemas antimisiles o los programas maliciosos. Una postura que vendría ligada a la de los investigadores G. Sartor y A. Ominici, que postulan la necesidad de distinguir entre la «capacidad de independencia» que tiene un sistema de armas para completar una tarea y la «independencia organizativa». Es decir, su capacidad para llevar a cabo una tarea dentro de la «infraestructura socio técnica global», sugiriendo que los efectos de un sistema de armas no solo dependerían de su diseño, sino también del uso que se le diese y la vulnerabilidad de aquellos a los que afectase (Carson 2020: 174-175; Sartor y Omicini 2016: 40).

La incapacidad de los algoritmos actuales de posibilitar el discernimiento moral como el de los seres humanos ha llevado a argumentar a los investigadores E. Rosert y F. Sauer, así como a la organización no gubernamental (ONG) Human Rights Watch, que los SAAL serían incapaces de discernir entre combatientes y civiles, violando, por tanto, el principio de distinción por lo que se deberían considerar como sistemas de armas indiscriminados y por tanto atentarían contra la dignidad humana. Esto sería debido a que el concepto de «civil» es un término complejo, dependiente del contexto en el que se establece la acción, no siendo trasladable a una aplicación informática (*software*), independientemente de que dicho algoritmo estuviese basado en reglas o a través del aprendizaje automático (*machine learning*). Es más, para dichos investigadores, el reconocer y aplicar el concepto de «civil» en el campo de batalla requeriría de juicios de valor, pero también de un cierto grado de conocimiento del entorno y una comprensión del contexto social, que la tecnología computacional actual sería incapaz de poseer (HRW 2012: 30-31; Rosert y Sauer 2019: 372-373).

Existen, no obstante, críticas con relación a la utilización de la «dignidad humana» como base para prohibir los SAAL. El investigador A. Saxton argumenta que el uso *de facto* de dichos sistemas no necesariamente significaría una violación inherente de la dignidad humana simplemente por su autonomía. El concepto de dignidad humana sería difícil de medir durante un conflicto, dado que la propia guerra sacrifica parte de la dignidad humana a través del sacrificio intencionado de vidas para alcanzar objetivos militares. Es más, según el investigador M. N. Schmitt, una prohibición total cercenaría la posibilidad de utilizar los SAAL para minimizar el daño «civil» en comparación con otros arsenales existentes. Postura que sería apoyada por el investigador R. Artkin, que argumenta que es muy probable que en un futuro se desarrollen SAAL que minimicen las bajas civiles con más capacidad que los propios humanos. No obstante, en nuestra opinión, para que ello fuese una realidad dichos sistemas necesitarían de un desarrollo y despliegue implementando el principio de precaución y que pudiese cumplir con los preceptos del DIH. En el caso que dicho sistema fuese más allá de la capacidad humana en su aplicación de dicha normativa humanitaria, entonces significaría que los SAAL estarían cumpliendo un rol humanitario y existiría un imperativo moral para su utilización (Artkin 2013: 1-3, 5; Saxton 2016; Schmitt 2012: 290-292).

Ahora bien, siguiendo la argumentación de los investigadores J. Foy y A. K. Krishnan, que nosotros compartimos, en la actualidad aún existen importantes desafíos para que un SAAL pueda observar el principio de distinción y como corolario que no se considerase un sistema de armas

indiscriminado. Dichos desafíos se podrían resumir en las siguientes inquietudes (Foy 2014: 57-59; Krishnan 2009: 98-99):

- Una capacidad de percepción de la máquina aún débil (*weak machine_perception*): La distinción requeriría de una evaluación a través de una información proveniente de sensores que necesitarían de una alta capacidad de discriminación. Una situación que se volvería más compleja cuando existiesen combatientes no uniformados;
- Un problema del marco de actuación (*frame problem*): En un entorno de batalla moderno de ritmo acelerado, un SAAL tendría problemas para interpretar toda la información existente. Esto implicaría la necesidad de programar qué información sería o no relevante. Una programación inadecuada podría llevar a una interpretación errónea de la información causando un ataque indiscriminado, que se vería incrementado si existiesen dudas sobre la legitimidad del objetivo. La dificultad estribaría en establecer en dichas situaciones el umbral de ataque dentro del algoritmo.
- Un algoritmo débil (*weak software*): El incremento de la complejidad de un algoritmo le hace menos predecible. Ningún programador comprende un algoritmo complejo al completo, dado que muchas veces se programan por módulos, por lo que, combinado con un entorno abierto, se podrían dar situaciones donde el SAAL aplicase la fuerza indiscriminadamente por un error de programación no anticipado o una situación no programada.

A dichas inquietudes el investigador E. K. Gade también añade la neblina que existe en la actualidad para distinguir un combatiente de un no combatiente, de acuerdo con los términos del DIH, que no equivaldría a ser un «civil» ya que dicho sujeto también podría participar en la «máquina de guerra». Por lo tanto, la decisión de establecer quién sería considerado un «no combatiente», sería subjetiva y dependería del contexto y el entorno en el que se desarrollase la acción, lo que dificultaría el desarrollo de algoritmos eficientes para los SAAL. Dichos desafíos sugieren la necesidad de tener una precaución extrema antes de desplegar y usar SAAL completamente autónomos, si se desea que cumplan el principio de distinción, por lo que sería importante investigar algoritmos de IA que pudiesen establecer elementos de control adecuados (Gade 2010: 227).

Si se tomase como ejemplo la actual guerra de Ucrania y la utilización de SAAL autónomos o «semi» autónomos, siguiendo las premisas de J. Foy y A. K. Krishan anteriormente expuestas, se podría intentar establecer si algunos de los sistemas de armas utilizados por los diversos contendien-

tes serían capaces de aplicar el principio de «distinción» o por el contrario se podrían considerar como sistemas de armas indiscriminados.

Un primer ejemplo podría ser la utilización por parte de la Federación Rusa de misiles tierra-aire S-300, originalmente diseñados para la defensa antiaérea, en el bombardeo en la zona de Zaporiyia, el 30 de septiembre de 2022, que dejó treinta civiles muertos. La reasignación de dichos sistemas antimisiles con IA para el bombardeo terrestre, tomando como base las inquietudes antes establecidas, implicaría un uso que acrecentaría la pérdida de precisión, dado que su algoritmo sería débil para dicho uso, el marco de actuación de dicho sistema de armas sería inadecuado y su sistema de percepción para dicho cometido también sería débil. Esto tendría como consecuencia que, en dicho escenario, dicho sistema podría ser considerado como un sistema de armas indiscriminado (Segura, 2022).

Alternativamente, en un principio no se podría establecer, siguiendo las mismas premisas, que la utilización por parte de la Federación Rusa de drones «suicidas» iraníes, de tipología Shahed 136 o el Geran-2 ruso (mismo sistema de armas), pudiesen ser considerados como indiscriminados, ya que son considerados efectivos y de gran precisión en su utilización contra objetivos fijos[6]. Por lo tanto, en teoría se podrían considerar como un sistema que seguiría teóricamente el principio de distinción. Tampoco el dron semiautónomo Bayraktar TB2 (turco) utilizado por Ucrania podría ser considerado como un sistema de armas indiscriminado, ya que tiene *software* que le permite en todo momento conocer sus coordenadas precisas y calcular la ruta óptima para alcanzar su objetivo (Noticias de Israel, 2022; Segura, 2022; Svitlyk, 2022).

Ahora bien, el principio de «distinción» también establece que las operaciones militares solo deben ser dirigidas contra objetivos militares y los civiles no pueden ser atacados si no participan directamente en las hostilidades, así como la prohibición de atacar o destruir los bienes indispensables para la supervivencia de la población civil, como las instalaciones y reservas de agua[7]. En el caso de los drones iraníes utilizados en Ucrania por parte de la Federación Rusa, dichos SAAL estarían siendo utilizados para destruir, entre otras, infraestructuras críticas civiles de electricidad y agua. Por lo tanto, en dicho escenario y aunque dichos sistemas de armas fuesen precisos, no seguirían el principio de «distinción», como ha puesto

[6] Los Shahed 136 utilizan un guiado de navegación por satélite (GPS o Glonass ruso) y otro sistema de navegación inercial de reserva en caso de interferencia sobre la señal de satélite, hasta que pueden volver a captar dicha señal (Noticias de Israel, 2022).

[7] Artículo 54.2 del Protocolo Adicional I a los Convenios de Ginebra de 1949, relativo a la protección de las víctimas de los conflictos armados internacionales, 1977 (CICR, 1977).

de manifiesto el CICR en su información sobre el DIH en el conflicto de Ucrania. Un uso de los drones iraníes que, además, podría estar incumpliendo los artículos 51.4 y 54.2 del PAI ya mencionados (CICR, 2022; Defense Post, 2022).

2.2. Proporcionalidad en los SAAL

Además de que un SAAL desarrolle adecuadamente la capacidad de establecer el principio de distinción, el principio de proporcionalidad requiere que el uso de cualquier sistema de armas sea evaluado para determinar entre la ventaja militar por su uso y el daño contra el estamento civil (personas y/u objetos), ya que el daño hacia dicho colectivo no debería ser excesivo con respecto a la ventaja militar obtenida. Una idea que vendría enmarcada, en el ámbito del *ius ad bellum,* dentro del marco de «guerra justa» y que implicaría que la destrucción perpetrada por una guerra no debería ser desproporcionada con relación al bien que dicha guerra debiese alcanzar. Incluso si existiese una causa justa, el recurrir a una guerra podría ser impropio si el daño que causase fuese excesivo. En cuanto al *ius in bello*, el daño colateral a los civiles estaría prohibido si resultase desproporcionado, lo que se consideraría un uso excesivo de la fuerza, con relación a los resultados obtenidos quedando reflejado en el artículo 51(5)(b) del PAI. También, el Estatuto de Roma de la Corte Penal Internacional (Artículo 8(2)(b)(i-vi)) lo consideraría, además, como «crímenes de guerra» (Anderson y Waxman, 2013: 41; Hurka, 2005: 34-36; Roff, 2015a: 39-40).

Siguiendo con el argumento de que los SAAL pudiesen salvar vidas, especialmente en el caso de una amenaza injusta y mencionando en el apartado anterior por el investigador M. N. Schmitt, dicha idea serviría de especial referencia con relación al análisis sobre el principio de proporcionalidad. En dicho contexto, en el caso de que un SAAL incurriese en muertes colaterales persiguiendo objetivos militares legítimos, dichas muertes serían desafortunadas, pero no prohibidas, siguiendo la doctrina del «efecto doble» (*double effect*)[8]. Se estaría entonces ante un escenario de «defensa propia» que solo sería aceptable si el daño fuese inminente y dirigido contra los intereses vitales del Estado, diferenciando entre ataque «anticipatorio» y ataque «preventivo», este último, prohibido. En dicho

[8] De acuerdo con dicho principio, a veces está permitido causar daño colateral (o «efecto doble») para obtener un buen resultado, aunque no estuviese permitido causar dicho daño como forma de alcanzar el mismo buen fin (McIntyre, 2018).

contexto habría que insistir, como argumenta el investigador H. M. Roff, en que se deberían ponderar todos los daños que dicha acción pudiese causar en un futuro y no solo los relativos al principio de «guerra justa» en el presente, dado que un SAAL incide tanto en el *ius ad bellum,* el *ius in bello* y el *ius post bellum.* Dicha premisa afectaría, en consecuencia, a los cálculos sobre el principio de proporcionalidad cuando se decidiese entrar en guerra (Roff, 2015a: 42-44, 47, 49-50; Schmitt, 2013: 176).

Paralelamente, habría que destacar que el principio de proporcionalidad no se puede definir en abstracto, como establece M. Wagner, pues solo tendría sentido desde un punto de vista contextualizado, es decir: una acción concreta, en un escenario concreto y en un tiempo concreto. En dicho marco sería importante establecer el significado del término «excesivo» del artículo 51(5)(b) del PAI, dado el potencial de un entorno cambiante dentro de la propia acción de combate. El artículo 57(2) del PAI sería la base para establecer dichos límites al requerir a los comandantes que tomasen precauciones para evitar o minimizar el daño o la pérdida de vidas indirectas. Destacaríamos, no obstante, la puntualización que diversos Estados han realizado con respecto a la aplicación del Artículo 51 del PAI, ya que consideran que la aplicación del principio de proporcionalidad dependería de la información que se tuviese a mano en cada momento[9]. En todo caso, M. Wagner argumenta que el principio de proporcionalidad sería demasiado impreciso, lo que crearía tensiones en su aplicación, como pondría de manifiesto el informe final del Tribunal Penal Internacional para la ex-Yugoslavia, sobre los bombardeos de la Organización del Tratado del Atlántico Norte (OTAN), que argumenta: «El problema principal con el principio de proporcionalidad no es si existe, sino qué significa y cómo se debe aplicar» (ICTY, 2000; Wagner, 2014: 1393-1397).

En el caso de Ucrania, como indica el profesor R. M. Dover, la Federación Rusa estaría reescribiendo las reglas de la guerra de asedio, destruyendo las capacidades energéticas para, por ejemplo, poder cocinar, calentarse o beber agua por parte de la población civil. La utilización de drones para llevar a cabo la destrucción de infraestructuras civiles críticas tendría, a su entender, una legalidad internacional cuestionable y sería una forma clara de incumplir el derecho internacional. No obstante, el poder establecer que Rusia estaría incumpliendo el principio de proporcionalidad en la guerra de Ucrania seguiría siendo complicado. Existiría una necesidad de contextualización importante. No sería lo mismo el utilizar drones para

[9] Postura que también manifiesta el Estado español en la ratificación de los Protocolos I y II a los Convenios de Ginebra de 1949 (*BOE,* 1989: 23828).

destruir infraestructura de «doble uso» (civil/militar), como por ejemplo los depósitos de combustible, como la destrucción de las infraestructuras de bombeo de agua para la población en centros urbanos, como en el caso de los ataques a dichas infraestructuras en Kiev y otras ciudades ucranianas (Dover, 2022).

Ahora bien, ¿podría en la actualidad un algoritmo de IA aplicar adecuadamente el principio de proporcionalidad a través de un análisis tan dependiente del contexto? El propio M. Wagner plantea dudas sobre dicha posibilidad, a no ser que dicho algoritmo fuese capaz de manejar un gran número de decisiones y no solo codificar escenarios individuales, codificando así un gran número de reglas que pudiesen tomar decisiones mientras sopesasen un sinfín de factores. Posiblemente, dados los avances de la IA y el uso de vectores efectivos, en un futuro a medio plazo sea posible avanzar en dicho sentido, posiblemente a través de la computación cuántica, para lograr una contrapartida aceptable entre la necesidad militar y la protección civil, aunque a corto plazo no parezca probable (Wagner, 2014: 1398-1399).

2.3. Rendición de cuentas y responsabilidad en los SAAL

Los principios de «rendición de cuentas» (*accountability*) y «responsabilidad» (*responsability*) son conceptos clave cuando se analizan los SAAL. El GGE del CCW sobre los SAAL de las NU, al establecer los posibles principios rectores especificaba que: «El ser humano debe mantener la responsabilidad por las decisiones que se adopten sobre el uso de los sistemas de armas, ya que la obligación de rendir cuentas no puede transferirse a las máquinas y dicha consideración debería tenerse en cuenta durante todo el ciclo de vida del sistema de armas». No obstante, según argumentan, entre otros, los investigadores I. Verdiesen *et al.*, P. M. Asaro o M. Wagner, los SAAL podrían crear un vacío en relación con la rendición de cuentas, circunstancias en donde ningún humano fuese culpable de las decisiones, acciones o efectos de dichos sistemas de armas. Idea que partiría de la premisa de que la comunidad internacional sería incapaz de verificar la legalidad de la acción, ni confirmar la autenticidad de la inteligencia utilizada en el proceso de «adquisición» de los objetivos, lo que supondría como resultado la impunidad de dicha acción (Asaro, 2012: 693; un, 2018b: 4; Verdiesen *et al.,* 2021: 138, 145; Wagner, 2014: 1371).

En todo caso, la rendición de cuentas estaría íntimamente ligada al concepto de control. Una idea argumentada por el investigador M. Bo-

vens, donde la rendición de cuentas sería una forma de control, aunque no todas las formas de control serían mecanismos de rendición de cuentas. Si se tomase como base la perspectiva sociotécnica del concepto de control, se estaría ante la premisa de que un agente controlaría a otro agente (que podría ser humano o un artefacto) a través de normas legales, sanciones o instrucciones políticas, y estaría intrínsecamente ligado al concepto de rendición de cuentas a través de un control que podría ser *ex ante,* durante o *ex post*. Dado que, en la actualidad, según argumentan I. Verdiesen *et al.,* los desarrolladores de algoritmos no dispondrían de un marco estructural de control adecuado*,* sería necesario el establecimiento de un entorno de gobernanza robusto, incluyendo el desarrollo de nuevas herramientas de control, como el diseño y desarrollo de los denominados Agentes Morales Artificiales (AMA) que se introducirían como «*software* moral» en los SAAL. Desafío que habría llevado a desarrollar la noción del «Control Humano Significativo» (CHS) (*Meaningful Human Control [MHC]*) en el debate sobre los SAAL (Bovens, 2007: 454; Verdiesen *et al,*. 2021: 147-148).

Sería la ONG Article 36 la primera en acuñar dicho término en su aportación al debate sobre los SAAL, en el foro del CCW de 2014. Posteriormente, en el 2016, la propia ONG presentó al debate una serie de elementos que consideraba necesarios para que existiese un CHS amplio de los SAAL: que la tecnología fuese predecible y transparente; que el usuario tuviese una información precisa; que existiese la posibilidad y la capacidad de una intervención humana oportuna y; que también hubiese alguna forma de rendición de cuentas. Además, el CHS debería estar integrado a través de algoritmos durante todo el ciclo del uso del SAAL en un conflicto: *ante bellum, in bello* y *post bellum*, así como en todas las fases de combate: táctica, operacional y estratégica. La necesidad de un CHS apropiado, también fue argumentada por otras ONG como Amnistía Internacional o Human Rights Watch (Article 36, 2014 y 2016; Chengeta, 2017: 855-856; HRW 2016).

En dicho contexto, el problema de fondo surge por la inexistencia de una definición consensuada del significado del término CHS en los foros internacionales, como en el caso del CCW. Dicho estancamiento ha propiciado que solo exista una definición abstracta de dicho término, lo que ha llevado al investigador M. Ekelhof a proponer la necesidad de comprender lo que dicho control significaría en el contexto de las operaciones militares. No solo sería necesario un control a nivel operativo, sino que sería crucial que dicho control se ejerciese en todas las fases, incluido el desarrollo, la formación y el despliegue de los SAAL, por lo que el CHS solo tendría

significado si se ejerciese dentro del marco general del control distribuido que enmarca a cualquier sistema de armas[10] (Ekelhof, 2019: 347).

Quedaría pendiente, no obstante, como se establecerían las diversas responsabilidades de los distintos actores (individual, mandos, Estado) que, como argumenta T. Chengeta, serían responsabilidades complementarias y no se deberían excluir unas de otras, aunque desde un punto de vista jurídico cada actor sería responsable por sus propias acciones durante las diferentes fases. Por lo tanto, para que un CHS se convirtiese en un estándar legal, se debería especificar para qué actores sería de aplicación dicho término. A nuestro entender, aunque el CHS fuese específico para cada fase, su concepto debería abarcar todo el «Marco Integrado de Supervisión Humana» de una forma holística y no únicamente, como argumentan T. Chengeta o la ONG Article 36, a nivel individual del operador (combatiente) como usuario final del SAAL. El CHS debería ejercerse en todas las fases del proceso de creación, despliegue, uso y revisión de cualquier sistema de armas con IA. Dicha premisa aseguraría el cumplimiento del derecho internacional vigente y cualquier otra norma estándar o buenas prácticas desarrolladas, tanto a nivel internacional, los Estados o la propia sociedad (Article 36, 2016; Chengeta, 2017: 868-869).

Tomando como referencia el argumento de T. Chengata sobre las responsabilidades complementarias y no excluyentes, si se tomase como ejemplo el sistema de armas Shahed 136 iraní, se podría establecer que, desde un punto de vista jurídico, cada actor sería responsable por sus posibles acciones durante sus distintas fases. Así, se podría argumentar que dicho sistema de armas no exhibiría un CHS adecuado ya desde su fase de desarrollo, pues no utilizarían un controlador humano dedicado una vez lanzados, lo que implicaría que no existiría la posibilidad y la capacidad de una intervención humana oportuna en caso de desviación del objetivo. En tal circunstancia el lanzador de dicho dron podría argumentar que la muerte de civiles por una explosión, por ejemplo, en un barrio residencial, podría ser un accidente trágico y no sería posible verificar cual era el verdadero objetivo del dron al no contar con un CHS adecuado en su diseño y construcción. Alternativamente, también se podría argumentar que el lanzamiento de un número significativo de drones hacia un objetivo particular, como la maquinaria de bombeo del agua de una población, no podría ser considerado un ataque accidental, dado que es bastante impro-

[10] En dicho contexto, I. Verdiesen *et al.* propusieron un «Marco Integrado de Supervisión Humana» (*Comprehensive Human Oversight Framework*) que pretende analizar las conexiones entre las diversas fases e identificar las posibles lagunas en los mecanismos de control (Verdiesen *et al.,* 2021: 151).

bable que un gran número de drones sufriesen la misma avería. En tal caso, el lanzador (la Federación Rusa) podría ser encausado de acuerdo con el DIH por el ataque a objetivos civiles, esgrimiendo el art. 54.2 del PAI. Así, dependiendo del escenario y la contextualización del momento se podría argumentar uno u otro caso (CICR, 1977; Kossov, 2022).

2.4. El artículo 36 del Protocolo Adicional I y los SAAL

En el marco técnico de los SAAL se establece el desarrollo de la fusión de sensores que permiten la imitación de las funciones del cerebro humano, combinando su información individual al mismo tiempo que interactúan con el entorno para, por ejemplo, calcular trayectorias y movimientos. Nuevas capacidades que establecen nuevos retos para la observancia del DIH por los nuevos sistemas de armas. Dos problemas principales podrían surgir: que el SAAL fuese incapaz de adherirse por sí mismo a lo establecido en el DIH, aunque el objetivo fuese legal; o, se evaluasen las funciones de dicho SAAL en cuanto a sus medios (*means*), de acuerdo con lo establecido en el Artículo 36 del PAI, considerándose no aceptables por incumplir dicho precepto. El problema, surgiría en la forma de analizar los métodos (*methods*) (despliegue y tácticas) que fuesen implementados para un SAAL determinado con relación al cumplimiento del DIH, teniendo en cuenta que dichos métodos tendrían un impacto real en la capacidad militar del armamento. Tanto para el CICR como para el investigador V. Sehrawat, entre otros, lo crucial sería el incorporar el estudio del «método de combate» (*method of warfare*) establecido como una parte intrínseca en el análisis (CICR, 2006: 4; Sehrawat, 2017: 41-43).

Desgraciadamente, en la actualidad no existe ninguna estandarización a nivel internacional para llevar a cabo el análisis de dichos métodos, siendo, un importante condicionante para la aplicación del Artículo 36 del PAI, pues cada Estado puede desarrollar sus propios procedimientos internos. Tampoco existe un consenso internacional sobre qué armamento sería susceptible de ser revisado de acuerdo con dicho artículo, aunque en general se establezca que estaría reservado para aquellos medios de combate destinados a causar daño a personas u objetos. Por ejemplo, el *Manual de Tallinn 2.0* define un armamento cibernético como: los medios de combate utilizados, diseñados o destinados a ser usados para causar daño o muerte a personas y/o daños a objetos. Para los investigadores G. D. Brown y A. O. Metcalf sería: «un artefacto desarrollado u obtenido para un uso primario de matar, mutilar, herir, dañar o destruir». Si se aceptase dicha premisa, se

excluirían aquellas herramientas destinadas únicamente a obtener información (Brown y Metcalf, 2014: 135; OTAN, 2017).

En cuanto a la revisión de un SAAL para que fuese compatible con el Artículo 36 del PAI, dejando a un lado la dimensión del nivel de control por parte de un humano sobre un SAAL analizado anteriormente, habría que tener en cuenta el nivel de sofisticación del algoritmo que determinaría el control sobre su comportamiento y como afrontaría las incertidumbres que surgiesen debido al entorno de uso. En dicho contexto, los SAAL podrían ser catalogados en tres categorías (Boulanin y Verbrugge, 2017: 18):

- Sistemas Reactivos: el sistema seguiría una serie de reglas usando la metodología condición-acción, que explícitamente prescribiría como un sistema reaccionaría a una información recibida de un sensor. Su comportamiento sería predecible si se conociesen las reglas que utiliza.
- Sistemas Deliberativos: el sistema utiliza un modelo del entorno, una función de valor que provee información sobre el objetivo deseado y un conjunto de reglas potenciales que le ayuda a buscar y planificar como llevar a cabo dicho objetivo. Para decidir una acción el sistema compararía las posibles consecuencias de las acciones factibles para encontrar la más apropiada. El comportamiento no sería completamente predecible, dado que, aunque el marco de actividad lo fuese, las acciones individuales puede que no.
- Sistemas de Aprendizaje: dichos sistemas podrían mejorar su rendimiento a lo largo del tiempo a través de la experiencia. Aprenderían a través de la abstracción de las relaciones estadísticas de los datos. El conocimiento aprendido serviría para volver a parametrizarse automáticamente y reprogramar parcialmente el sistema. En dicho contexto el comportamiento podría ser impredecible si los parámetros de aprendizaje (datos de entrada) no fuesen lo suficientemente conocidos y comprendidos por un operador.

Una tercera dimensión sería la tipología de las decisiones y las funciones que se automatizarían dentro de un SAAL (movilidad, selección de objetivos, inteligencia, interoperabilidad, detección de fallos, etc.). Los parámetros de autonomía variarían en cada uno de ellos y las implicaciones jurídicas serían diferentes. Por lo tanto, según establecen M. Boulanin y V. Verbrugge, se debería establecer una lista de comprobación, con dos preguntas fundamentales, para la revisión de los medios y los métodos de combate:

- Con relación a las características técnicas, capacidades y efectos intencionados en condiciones normales de uso: ¿se podría establecer que el SAAL sería capaz de cumplir con los preceptos del DIH?
- Si el SAAL pudiese seleccionar y disparar de forma autónoma: ¿en qué circunstancias podría el uso del sistema violar el DIH?

Dependiendo de la respuesta a dichas preguntas, la revisión podría establecer restricciones de uso o proponer recomendaciones de control (hombre-máquina y mando-y-control), de acuerdo con el Artículo 36 del PAI. Al mismo tiempo, el(los) algoritmo(s) de dicho sistema se debería(n) probar y evaluar, incluyendo la revisión de los mecanismos existentes para minimizar un uso no intencionado o un ciberataque al sistema (Boulanin y Verbrugge, 2017: 20-23; CICR, 2006: 24; Sehrawat, 2017: 48-49).

Si se tomase como ejemplo el sistema de armas de tipología iraní Shahed 136, un posible sistema reactivo, habría que estudiar si dicho sistema cumpliría con las directrices establecidas por el Artículo 36 del PAI con relación al DIH. De acuerdo con dicha revisión, se podrían establecer restricciones de uso, por ejemplo, a través de restricciones de venta a terceros países o proponer recomendaciones de control revisando los mecanismos existentes para que pudiese existir un CHS en todas las fases de su desarrollo y uso. Del mismo modo habría que realizar dichas comprobaciones sobre los drones turcos STM Kargu II o el Bayraktar TB2, sistemas deliberativos, en caso de su posible uso por parte de Ucrania. Análisis que también se extendería a otros SAAL que se utilizasen por parte de ambos contendientes durante dicho conflicto (*Noticias de Israel*, 2022; STM, 2022; Svitlyk, 2022).

2.5. El impacto de la ciberguerra en el DIH

Todas las leyes y en especial las relativas al DIH fueron desarrolladas para un mundo analógico y en la actualidad tiene bastantes problemas para adaptarse al mundo digital. Dado que no se vislumbra un nuevo tratado internacional en un futuro cercano en el ciberespacio, se tendrá que ir a un conflicto armado con el marco jurídico existente y no con el que se desearía tener, con la problemática añadida de su aplicabilidad real[11]. Ahora

[11] Se podría subrayar, como ejemplo, que las grandes potencias como Estados Unidos de América, la Federación Rusa y China no forman parte ni reconocen la jurisdicción de la Corte Penal Internacional (CPI) con sede en La Haya, que investiga y procesa, entre otros, los crímenes de guerra. Aunque se investigasen las posibles acciones contrarias al DIH, sería difícil aplicar las posibles sanciones (Ámsterdam, 2023).

bien, en la actualidad no existe consenso entre los Estados sobre cómo aplicar, en la práctica, el derecho internacional existente del *ius ad bellum* y el *ius in bello* en el ciberespacio. Para tratar de soslayar dicha problemática se han desarrollado instrumentos jurídicos no vinculantes (*soft law*), como los denominados *Manuales de Tallinn* de la OTAN[12]. Con relación al *Manual de Tallinn* sobre la ciberguerra de 2013, la Regla 20.ª sobre la aplicabilidad del DICA establece que: «las operaciones cibernéticas ejecutadas en el contexto de un conflicto armado están sujetos a la Ley de Conflictos Armados (DICA)». Dado que este no regula las actividades en el ciberespacio, se deberá tener en cuenta la «cláusula Martens» de la IV Convención de La Haya de 1907[13]. En cuanto a las Convenciones de Ginebra de 1949 y el PAI de 1977, con relación a la «responsabilidad criminal» de los «comandantes y otros superiores», la Regla 24(a) establece que «los comandantes y otros superiores son criminalmente responsables por ordenar operaciones cibernéticas constituyentes en crímenes de guerra». Sobre el concepto de «levantamiento en masa», la Regla 27 establece que «en un conflicto internacional armado, los habitantes de un territorio no ocupado que lleva a cabo operaciones cibernéticas como parte de un "levantamiento en masa" gozarán de inmunidad de combatiente y el estado de prisionero de guerra». En cuanto al concepto de los «civiles», la Regla 29 establece que «no está prohibido que los civiles participen directamente en operaciones cibernéticas hostiles, pero perderán la protección sobre posibles "ataques" mientras participen». Siguiendo con el mismo concepto, la Regla 32 establece que «la población civil como tal, al mismo tiempo que los individuos civiles, no serán objetivo de un ciberataque» (OTAN, 2013: 68, 80, 88, 90, 97).

Una mención especial se establece con respecto a los objetos de «doble uso» (civil y militar). La Regla 39 establece que «un objeto utilizado tanto para propósitos civiles y militares –incluyendo ordenadores, redes y la infraestructura cibernética– es un objetivo militar». La Regla 43 prohíbe el uso de medios y de métodos de ciberguerra que sean de naturaleza indiscriminada, basándose en los Artículos 51(4)(b) y (c) del PAI. La Regla

[12] Sobre «las Leyes Internacionales Aplicables a la Ciberguerra» y el «Derecho Internacional Aplicable a las Operaciones Cibernéticas» (OTAN, 2013; 2017).

[13] Cláusula Martens (Convención de la Haya IV 1907): Hasta que un código más completo de las leyes de guerra se haya publicado, las Altas Partes Contratantes juzgan oportuno declarar que, en los casos no incluidos en las disposiciones reglamentarias adoptadas por ellas, las poblaciones y los beligerantes quedan bajo la protección y el imperio de los principios de la ley de las naciones, tal como y resultan de los usos establecidos entre naciones civilizadas, de las leyes de la humanidad y los dictados de la conciencia pública (Roberts y Guelf, 1989: 45).

44 establece que «está prohibido el uso de trampas explosivas cibernéticas asociadas con determinados objetos en el DICA». También quedan prohibidas las represalias, de acuerdo con la PAI, a través de la Regla 47. La Regla 70, indica que «el personal médico y religioso, las unidades y los transportes médicos deben ser respetados y protegidos y, en particular, no deben ser un objetivo de un ciberataque». Se complementa con la Regla 71 sobre los ordenadores, sistemas y redes médicas. Por último, la Regla 81 establece que «se prohíbe el ataque, la destrucción, el remover o el hacer inservible aquellos objetos indispensables para la supervivencia de la población civil, por medio de operaciones cibernéticas» (OTAN, 2013: 13, 121-122, 126-127, 167, 169, 185).

En cuanto al *Manual de Tallinn 2.0*, un concepto extremadamente importante es la doctrina de la «responsabilidad», de acuerdo con los artículos establecidos por la «Comisión de la Responsabilidad de los Estados» de las Naciones Unidas. La Regla 14 establece que un «Estado tiene responsabilidad internacional sobre cualquier acto cibernético atribuible a dicho Estado que constituya una violación de una obligación legal internacional». La Regla 15, toma en consideración los Artículos 4.º y 5.º sobre la «Responsabilidad de los Estados», estableciendo que también se consideran órganos de los Estados a los actores (tanto individuos como organizaciones) que, aunque por ley no formen parte de un Estado, sí tengan una «dependencia completa» del mismo. Por otro lado, la Regla 17 trata de los actos «por delegación» (*proxy*). De acuerdo con las leyes internacionales, las operaciones cibernéticas llevadas a cabo por actores no estatales, pero que estén bajo «un control efectivo» de un Estado, entonces dichos actos serían atribuibles a dicho Estado. En cuanto al aspecto de las «contramedidas», los Expertos estuvieron de acuerdo que dichas contramedidas no podían violar una norma perentoria y que deberían ser proporcionales al daño recibido, aunque no existiría la necesidad de que dichas contramedidas cibernéticas tuviesen como objetivo el mismo órgano estatal que hubiese violado la ley internacional. No obstante, volvemos a subrayar que dichos manuales son de aplicación voluntaria por los Estados y se limitan a una posición occidental (Jensen, 2017: 750-751, 754).

A nivel global, de especial relevancia ha sido la adopción por consenso, en 2019, por parte del CCW, pero sin validez jurídica vinculante, de los principios rectores formulados por el GGE sobre los SAAL. Especialmente, habría que subrayar el primer principio rector que determina que: «El DIH sigue aplicándose plenamente a todos los sistemas de armas, incluido el posible desarrollo y uso de los SAAL», pues tiene un gran calado para el mundo cibernético, sobre todo porque fueron adoptados, entre otros Es-

tados, por las grandes potencias: Rusia, China y USA. En dicho contexto la posición de Rusia queda reflejada en su «Documento de Trabajo» sobre la aplicabilidad de dichos principios rectores de 2019[14]. En cuanto a China, aunque no existe una postura sobre dichos principios, en 2018 desarrolló su posición sobre las reglas humanitarias internacionales de los SAAL, que deberían estar de acuerdo con las Convenciones de Ginebra de 1949 y los PAI y PAII de 1977, con una mención especial de los principios de precaución, distinción y proporcionalidad. Finalmente, con relación a los Estados Unidos de América (USA), sus «Comentarios sobre los Principios Rectores», de 2020, abogan por la necesidad de que el GGE del CCW sobre los SAAL aclare los requerimientos del DIH y su aplicabilidad sobre las tecnologías emergentes, incluyendo la idea de que un SAAL fuese incapaz de ser usado de acuerdo con los principios de distinción y proporcionalidad, pues entonces dicho sistema sería considerado ilegal (NU, 2018c: 2; NU, 2019a:. 11-12; NU, 2019c; NU, 2020: 2).

Como continuación, el informe del GGE del CCW sobre los SAAL de 2022 establece que cualquier acto ilegal de un Estado, incluyendo los relativos al uso potencial de SAAL, implicaría la responsabilidad internacional de dicho Estado de acuerdo con el derecho internacional. También destacaríamos el documento de trabajo presentado por Finlandia, Francia, Alemania, los Países Bajos, Noruega, España y Suecia durante dichas reuniones. Dichos países abogarían por un enfoque a dos niveles (*two-tier approach*): prohibir los SAAL completamente autónomos que operasen completamente fuera del control humano y de una cadena de mando responsable, así como regular otros SAAL con diversos grados de autonomía para asegurar su adhesión a las reglas y principios del DIH durante todas las fases de desarrollo, despliegue y uso (NU, 2022a: 19; 2022b).

3. Conclusiones

A la vista de la situación actual, no se puede obviar de que existe una debilidad del sistema de gobernanza global, pues se asiste a una «gobernanza jurídica asimétrica» y que, como aprecia el investigador K. Watkin, es una lucha continua «en las fronteras de la ley», situación que también sería aplicable a la guerra de Ucrania. Aunque las NU en sus resoluciones han establecido que el derecho internacional y la Carta de las Naciones

[14] Una postura que debe ser matizada para la Federación Rusa, reiterada durante las conversaciones del GGE sobre los SAAL de 2022, al excluir de la definición de SAAL a los vehículos aéreos no tripulados (como los drones) y los sistemas de armas altamente autónomos ya existentes. La nueva postura, en consecuencia, podría ser un intento de blindaje ante posibles futuras acciones penales internacionales, con relación a los SAAL utilizados en la guerra de Ucrania, que había comenzado a principios de dicho año (NU, 2022c).

Unidas son de aplicación al ciberespacio, la problemática surge cuando se intenta establecer la forma de su aplicación y si para ello es necesario establecer un nuevo instrumento jurídico global. Teniendo en cuenta los intercambios entre Estados, la situación geopolítica actual, los debates en la NU estancados y la reticencia de las grandes potencias a aceptar la jurisdicción de los tribunales penales internacionales, como en el caso del CPI, será difícil que a corto o medio plazo se establezca un nuevo tratado, pues los Estados prefieren medidas internas a nivel estatal y solo contemplar el intercambio de información de manera voluntaria. En dicho contexto, la única posibilidad actual sería el desarrollo de instrumentos jurídicos no vinculantes (*soft law*), que en un futuro se pudiesen ser incorporados al derecho consuetudinario, mientras se intenta diseñar nuevos algoritmos de IA, como los AMA, que pudiesen servir de vehículo para implantar el DIH en los SAAL (Watkin, 2016).

BIBLIOGRAFÍA

Fuentes primarias

Boletín Oficial del Estado (*BOE*). *Instrumentos de Ratificación de los Protocolos I y II adicionales a los Convenios de Ginebra de 12 de agosto de 1949, relativos a la protección de las víctimas de los conflictos armados internacionales y sin carácter internacional, hechos en Ginebra el 8 de junio de 1977*. *BOE*, 1989, 177, pp. 23828-23863.

Comité Internacional de la Cruz Roja (CICR). (1977). *Protocolo I adicional a los Convenios de Ginebra de 1949 relativo a la protección de las víctimas de los conflictos armados internacionales 1977*. Organización de las Naciones Unidas. [Consulta: 4 de febrero de 2022]. Disponible en: https://www.icrc.org/es/document/protocolo-i-adicional-convenios-ginebra-1949-proteccion-victimas-conflictos-armados-internacionales-1977

Comité Internacional de la Cruz Roja (CICR). (2005). *Customary International Humanitarian Law (Volume II: Practice, Part I and II)*. Henckaerts J. M. y Doswald-Beck, L. (eds.). Comité Internacional de la Cruz Roja. Cambridge University Press. [Consulta: 4 de febrero de 2022]. Disponible en: https://www.icrc.org/en/doc/assets/files/other/customary-international-humanitarian-law-ii-icrc-eng.pdf

Comité Internacional de la Cruz Roja (CICR). (2006). *A Guide to the Legal Review of New Weapons, Means and Methods of Warfare.*

Measures to Implement Article 36 of Additional Protocol I of 1977. Comité Internacional de la Cruz Roja. [Consulta: 10 de febrero de 2022]. Disponible en: https://www.icrc.org/en/doc/assets/files/other/icrc_002_0902.pdf

Comité Internacional de la Cruz Roja (CICR). (2007). *El derecho internacional humanitario consuetudinario (Volúmen I: Normas)*. En: Henckaerts, J. M. y Doswald-Beck, L. (eds.). Comité Internacional de la Cruz Roja, Centro de Apoyo en Comunicación para América Latina y el Caribe. [Consulta: 4 de febrero de 2022]. Disponible: en https://www.icrc.org/es/doc/assets/files/other/icrc_003_pcustom.pdf

Comité Internacional de la Cruz Roja (CICR). (2016). *Expert Meeting: Autonomous Weapons Systems. Implications of Increasing Autonomy, in the Critical Function of Weapons (Versoix, 15-16 marzo 2016)*. Comité Internacional de la Cruz Roja. [Consulta: 4 de febrero 2022 a las 6:40 horas). https://shop.icrc.org/autonomous-weapon-systems-implications-of-increasing-autonomy-in-the-critical-functions-of-weapons-print-en

Comité Internacional de la Cruz Roja (CICR). (2022). *Armed Conflict in Ukraine: a reca pon basic IHL rules.* 22 de marzo. [Consulta: 12 de noviembre de 2022]. Disponible en: https://blogs.icrc.org/law-and-policy/2022/03/17/armed-conflict-in-ukraine-a-recap-of-basic-ihl-rules/

Consejo de la Unión Europea (CUE). (2014). *Rome Declaration on Responsible Research and Innovation in Europe (21 November 2014).* Consejo de la Unión Europea. [Consulta: 14 de febrero de 2022]. Disponible en: https://www.sis-rri-conference.eu/wp-content/uploads/2014/12/RomeDeclaration_Final.pdf

International Criminal Tribunal for the Former Yugoslavia. (ICTY). (2000). *Final Report to the Prosecutorby the Committee Established to Review the Nato Bombing Campaign Against the Federal Republic of Yugoslavia.* Naciones Unidas. [Consulta: el 8 de febrero de 2022]. Disponible en: https://www.icty.org/en/press/final-report-prosecutor-committee-established-review-nato-bombing-campaign-against-federal#IVA64d

Naciones Unidas (NU). (2018a). *Categorizing lethal autonomous weapons systems – A technical and legal perspective in understanding LAWS (Submitted by Estonia & Finland)*, Organización de las Naciones Unidas, [Consulta: 6 de febrero de 2022]. Disponible en: https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2018/gge/documents/GGE.2-WP2.pdf

Naciones Unidas (NU). (2018b). *Informe de 2018 del Grupo de Expertos Gubernamentales sobre Sistemas Armamentísticos Autónomos Letales (SAAL) (23 de octubre 2018)*. Organización de las Naciones Unidas. [Consulta: 10 de febrero de 2022]. Disponible en: https://undocs.org/es/CCW/GGE.1/2018/3

Naciones Unidas (NU). (2018c). *Position Paper submitted by China,* Organización de las Naciones Unidas. [Consulta: 12 de febrero de 2022]. Disponible en: https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2018/gge/documents/GGE.1-WP7.pdf

Naciones Unidas (NU). (2019a). *Informe final de la Reunión de las Altas Partes Contratantes en la Convención sobre Prohibiciones o Restricciones del Empleo de Ciertas Armas Convencionales que Puedan Considerarse Excesivamente Nocivas o de Efectos Indiscriminados*. Organización de las Naciones Unidas. [Consulta: 12 de febrero 2022]. Disponible en: https://undocs.org/es/CCW/MSP/2019/9

Naciones Unidas (NU). (2019b). *Potential opportunities and limitations of military uses of lethal autonomous weapons systems (Submitted by the Russian Federation)*. Organización de las Naciones Unidas. [Consulta: 6 de febrero de 2022]. Disponible en: https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2019/gge/Documents/GGE.2-WP1.pdf

Naciones Unidas (NU). (2019c). *Working Paper of the Russian Federation National Implementation of the Guiding Principles on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*. [Consulta: 12 de febrero de 2022]. Disponible en: https://documents.unoda.org/wp-content/uploads/2020/09/Ru-Commentaries-on-GGE-on-LAWS-guiding-principles1.pdf

Naciones Unidas (NU) (2020). *U.S. Commentaries of the Guiding Principles.* Organización de las Naciones Unidas. [Consulta: 12 de febrero de 2022]. Disponible en: https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2020/gge/documents/US_2020.pdf

Naciones Unidas (NU). (2022a). *Report of the 2022 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems.* Organización de las Naciones Unidas. [Consulta: 12 de noviembre de 2022]. Disponible en: https://meetings.unoda.org/ccw/convention-certain-conventional-weapons-group-governmental-experts-2022

Naciones Unidas (NU). (2022b). *Working paper submitted by Finland, France, Germany, the Netherlands, Norway, Spain, and Sweden to the 2022 Chair of the Group of Governmental Experts (GGE) on emerging technologies in the area of lethal autonomous weapons systems (LAWS)*. Organización de las Naciones Unidas. [Consulta: 15 de noviembre de 2022]. Disponible en: https://meetings.unoda.org/meeting/57989/documents?f%5B0%5D=author_documents_%3AMultiple%20States

Naciones Unidas (NU). (2022c). *Working paper of the Russian Federation 'Application of International Law to Lethal Autonomous Weapons Systems (LAWS)'*. Organización de las Naciones Unidas. [Consulta: 1 de marzo de 2022]. Disponible en: https://documents.unoda.org/wp-content/uploads/2022/07/WP-Russian-Federation_EN.pdf

Organización del Tratado del Atlántico Norte (OTAN). (2013). *Tallinn Manual on the International Law Applicable to Cyber Warfare – Prepared by the National Group of Experts at the invitation of the NATO-CCDCOE*. Cambridge University Press.

Organización del Tratado del Atlántico Norte (OTAN). (2017). *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations – Prepared by the National Group of Experts at the invitation of the NATO-CCDCOE*. Cambridge University Press.

Fuentes secundarias

Allen, C., Varner, G. y Zinser, J. (2000). Prolegomena to any future artificial moral agent. *Journal of Experimental Theory in Artificial Intelligence*. 12, pp. 251-261.

Amodei, D. *et al.* (2016). *Concrete Problems in AI Safety,* Cornell University. [Consulta: 17 de febrero de 2022]. Disponible en: https://arxiv.org/pdf/1606.06565.pdf

Ámsterdam, L. R. B. (2023). Hipocresía en la justicia internacional. La culpa de que no se pueda juzgar a Putin es también de EE. UU. (y Francia o Reino Unido). *El Confidencial 22 de marzo.* [Consulta: 31 de marzo de 2023]. Disponible en: https://www.elconfidencial.com/mundo/2023-03-22/culpa-juzgar-putin-eeuu-francia-reino-unido_3596963/

Anderson, K. y Waxman, M. (2013). Law and Ethics for Robot Soldiers. *Policy Review.* 176, pp. 35-49.

Arkin, R. (2007). *Governing Lethal Behaviour: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture.* Technical Report GIT-GVU-07-11, Georgia Institute of Technology. [Consulta: 17 de febrero de 2022]. Disponible en: https://www.cc.gatech.edu/ai/robot-lab/online-publications/formalizationv35.pdf

Arkin, R. (2013). Lethal Autonomous Systems and the Plight of the Non-Combatant. *AISB Quaterly.* [Consulta: 5 de febrero de 2022]. https://www.researchgate.net/publication/319912057_Lethal_Autonomous_Systems_and_the_Plight_of_the_Non-combatant

Article 36. (2014). Key areas for debate on autonomous weapons systems. *Memorandum for delegates at the Convention on Certain Conventional Weapons (CCW). Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS).* Geneva, 13-16 May. [Consulta: 10 de febrero 2022]. Disponible en: https://article36.org/wp-content/uploads/2014/05/A36-CCW-May-2014.pdf

Article 36. (2016). *Meaningful Human Control, Artificial Intelligence and Autonomous Weapons. Briefing paper for delegates at the Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS).* Geneva, 11-15 April. [Consulta: 10 de febrero de 2022]. Disponible en: https://article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf

Asaro, P. M. (2012). On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making. *International Review of the Red Cross.* 94(886), pp. 687-709.

Benjamins, R. (2021). A choices framework for the responsible use of AI. *Artificial Intelligence and Ethics.* 1, pp. 49-53.

Bode, I. y Huelss, H. (2018). Autonomous weapons systems and changing norms in international relations. *Review of International Studies.* Vol. 44(3), pp. 393-413.

Boulinin V. y Verbruggen, M. (2017). *Article 36 Reviews. Dealing with the challenges posed by emerging technologies.* Stockholm International Peace Research Institute.

Brown, G. D. y Metcalf, A. O. (2014). Easier Said than Done: Legal Reviews of Cyber Weapons. *Journal of National Security Law and Policy.* 7, pp. 115-138.

Carson, J. F. (2020). Defining semi-autonomous, automated and autonomous weapons systems in order to understand their ethical challenges. *Digital War.* 1, pp. 173-177.

Cervantes, J. A. *et al*. (2020). Artificial Moral Agents: A Survey of the Current Status. *Science & Engineering Ethics.* 26, pp. 501-532.

Chengeta, T. (2017). Defining the emerging notion of "Meaningful Human Control" in Weapons Systems. *International Law and Politics.* Vol. 49, pp. 833-890.

Defense Post (THE). (2022). russia's use of Iranian Drones in Ukraine 'appalling': Blinken. *The Defense Post 28 de octubre.* [Consulta: 12 de febrero de 2022]. Disponible en: https://www.thedefensepost.com/2022/10/28/russia-use-iranian-drones-ukraine-appalling/

Dover, R. M. (2022). Ukraine War: Putin is rewriting the rules of siege warfare this Winter. *The Conversation.* 3 de noviembre. [Consulta: 12 de noviembre de 2022]. Disponible en: https://theconversation.com/ukraine-war-putin-is-rewriting-the-rules-of-siege-warfare-this-winter-193425

Ekelhof, M. (2019). Moving beyond semantics on Autonomous Weapons: Meaningful Human Control in Operation. *Global Policy.* 10(3), pp. 343-348.

Floridi, L. y Sanders, J. W. (2004). On the Morality of Artificial Agents. *Minds and Machines: Journal for Artificial Intelligence, Philosophy and Cognitive Science.*14(3), pp. 349-379.

Fossa, F. (2018). Artificial Moral Agents: moral mentors or sensible tools? *Ethics and Information Technology.* 20, pp. 115-126.

Foy, J. (2014). Autonomous Weapons Systems: Taking the human out of International Humanitarian Law. *Delhousie Journal of Legal Studies.* 23, pp. 47-70.

Gade, E. K. (2010). Defining the Non-Combatant: How do we determine who is worthy of protection in violent conflict. *Journal of Military Ethics.* 3, pp. 219-242.

Galliott, J. y Scholz, J. (2018). Artificial Intelligence in Weapons. The moral imperative for Minimally-Just Autonomy. *Journal of Indo-Pacific Affairs.* Winter, pp. 57-67.

Geib, R. y Lahmann, H. (2012). Cyber warfare: applying the principle of distinction in an interconnected space. *Israel Law Review.* Vol. 45(3), pp. 381-399.

Heyns, C. (2016). Autonomous Weapons Systems: living a dignified life and dying a dignified death. En: Butha, N. C. *et al.* (eds.). *Autonomous Weapons Systems. Law, Ethics, Policy (ebook)*. Cambridge University Press, pp. 4-20.

Holland Michel, A. (2021). *Known Unknowns. Data Issues and Military Autonomous Systems.* United Nations Institute for Disarmament Research (UNIDIR).

Human Rights Watch (HRW). (2012). *Losing Humanity. The case against killer robots.* Human Rights Watch. [Consulta: 5 de febrero de 2022]. Disponible en: https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots

Human Rights Watch (HRW). (2016). *Killer Robots and the Concept of Meaningful Human Control. Memorandum to Convention on Conventional Weapons (CCW) Delegates.* April. Human Rights Watch. [Consulta: 10 de febrero de 2022]. Disponible en: https://www.hrw.org/news/2016/04/11/killer-robots-and-concept-meaningful-human-control

Hurka, T. (2005). Proportionality in the Morality of War. *Philosophy & Public Affairs.* 33(1), pp. 34-66.

Jensen, E. T. (2017). The Tallinn Manual 2.0: Highlights and Insights. *Georgetown Journal of International Law*. 48, pp. 735-778.

Johnson, D. G. (2006). Computer Systems: Moral entities but not moral agents. *Ethics and Information Technology.* 8, pp. 195-204.

Klincewicz, M. (2015). Autonomous Weapons Systems, the Frame Problem and Computer Security. *Journal of Military Ethics.* 14(2), pp. 162-176.

Kossov, I. (2022). How Russia uses Iranian drones to try to overwhelm Ukrainie's air defence. *The Kyiv Independent.* 24 de octubre. [Consulta: 13 de noviembre de 2022]. Disponible en: https://kyivindependent.com/national/russias-gambit-to-exhaust-ukraines-air-defense-with-iranian-kamikaze-drones

Krishnan, A. (2009). *Killer Robots. Legality and Ethicality of Autonomous Weapons.* Routledge.

Lin, P., Bekey, G. y Abney, K. (2008). *Autonomous Military Robotics: Risk, Ethics and Design.* Ethics & Emerging Science Group, California Politechnic State University. [Consulta: 17 febrero 2022]. Disponible en: https://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1001&context=phil_fac

Lorenzo Ponce de León, R. (2012). Las Reglas de enfrentamiento (ROE) como paradigma del Estado de derecho en operaciones militares. *Revista Española de Derecho Militar.* 99, pp. 37-220.

Malle, B. y Scheutz, M. (2014). Moral competence in social robots. *IEEE International Symposium on Ethics in Engineering, Science*

and Technology. [Consulta: 13 de febrero de 2022]. Disponible en: https://hrilab.tufts.edu/publications/malleschеutz14ieee.pdf

Marchant, G. E. y Allenby, B. (2017). Soft law: new tools for governing emerging technologies. *Bulletin of the Atomic Scientists.* 73(2), pp. 108-114.

Marín Martínez, A. P. (2021). *Ciberética, Agentes Morales Artificiales y Responsabilidad Jurídica Internacional (inédita).* Tesis Doctoral, Facultad de Derecho Universidad de León.

Mcintyre, A. (2018). Doctrine of Double Effect, *Standford Encyclopedia of Philosophy.* Standford University. [Consulta: 8 de febrero de 2022]. Disponible en: https://plato.stanford.edu/entries/double-effect/#:~:text=According%20to%20the%20principle%20of,about%20the%20same%20good%20end.

Mittelstadt, B. D. *et al.* (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society.* 3(2), pp. 1-21.

Noticias De Israel. (2022). El dron kamikaze Shahed-136 de Irán: Todo lo que necesita saber. *Noticias de Israel.* 21 de octubre. [Consulta: 12 de noviembre de 2022]. Disponible en: https://israelnoticias.com/militar/el-dron-kamikaze-shahed-136-de-iran-todo-lo-que-necesita-saber/

Roff, H. M. (2015a). Lethal Autonomous Weapons and Jus Ad Bellum proportionality. *Case Western Reserve Journal of International Law.* Vol. 47-1, pp. 37-52.

Roff, H. M. (2015b). Autonomous or 'Semi' Autonomous Weapons? A distinction without difference. *Huffington Post*. Verizon Media. [Consulta: 4 de febrero 2022]. Disponible en: https://www.huffpost.com/entry/autonomous-or-semi-autono_b_6487268#:~:text=First%2C%20autonomous%20weapons%20are%20those,intervention%20by%20a%20human%20operator.&text=Semi%2Dautonomous%20weapons%20have%20at,not%20three%2C%20of%20these%20functions.

Rosert, E. y Sauer, F. (2019). Prohibiting Autonomous Weapons: Put Human Dignity First. *Global Policy*. Vol. 10(3), pp. 370-375.

Sartor, G. y Omicini, A. (2016). The autonomy of technological systems and responsibilities for their use. En: Bhuta, N. (ed.). *Autonomous Weapons Systems: Law, Ethics, Policy*. Cambridge University Press, pp. 39-74.

Saxton, A. (2016). (Un)Dignified Killer Robots? The Problem with the Human Dignity Argument. *Lawfare.* [Consulta: 5 de febrero de 2022]. Disponible en: https://www.lawfareblog.com/undignified-

killer-robots-problem-human-dignity-argumentSchmitt, M. N. (2012). 'Attack' as a Term of Art in International Law: The Cyber Operations Context. En: Czosseck, C., Otis, R. y Ziolkowski, K. (eds.). *2012 4th International Conference on Cyber Conflict*. Organización del Tratado del Atlántico Norte (OTAN) – CCDCOE, pp. 283-293.

Segura, C. (2022). Iranian 'suicide' drones: russia's new favorite weapon in Ukraine war. *El País Internacional* 12 de octubre. [Consulta: 12 de noviembre de 2022]. Disponible en: https://english.elpais.com/international/2022-10-12/iranian-suicide-drones-russias-new-favorite-weapon-in-ukraine-war.html

Sehrawat, V. (2017). Autonomous weapon systems: Law of armed conflict (LOAC) and other legal challenges. *Computer Law and Security Review.* 33, pp. 38-56.

Sharkey, A. (2017). Can robots be responsible moral agents? And why should we care?. *Connection Science.* 29(3), pp. 210-216.

Svitlyk, Y. (2022). Invasion of Ukraine: Bayraktar TB2 strike UAV review. *Root Nation* 3 de marzo. [Consulta: 12 de noviembre de 2022]. Disponible en: https://root-nation.com/en/articles-en/weapons-en/en-bayraktar-tb2-drone-review/

Torrance, S. (2008). Ethics and Consciousness in Artificial Agents. *Artificial Intelligence and Society.* 22, pp. 495-521.

Ulam, S. (1958). Tribute to John von Neumann. *Bulletin of the American Mathematical Society.* Vol. 64(3) part 2, pp. 1-49.

Verdiesen, I., Santoni de Sio, F. y Dignum, V. (2021). Accountability and Control over Autonomous Weapons Systems: A Framework for Comprehensive Human Oversight. *Minds & Machines.* Vol. 31, pp. 137-163.

Wagner, M. (2014). The Dehumanization of International Humanitarian Law: Legal, Ethical and Political Implications of AutonomousWeapons Systems. *Vanderbilt Journal of Transitional Law.* 47, 1371-1424.

Wallach, W. y Marchant, G. (2019). Towar the Agile and Comprehensive International Governance of AI and Robotics. *Proceedings of the IEEE.* 107(3), pp. 505-508.

Watkin, K. (2016). *Fighting at the Legal Boundaries: Controlling the Use of Force in Contemporary Conflict.* Oxford University Press.

Westhues, A. (2020). *Sistemas de Armas Autónomas Letales: ¿Autónomas o Automatizadas?*. Trabajo de Fin de Máster. Máster

Universitario en Paz, Seguridad y Defensa. Instituto Universitario General Gutiérrez Mellado.

Xeridia. (2021). Redes Neuronales Artificiales: Qué son y cómo se entrenan. *Xeridia.* [Consulta: 17 de febrero de 2022]. Disponible en: https://www.xeridia.com/blog/redes-neuronales-artificiales-que-son-y-como-se-entrenan-parte-i

Young, J. E., Sharlin, E. y Igarasghi, T. (2011). What is mixed reality anyway? Considering the boundaries of mixed reality in the context of robots. En: Wang, X. (ed.). *Mixed Reality and Human Robot Interaction.* Springer, pp. 1-11. [Consulta: 4 de febrero de 2022]. Disponible en: https://www.researchgate.net/publication/226611963_What_Is_Mixed_Reality_Anyway_Considering_the_Boundaries_of_Mixed_Reality_in_the_Context_of_Robots

Zieba, S. *et al*. (2010). Principles of adjustable autonomy: a framework for resilient human-machine cooperation. *Cognition, Technology & Work.* 12, pp. 193-203.