La Ciberdefensa ofensiva y la Inteligencia Artificial

Offensive Cyber Defense and Artificial Intelligence

Oscar Niss
(GBA - UNICEN)
oscarniss@gmail.com

Resumen: La Ciberdefensa es un Área de Capacidad nueva dentro de las Fuerzas Armadas. A diferencia de los dominios tradicionales donde se desarrollan los conflictos armados – tierra, mar y aire - el *Ciberespacio* posee, además de su anclaje territorial, una importante componente de configuración que le da las características propias de la virtualidad. Sin embargo, no por ello está exento de efectos producidos en el mundo físico ante una operación iniciada en el dominio virtual. En el estado del arte actual, la protección del *Ciberespacio* es mayoritariamente defensiva, por lo que a priori no requeriría de elementos ofensivos para el cumplimiento del objetivo, sin embargo el devenir conceptual y tecnológico está migrando hacia sistemas proactivos y no reactivos, impulsados por tecnologías disruptivas como la Inteligencia Artificial (IA). En ese sentido, el uso dual – civil y militar - de tecnologías para la protección cibernética, conllevaría el riesgo de la propagación de ataques automatizados mediante armas dotadas de IA y sus posibles efectos indirectos en el mundo físico, por lo que es necesario explorar el aspecto normativo de su desarrollo y empleo en escenarios bélicos. Este artículo pretende plantear interrogantes e inquietudes frente al desafío de la irrupción de esta nueva tecnología y su empleo de uso dual, civil y militar.

Abstract: Cyber Defense is a new Capability area within the Armed Forces. Unlike the traditional domains where armed conflicts take place – land, sea and air – Cyberspace has, in addition to its territorial anchorage, an important configuration component that gives it the characteristics of virtuality. However, it is not exempt from effects produced in the physical world when an operation initiated in the virtual domain. In the current state of the art, the protection of Cyberspace is mostly defensive, so in principle it would not require offensive elements to achieve the objective, however the conceptual and technological development is migrating towards proactive and non-reactive systems, driven by technologies. disruptive technologies such as Artificial Intelligence (AI). In this sense, the dual use – civil and military – of technologies for cyber protection would entail the risk of the spread of automated attacks using weapons equipped with AI and their possible indirect effects on the physical world, so it is necessary to explore the normative aspect of its development and use in war scenarios. This article aims to raise questions and concerns regarding the challenge of the emergence of this new technology and its dual use, civil and military.

Palabras Clave: Ciberdefensa Ofensiva; Ciberarmas; Inteligencia Artificial
Keywords: Offensive Cyber Defense; Cyber weapons; Artificial intelligence

Received May 2024; Accepted June 2024; Published July 2024
https://doi.org/10.24215/15146774e061

ISSN 1514-6774

1 Introducción

Sin duda, aunque no sin esfuerzo, podemos encontrar paralelismos o puntos de contacto entre lo que acontece en el mundo físico y lo que transcurre en el ambiente creado por el hombre denominado *Ciberespacio*.

Es interesante recordar que tal definición proviene de la literatura de Ciencia Ficción, en particular en obras del autor *William Gibson*, que luego encontró anclaje en las *Tecnologías de la Información y la Comunicación* (TIC). A diferencia de lo imaginado por Gibson, un ambiente enteramente simbólico a la postre, el de la realidad, el que usamos a diario tiene un enorme componente físico, palpable, amortizable en los balances si se quiere. Otro detalle, no de menor importancia, es que gran parte de ese componente físico, contenedor del lógico, tiene anclaje territorial, por ende, podemos delimitarlo en un adentro y un afuera de las fronteras. Estos atributos no hacen sino manifestarnos un ambiente tensionado y atravesado incluso por aspectos vinculados a la geopolítica, cuando se ha convertido en un activo de importancia estratégica para el desarrollo de cualquier sociedad. Así lo entiende la normativa Argentina cuando expresa, en su segunda propuesta de *Estrategia Nacional de Ciberseguridad*, que el *Ciberespacio* está concurrido por:

> Servicios esenciales para la vida de las personas y para la economía, como la energía, el agua, el transporte, las comunicaciones, la educación, la salud, el comercio y los servicios financieros, entre otros, tienen en la actualidad una fuerte dependencia de las redes informáticas. (República Argentina PEN, 2023).

De acuerdo a lo expresado por este país, que va en línea con lo que entienden la mayoría de las naciones, se puede decir que el *Ciberespacio* atañe al quehacer de los gobiernos y al deber ser de los Estados, tanto que la Organización de las Naciones Unidas (ONU) publicó en uno de sus informes:

> El Grupo [de Expertos Gubernamentales] reafirma las evaluaciones recomendaciones sobre el derecho internacional de los informes de los anteriores Grupos de Expertos Gubernamentales, en particular que el derecho internacional, y en particular la Carta de las Naciones Unidas, es aplicable y esencial para mantener la paz y la estabilidad y para promover un entorno de TIC abierto, seguro, estable, accesible y pacífico. (ONU, 2021).

Así, el *Ciberespacio*, cuya principal red de comunicaciones y almacenamiento de datos es conocida como *Internet*, pasó de ser una herramienta exclusivamente de uso militar en los años ’60, para luego compartir conocimiento entre universidades en los ’80, de comercio a partir de los ’90 y de difusión de contenidos no formales a partir de las redes sociales. Este brevísimo derrotero podría dar una idea de cuáles serían los *Stakeholders* en cada momento.

En ese contexto, las naciones no tardaron en abordar el tema de la soberanía en el *Ciberespacio*. De las muchas definiciones[1] de este nuevo dominio, tomaremos la de la Estrategia Nacional de Ciberseguridad Argentina en su segunda versión:

[1] Téngase en consideración que hay tantas definiciones de Ciberespacio como autor haya. Algunas de ellas incluyen a las personas como parte de ese espacio. En nuestra considera-

> Entorno global compuesto por las infraestructuras de tecnología de la información, incluida Internet, las redes y los sistemas de información y de telecomunicaciones y la dimensión lógica creada a partir del uso de los mismos, que tiene como características esenciales, su dimensión transfronteriza, sin perjuicio de la soberanía de los Estados, su masividad y su vertiginosa y constante evolución. (República Argentina PEN, 2023).

Comprender la naturaleza del dominio *Cibernético* es necesario para poder abordar cuestiones de interés nacional, como son las regulaciones sobre los asuntos de índole soberana[2] y de defensa del territorio, extendiéndolo desde los dominios tradicionales tierra, mar, aire y espacio, al *Ciberespacio*. En ese sentido, se abordan con distintos resultados, temas relacionados al derecho internacional, a la diplomacia y a una nueva área de capacidad en las Fuerzas Armadas.

Esta nueva etapa, en términos de historia del *Ciberespacio*, donde las naciones hacen valer sus derechos soberanos, como puede suponerse, representa un enorme desafío intelectual, tanto desde la dimensión tecnológica como normativa legal. En este último aspecto la Ley 27.078 llamada Argentina Digital, sancionada en diciembre del 2014 adelantaba, en su ARTÍCULO 2°, que "las disposiciones de la presente ley tienen como finalidad garantizar el derecho humano a las comunicaciones y a las telecomunicaciones, reconocer a las Tecnologías de la Información y las Comunicaciones (TIC) como un factor preponderante en la independencia tecnológica…"[3]. Ya se introducían en la norma nacional los conceptos de independencia y derechos humanos en el Ciberespacio, cuestión abordada luego en organismos internacionales.

Al hablar de soberanía, se abre un debate sobre los conceptos consensuados por el concierto de las naciones en los ambientes tradicionales, pero que no están del todo claros en la dimensión Ciberespacial: cuándo existe violación de soberanía, cómo se perfecciona un ataque, cuándo hay uso de la fuerza, la aplicación del *Derecho Internacional Humanitario* (DIH), la no interferencia en los asuntos de otros Estados, las opciones de respuesta ante una violación de soberanía, entre otros desafíos.[4]

---

ción, de ser así, debiera incluirse a las personas en los ambientes tradicionales como tierra, mar y aire desde la dimensión de la Defensa.

[2] En la primera versión de la Estrategia Nacional de Ciberseguridad Argentina del año 2019 se expresa que "Internet representa un dominio global e intangible y un flujo infinito de datos sobre el cual no se ejerce dominio ni soberanía". Nótese las distintas concepciones acorde al posicionamiento del país en ese espacio. Incluso es notable que esa definición vaya en contra de lo expresado por la comunidad internacional en el seno de la ONU, en los distintos informes del GEG.

[3] Nótese que la primera versión de la Estrategia Nacional de Ciberseguridad del año 2019 RESOL-2019-829-APN-SGM#JGM, se contrapone a una norma de mayor orden como la Ley 27078 del año 2014, en lo relacionado a la soberanía en cuánto independencia tecnológica.

[4] En nuestra opinión el debate sobre aspectos normativos, legales y de soberanía en el Ciberespacio llegó tarde. Hay un fuerte *estatus quo* sostenido por algunas naciones y por corporaciones tecnológicas que deja poco lugar al debate, o lo vacía demasiado. El caso de la IA puede correr riesgo de seguir el mismo camino.

El debate sobre los temas planteados es llevado a cabo en el seno de la ONU, con la conformación de un *Grupo de Expertos Gubernamentales* (GEG) *sobre los avances en la información y las telecomunicaciones en el contexto de la seguridad internacional*, emitiendo un primer informe en el año 2010. Han existido tres GEG que han examinado las amenazas existentes y potenciales en el ámbito Cibernético y las posibles medidas de cooperación entre naciones para abordarlos. En su introducción expresa que "cada vez son más numerosos los informes de que los Estados están desarrollando tecnologías de la información y las comunicaciones como instrumentos de guerra y para fines de inteligencia y políticos" (ONU A/65/201, 2010). Como puede observarse si hablamos de instrumentos de guerra, necesariamente hay que abordar de qué trataría ese instrumento en el nuevo ambiente.

En ese mismo sentido, la Resolución Ministerial 105/2023 del Ministerio de Defensa de la República Argentina, aprobó la *Política de Ciberdefensa*, que responde a las directivas del Presidente de la Nación, que mediante Decreto 457/2021 denominado *Directiva Política de Defensa Nacional* (DPDN) instruye a ese Ministerio a "contemplar e incluir el desarrollo doctrinario, planeamiento, diseño y elaboración de la política de Ciberdefensa en el nuevo Ciclo de Planeamiento de la Defensa Nacional"[5]. Como se ve, las diferentes dimensiones de la problemática planteada por el *Ciberespacio*, son reconocidas por las naciones y así lo expresan en sus normas. En este caso, por primera vez en nuestro país, el Ciclo de Planeamiento incluye esta dimensión de manera expresa y no sólo en sus considerandos.

El Decreto 457/2021 reconoce que "si bien las acciones de ciberguerra poseen su origen en el ámbito virtual de los sistemas informáticos y las redes de comunicación, también pueden impactar sobre el mundo físico" (ARG DCTO-2021-457-APN-PTE). La Ley de Defensa 23.554/1988 enmarca que en su conjunto que el planeamiento debe estar basado en una estrategia defensiva y disuasiva, autónoma y cooperativa. Esta situación se da por excelencia en la defensa del Ciberespacio. Sin embargo, la cuestión ofensiva en este ambiente está expresada en toda documentación ya sea nacional o internacional, como iniciativa o como preocupación. En ese sentido, la *Política de Ciberdefensa* Argentina expresa:

> Es necesario el resguardo soberano del mismo, asegurando la confidencialidad, integridad y disponibilidad de la información contenida, tanto administrativa, como las operacionales de los sistemas de mando y control, incluido los sistemas de armas que por su tecnología así lo requieran. (Ministerio de Defensa RESOL-2023-105-APN-MD, 2022)

[5] Hay que hacer una consideración, en el sentido que la Ciberdefensa alude a una capacidad de las Fuerzas Armadas, esto ocurre en casi todos los países; en cambio cuando hablamos de Ciberseguridad se hace referencia a políticas y normas para lograr un estado de menor exposición al riesgo cibernético que ocurre en una red informática, donde se incluyen procedimientos, tecnología y formación de recursos humanos. Como en el caso del Ciberespacio, hay más de una definición. Por otro lado, el Ciberdelito, alude a conductas tipificadas penalmente que se cometen mediante dispositivos electrónicos o contra un sistema informático.

Sumado al problema ya planteado de la conformación del Ciberespacio debemos agregar lo que las tecnologías disruptivas presentan como desafío, tanto al momento de su adopción como al momento de normar. Esto ocurre con el advenimiento de la quinta generación de tecnologías de telefonía móvil (5G), el entrelazamiento cuántico para las comunicaciones, la computación cuántica y la Inteligencia Artificial (IA), entre otras. Sabido es que la velocidad de difusión de la tecnología crece, particularmente en los últimos años, dado el notable acortamiento del ciclo de vida de los productos como consecuencia del incesante y creciente flujo de innovación tecnológica. Este ritmo complejiza aun más los intentos regulatorios y de *control* por parte de los Estados, desafiando su comprensión y plexo normativo.

En particular la IA incide de manera similar a lo planteado en su momento para el *Ciberespacio*, en lo referido a la construcción de sentido, su *significante* y el ulterior debate ético y normativo. La DPDN toma nota de la problemática expresando que:

> Se están perfeccionando los usos militares de las tecnologías robóticas, cibernéticas, de inteligencia artificial y de sensores remotos. Las innovaciones científicas y tecnológicas de años recientes exhiben un paradigma de avanzada, que ha llevado a los sectores de defensa de varios países a ensayar estrategias que perfeccionan la protección ante potenciales ataques. En casos específicos, esas innovaciones se han traducido en capacidades militares ofensivas. (ARG DCTO-2021-457-APN-PTE, p 39)

Sin duda la IA agrega un componente harto complejo a la problemática conceptual y de uso de las Ciberarmas dotadas de autonomía. Pero, para poder avanzar en algunas aproximaciones es necesario respondernos algunas preguntas, ¿De qué trata una Ciberarma y qué características la dotan de automatismo? ¿Cuándo existe uso de la fuerza en el espacio Cibernético? ¿Podría una Ciberarma autónoma distinguir objetivos militares de objetivos civiles? ¿De quién sería la responsabilidad de un Ciberataque autónomo?

2 Desafíos, Ciberarmas y Automatismo

Debemos en primer lugar, procurar discernir qué es una *Ciberarma*. Los paralelismos con el mundo físico, si bien son de utilidad para dar forma a objetos intangibles, no siempre deparan una feliz estrategia para la comprensión, pero pueden acercarnos a una interpretación. Una definición sencilla de *Ciberarma* nos la proporciona *la Junta Interamericana de Defensa* (JID): “Software específicamente diseñado para causar un daño o efecto perjudicial a un elemento del ciberespacio pudiendo tener consecuencias físicas en los ámbitos de operaciones convencionales” (Junta Interamericana de Defensa, p. 13).

Sin embargo, el caso que nos ocupa del automatismo, requiere profundizar en esa definición para permitirnos un posterior abordaje de las responsabilidades. Haciendo un paralelismo con las armas convencionales, podríamos asociar una *Ciberarma* con la munición y concebirla como una única pieza o bien separarla al estilo convencional. Podríamos decir que el arma o vector de ataque es por ejemplo un correo electrónico que lleva la munición o *carga explosiva* como archivo adjunto; también hay

quienes consideran al usuario de una computadora como vector, siendo en este caso responsable de trasladar la carga explosiva mediante un desprevenido acceso a un enlace o link que la descargaría. Entendemos en estos casos como vector, al elemento que porta o que traslada la *carga explosiva* hacia el blanco, siendo la carga explosiva en el mundo cibernético, una *Pieza de Software Maliciosa* (PSM). También podríamos considerar como arma o vector, a un archivo válido adjunto a un correo electrónico, que en el interior de su código llevase una serie de instrucciones maliciosas[6]. Resumiendo, la *Ciberarma*, de alguna manera, trasladará la PSM hacia el objetivo, en general hacia el interior de un dispositivo tecnológico con el fin de producir determinados efectos.

La complejidad, si hasta aquí no lo es, aparece cuando a la *Ciberarma* la dotamos de automatismo, con algoritmos de IA y *Maching Learning* (IA/ML). Tenemos que diferenciar aquí un concepto. Hay armas convencionales, de efectos exclusivamente *kinéticos*, como por ejemplo el *sistema antitanque Spike LR2* con inteligencia artificial de la firma *Rafael Advanced Defense Systems*. Este arma dispone de un buscador electro óptico que incluye un sensor infrarrojo no refrigerado y un sensor diurno de color de alta definición para el apartado del guiado del misil. Según la firma, "este nuevo buscador incluye capacidades de rastreo de objetivos mediante IA para mantener el bloqueo del objetivo, casi sin necesidad de intervención del tirador". Abundan los ejemplos de *sistemas de armas*[7] que han sido modernizados con elementos TIC e IA. Drones y hasta elementos para la logística militar como la iniciativa del Ejército Argentino, a través de su Facultad de Ingeniería, para el desarrollo de un *Vehículo Autónomo de Exploración* (VAE). Si bien estos elementos podrían clasificarse como Sistemas de Armas Autónomos ya que buscan, identifican y completan su misión de manera independiente, sin intervención humana, en este artículo nos referiremos a lo que definimos como *Sistemas de Ciberarmas Autónomos* (SCA), esto es, que cumplen su misión desde el *Ciberespacio* y siendo su *Ciberterreno Clave*[8] las capas cognitiva, lógica o TIC del Ciberespacio, independientemente del efecto *kinético* que pueda provocar en el mundo físico.

---

6 Llamamos pieza o instrucciones maliciosas a código de programación que pretende provocar un daño en la computadora o dispositivo al que accedió.

7 Sistema de Armas es el conjunto de medios, elementos asociados, técnicas y procedimientos, cuya integración conforma un instrumento de combate eficaz para el logro de un efecto determinado. Cuando se habla de un efecto determinado, no implica el empleo de las armas necesariamente. De ahí que un radar, considerado sistema de armas de apoyo operativo, es capaz de realizar efectos tales como vigilar, detectar, alertar, dirigir, sin que en ningún caso implique el empleo de armamento. (fuente FAA).

8 El Ciber Terreno Clave (CTC) es el conjunto de elementos del ciberespacio, en cualquiera de sus capas (humana, ciberhumana, cognitiva, lógica, TIC y geográfica), que facilitan las actividades, operaciones o funciones esenciales para la misión y cuya destrucción, interrupción o captura generaría una ventaja operativa para el adversario. (Fuente JID – Guía de Ciberdefensa 2020)

Son entonces, *Sistemas de Ciberarmas Autónomos* (SCA), los que son empeñados en *Operaciones Multidominio*[9] desde el dominio Cibernético, en sus capas Lógica y/o TIC. Si bien los SCA pueden influir en la maniobra operacional a través de efectos *kinéticos*, estos siempre tienen su origen en la dimensión Cibernética.

Estos sistemas de armas entonces, que se empeñan desde el entorno ciberespacial, pueden estar dotadas de automatismo mediante lógicas de IA/ML. En este punto es necesario distinguir "entre sistemas automáticos y sistemas autónomos explicando que los primeros funcionan con instrucciones pre programadas para llevar a cabo una tarea específica, mientras que los segundos actúan dinámicamente para decidir, cuándo y cómo llevar a cabo una tarea" (ICRC, 2014, p. 5).

Los sistemas automáticos actúan basándose en instrucciones deterministas (basadas en reglas), mientras que los sistemas autónomos con IA actúan sobre la base de un razonamiento estocástico (basado en probabilidades), que introduce incertidumbre, pero también utiliza otras técnicas, como el razonamiento deductivo, el razonamiento inductivo y el aprendizaje automático, entre otros. El razonamiento estocástico se utiliza por ejemplo para modelar la variabilidad sobre la captura de imágenes o sonidos; por ejemplo, detectar patrones, y así asignar una probabilidad. El razonamiento deductivo, basado en el aprendizaje previo, agrega eficacia, permitiendo a la IA resolver problemas específicos que requieran de inferencias precisas y más exactas. Podemos seguir agregando características al método resolutivo de una IA/ML, aunque aun así podrían conllevar un margen de error probable e impreciso en la etapa de diseño de la pieza. En esa instancia se determina el tipo de algoritmo de reconocimiento utilizado, se plasman reglas y se procede a su entrenamiento[10]. El reconocimiento preciso de patrones es necesario tanto para un SCA como para un SAA.

Otra característica no menos importante de las SCA es que el tiempo transcurrido entre la inyección de la munición en la infraestructura a atacar y los efectos, pueden variar entre fracciones de segundos a años. En ese sentido, una *Ciberoperación*, al igual que una operación convencional, tiene una cadena de etapas: el reconocimiento del objetivo o infraestructura a atacar; la preparación de la *Pieza de Software Maliciosa*; la entrega; la explotación a fin de posicionarse o encontrar objetivos; la instalación una vez encontrado; la fase de mando y control (C2) donde puede recibir instrucciones desde un sistema externo que puede estar operado por una IA/ML; y por último las acciones sobre el objetivo. Las etapas previas a la inyección y las posteriores pueden durar un tiempo indefinido. Las acciones sobre el objetivo, por ejemplo si hablamos de una operación de *exfiltrado* de información, puede durar meses. En casos, el tiempo para producir los efectos, puede ser de años incluso, ya que la pieza permane-

[9] El foco de las Operaciones Multi-dominio "se centra en todas las dimensiones del campo de batalla y no en una amenaza determinada" (Fuente - Johnson, 2018, p.6).

[10] Sin embargo, otros factores que intervienen en la efectividad para el caso de Sistemas de Armas Autónomos son los dispositivos sensores utilizados, que captarán imágenes, sonidos, condiciones ambientales y por supuesto el entorno o ambiente en que se encuentre el objetivo estando en operaciones.

ce en estado latente hasta que se decide su activación, tal es el caso de las *Amenazas Persistentes Avanzadas* (APT)[11].

Resumiendo, los SCA, buscan, identifican y defienden / atacan objetivos de manera independiente, sin intervención humana, produciendo efectos primarios en el ambiente Ciberespacial o efectos *kinéticos* secundarios, utilizando técnicas de IA/ML.

3 Uso de la Fuerza en el Ciberespacio

Habiendo hecho una aproximación a lo que constituiría una Ciberarma o un SCA, debemos avanzar sobre en qué casos se constituiría el *uso de la fuerza* en este dominio, ya que no todo ataque conlleva esa calificación. Claramente un GEG se expresó en ese sentido:

> Al examinar la aplicación del derecho internacional a la utilización de las TIC por los Estados, el Grupo consideró de importancia fundamental los compromisos de los Estados con los siguientes principios de la Carta y de otro derecho internacional: (...) abstenerse en sus relaciones internacionales de la amenaza o el uso de la fuerza contra la integridad territorial o la independencia política de cualquier Estado, o de cualquier otra manera incompatible con los propósitos de las Naciones Unidas; (...). (ONU A/70/174)

A fin de entender el uso de la fuerza en el Ciberespacio, Noruega hizo una contribución en un informe de la ONU, apoyada por muchos países miembros donde entiende que:

> Una *Ciberoperación*[12] puede constituir el uso de la fuerza o incluso un ataque armado si su escala y sus efectos son comparables a los del uso de la fuerza o de un ataque armado por medios convencionales. Esto debe determinarse sobre la base de una evaluación caso por caso teniendo en cuenta las circunstancias específicas. (Sentencia de 27 de junio de 1986, ICJ Rep., 1986, p. 14)

Está clara la complejidad desde el momento que Noruega propone el tratamiento *caso por caso*. Sin embargo, podemos procurar conceptualizar la constitución de las operaciones en el Ciberespacio como manera de entender luego el uso de la fuerza. En ese sentido, las *Ciberoperaciones* pueden restringir sus efectos al dominio cibernético o bien extenderlas al mundo físico, de manera indirecta. Las *Operaciones Multi-Dominio*[13] se diferencian por cuanto integran el espacio físico y el *Ciberespacio* como dimensiones del campo de batalla de carácter no lineal y sin límites físicos ni geográficos convencionales. Operar en el Ciberespacio y en el espacio presenta un gran desafío para los efectos no *kinéticos* y fuegos no letales, por cuanto son extrema-

[11] APT por su denominación en Ingles *Advanced Persistent Threats*, se trata de un Ciberataque que se prolonga en el tiempo y dirigido por un proceso de *Mando y Control* en el que el atacante obtiene acceso a una infraestructura y permanece sin ser detectado por un período indefinido.

[12] Las Ciberoperaciones son acciones militares planificadas, organizadas, coordinadas y llevadas a cabo por unidades de ciberdefensa con la finalidad de lograr efectos en el ciberespacio, así como en los otros ámbitos de operaciones.

[13] En esencia hablamos de operaciones que se desarrollan en los dominios físico y cibernético.

damente difíciles de medir en forma cuantitativa y física[14]. "Por ende, el proceso de Evaluación del daño en Batalla (EDB) resulta extremadamente complejo para medir el nivel de rendimiento (NDR) de un Ciberataque sobre un nodo crítico adversario" (Miranda, p. 7).

Vemos así la problemática de cuantificar los efectos producidos a fin, en nuestro caso, de determinar si hubo o no uso de la fuerza. Como vemos, no basta con utilizar un Ciberarma sino que se deben evaluar los efectos buscados o producidos. Para ello, una serie de indicadores propuestos en los debates académicos que cooperan en la determinación del uso de la fuerza, podrían sintetizarse en los siguientes:

i. el origen de la operación y la naturaleza del instigador (militar o no);
ii. la naturaleza del objetivo previsto, por ejemplo, el carácter militar de la infraestructura atacada;
iii. el alcance de la intrusión/gravedad del ataque;
iv. los efectos reales o previstos de la operación;
v. la inmediatez de los efectos;
vi. la profundidad de penetración de la infraestructura cibernética.

Es importante destacar que un GEG reconoció la aplicación del Artículo 2 de la Carta de Naciones Unidas (ONU), donde en resumidas cuentas se insta a los países miembros a: respetar la jurisdicción de los Estados sobre el territorio, incluida la infraestructura de TIC que se encuentra allí; la prohibición del uso de la fuerza; la prohibición de la intervención interna de la intervención interna de otros Estados; la obligación de respetar el territorio soberano de otros Estados; la obligación de no permitir a sabiendas que su territorio sea utilizado para actos contrarios a los derechos de otros Estados; y la obligación de respetar los derechos humanos.

Nótese, del párrafo anterior, la cantidad de líneas de investigación o debates que se desprenden para el caso que nos ocupa. Porque la IA/ML puede utilizarse en cualquiera de los casos que la ONU toma en consideración. Pero para este artículo, nos centraremos en la *prohibición del uso de la fuerza*.

Reforzando lo dicho en el apartado anterior, debemos tener en cuenta que una vez catalogada una pieza de software (PSM) como *Ciberarma*, debemos analizar una serie de indicadores para determinar, luego de una operación, si su empleo constituyó el uso de la fuerza o no.

Si bien nos estamos centrando en el uso de Ciberarmas por parte de los Estados, debemos considerar que la mayoría de ellas están diseñadas para *uso dual*, esto es civil y militar. También, como en lo convencional, existe apropiación de Ciberarmas diseñadas para uso exclusivo militar, comercializada y utilizada en operaciones civiles por parte de delincuentes o grupos para estatales. Sin embargo, a diferencia de las armas convencionales, debemos tener en cuenta que su diseño y producción está al alcance de un individuo con los conocimientos necesarios, con bajísimos costos y además con ayuda hoy de la IA. Capítulo aparte es el uso de la IA para el diseño de armas, que ameritaría un proyecto de investigación, donde sería interesante preguntarse sobre la regulación de la IA para estos casos.

[14] Adaptación del texto OPERACIONES MULTI-DOMINIO: SOLUCIONES TÁCTICAS PARA DESAFÍOS ESTRATÉGICOS Y OPERACIONALES de Osvaldo Alaniz Miranda.

Resumiendo, para determinar si hubo *uso de la fuerza*, materia condenable en ese caso, se deben analizar cada uno de los factores que determinan el umbral. Este análisis es similar y representa similar complejidad sea que se trate de un Ciberarma con o sin IA/ML.

4 Identificación de Objetivos y Responsabilidades

Un elemento estrechamente vinculado al concepto y diseño de un SCA está relacionado a la obligación del sistema de distinguir objetivos militares de civiles. Recordemos los Convenios de La Haya de 1907, cuya finalidad primordial consiste en limitar la guerra a ataques contra objetivos necesarios para el resultado de las operaciones militares. La población civil, por consiguiente, debe de ser protegida contra los ataques militares. Esto también es válido en el ambiente Cibernético, aunque algunos informes consideran que es importante distinguir los dos aspectos principales de la aplicabilidad del DIH: uno se refiere a la si el DIH rige las operaciones cibernéticas que complementan las operaciones militares kinéticas existentes durante los conflictos armados; y el otro a la cuestión de si las operaciones cibernéticas por sí mismas - sin operaciones kinéticas- pueden estar reguladas por el DIH.

Otro elemento de consideración es el *Manual de Tallin*[15], que tiene como finalidad reafirmar y aclarar el derecho internacional que regula la guerra cibernética, incluidos el derecho que rige el empleo de la fuerza entre Estados (*jus ad bellum*) y el derecho que rige la conducta de las partes en conflictos armados internacionales y no internacionales (*jus in bello*). El manual no aborda las actividades cibernéticas que tengan lugar por debajo del umbral del "empleo de la fuerza" o de un conflicto armado, ni tampoco analiza la cuestión de los derechos humanos. Por todo ello es clara la determinación del uso de la fuerza en el ambiente cibernético a fin de considerar las normas que aplican a los conflictos armados.

En ese sentido, los tratados imponen distinguir los objetivos civiles de los militares, que son los necesarios para el cumplimiento de una misión, por lo que los SCA también deben incorporar esta característica propia de la etapa de diseño, ya que un arma con IA/ML no *enseñada* a distinguir el tipo de objetivo, no lo hará al momento de una misión. Debemos distinguir aquí también las armas empleadas en una *Ciberoperación Ofensiva*, de una Defensiva[16]. El SCA con capacidad de producir efectos que superen el umbral del uso de la fuerza, en general son de tipo ofensivos.

A fin de evaluar la capacidad de un SCA de reconocer el tipo de objetivo, es interesante observar el modelo de *Cyber Kill Chain*[17], que propone la estructura de un Ciberataque.

---

[15] Téngase en cuenta que el Manual de Tallin es un compendio académico no vinculante.

[16] Desde el punto de vista del planeamiento existen distintos tipos de Ciberoperaciones: defensivas, ofensivas, reactivas, anticipativas, exploratorias y otras.

[17] Cadena de Ciberexterminio (Cyber Kill Chain) es un modelo diseñado por la empresa Lockheed Martin. En general de uso militar que identifica la estructura de un ataque, también se lo utiliza para los Ciberataques.

En ella, las etapas de un Ciberataque tienen tres momentos que lo ordenan: *preparatorio, intrusión y acciones.* Cada una de las etapas podría hacer uso de distintos sistemas de armas e incluso todas podrían estar dotadas de IA/ML. Por supuesto estamos hablando aquí de *Operaciones* complejas en el *Ciberespacio*, también conocidas como *Amenazas Persistentes Avanzadas* (APT), que ya hemos mencionado. Una operación de este tipo es llevada a cabo por unidades con algún grado de sofisticación importante o bien por Estados y están dirigidas a un objetivo específico.

En ese sentido, bajo la hipótesis que planteamos donde estas operaciones necesitan de estas etapas para cumplirse, podríamos distinguir en cada una de ellas una dinámica y misión diferenciadas, que juntas hace a la operación completa. No es intención de este artículo profundizar sobre cada una, pero sí procurar discernir dónde la IA/ML debiera saber distinguir los distintos tipos de objetivos.

Así, en la etapa preparatoria de *reconocimiento* el atacante realiza "una huella digital (*fingerprinting*) del objetivo para crear un esquema o mapa de sus redes y sistemas TIC, estructura organizativa, relaciones, comunicaciones y afiliaciones e identificar sus vulnerabilidades, tanto técnicas como humanas, para posteriormente poder infiltrarse y explotar la red" (Junta Interamericana de Defensa, p. 33).

El conjunto de herramientas que se utilizan en esa etapa de inteligencia no necesariamente constituye un sistema de armas, de hecho por lo general no suele calificarse así al conjunto de recursos utilizados para ese fin. Sin embargo, sí la información recolectada es el *insumo* básico para la construcción de un SCA propiamente dicho. Pero avancemos a la siguiente etapa, la de la *weaponization*[18] o de preparación del arma, veamos que dice la JID:

> En la fase de preparación, en base a la información obtenida en la fase de reconocimiento, el conocimiento detallado de los recursos propios y los tipos de efectos deseados, se planifica el ciberataque, se seleccionan las herramientas más eficaces y se produce el armado de la carga útil (malware y exploits idóneos para explotar las vulnerabilidades conocidas o desconocidas) en los vectores de ataque (documentos *pdf* o *word*, dominios web comprometidos, emails suplantados, dispositivos de memorias *usb*, etc.). (Junta Interamericana de Defensa, p. 33)

Si pensamos que estas dos etapas, como las demas, podrían estár utilizando tecnologías de IA y su subconjunto de *Maching Learning*[19](ML), podríamos suponer con poco margen de error que serían claves para la identificación de objetivos. Una herramienta cuya misión es el reconocimiento de infraestructuras tecnológicas que luego serán los objetivos del ataque, debiera poder reconocer la *propiedad* militar del objeto estudiado. Sin duda esto es más claro de entender cuando hablamos de Sistemas de Armas Autónomos que opera en la capa *kinética*; distinguir una base militar de un hospital no parecería del todo complejo. Pero aquí estamos analizando los SCA donde el objetivo está en el ambiente cibernético, aunque pudiendo tener efectos físicos indirectos. Sin duda, la etapa de diseño de las herramientas de *reconocimiento*, como

[18] *Weaponization*: militarización, en contextos de operaciones militares, es el término utilizado para la preparación del artefacto.

[19] Técnica utilizada para el aprendizaje automatizado de algoritmos en base la experiencia, son un subconjunto de la IA.

la de la *munición*, dotadas de IA/ML, deben contemplar en sus algoritmos la capacidad de reconocimiento y aprendizaje posterior para la distinción de los dos objetivos con alta presición.

En ese sentido, los ejércitos de las principales potencias como EEUU estudian la evaluación de infraestructura cibernética automatizada, así lo expresa un informe de la *United States Special Operations Command* (USSOCOM): "la SOF [Fuerza de Operaciones Especiales] está interesada en tecnologías que puedan proporcionar la detección automática, geolocalización y caracterización del terreno cibernético dentro de determinadas áreas de interés". (Agency: United States Special Operations Command, 2020, p. 12)

De avanzar en ese sentido, donde la *detección automática* de objetivos proporcione un aceptable grado de certeza, y de estar bien diseñados y entrenados, los sistemas autónomos debieran tener *menos posibilidades de ser utilizados de modo mal intencionado o erróneo contra objetivos civiles*, ya que las caracterizaciones precisas de los terrenos o *Ciberterrenos clave* serían más certeras.

Sin embargo, otra complejidad que se presenta, es que existen infraestructuras que podrían ser *subsidiarias* al Sistema de Defensa, constituyendo un *objetivo* en momentos de una operación militar. En estos casos la supresión de un objetivo civil sería necesaria para el cumplimiento de una misión militar. Queda abierto el interrogante ¿quién tomaría esa decisión?. Al igual que en las armas convencionales, las responsabilidades en su diseño y empleo están bastante claramente divididas; con las *Ciberarmas* ocurre lo mismo, pero sin duda la IA/ML inclinaría más hacia el lado del diseñador, la balanza de la responsabilidad, ya que *–a priori-* no sería tan sencillo torcer su aprendizaje y dirigirla contra un objetivo para el cual no fue diseñada.

Para finalizar, las pautas de reconocimiento de objetivos y su distinción, incluidas en la etapa de *diseño*, constituyen el punto clave donde poner la atención al momento de exigir el cumplimiento de pactos internacionales en lo referente al DIH.

5 Conclusión

Como se ha visto a lo largo de este artículo, la complejidad presentada por las tecnologías de IA/ML aplicadas a la *Ciberdefensa Ofensiva* en operaciones militares, exceden al análisis que pueda hacerse desde una sola dimensión y requieren de un abordaje multidisciplinario.

Las *Ciberarmas* con IA/ML o SCA - si bien su empleo debiera estar restringido en algunos casos al uso militar - se venden en mercados de la *Deep o Dark Web*[20] al que puede acceder una organización civil. Por otro lado el diseño desaprensivo de un Malware utilizando una IA está disponible para cualquier persona con algunos conocimientos técnicos, si bien no representan una verdadera amenaza a infraestructuras de magnitud, se debe considerar la regulación de las funciones de las IA en cuanto al diseño de piezas que puedan ocasionar daños al ser humano.

[20] Denominación asignada a la web profunda no indexada por los buscadores convencionales.

El conocido, por estos días, *CHAT-GPT* de *OpenAI*, consultado sobre el sesgo en la programación de los algoritmos de las *Ciberarmas* dice que:

> Es importante que quienes diseñan y programan las Ciberarmas con inteligencia artificial tomen medidas para minimizar el sesgo en los datos y en el algoritmo, y para garantizar que las decisiones que tome la Ciberarma sean justas y éticas. Esto implica la necesidad de implementar procesos rigurosos de verificación y validación para evaluar la efectividad y la imparcialidad de las Ciberarmas antes de ser utilizadas. (CHAT-GPT)

El problema seguramente sea quién *verifica y valida*. Quizá para el caso de los *Sistemas de Ciberarmas Autónomos*, deban pensarse en organismos o acuerdos específicos entre las naciones. Como se dijo al comienzo, las tecnologías disruptivas se *llevan por delante* las normas, corren delante de ellas. Aún no se canceló el debate sobre la *Paz y Seguridad en el Ciberespacio*, dado en el seno de la ONU y ya se le agrega una complejidad más con la irrupción de la IA/ML.

Otra dimensión de la IA/ML para uso militar, que excede este trabajo, es la generación de información falsa para interferir en la maniobra del oponente. Basta citar el informe del *United States Special Operations Command* (USSOCOM) donde propone el uso de estas nuevas tecnologías para desarrollar "operaciones de influencia, engaño digital, interrupción de la comunicación y campañas de desinformación en el borde táctico y los niveles operativos" (*Agency: United States Special Operations Command*, 2020, pág. 16). Se abre aquí una línea de investigación que llega a los límites de la convivencia entre naciones, si se quiere. “Es una tecnología peligrosa”, dijo Rizzuto, investigador del *Atlantic Council*. “No se puede moderar esta tecnología de la misma manera que abordamos otros tipos de contenido en Internet”, dijo. “Los *deepfakes* como tecnología tienen más en común con las conversaciones sobre la no proliferación nuclear”.

Esta tecnología de la IA/ML, disruptiva por excelencia, plantea nuevos paradigmas en el desarrollo de la *Ciberdefensa Ofensiva* y los conflictos armados, que deben ser abordados con la suficiente antelación que otras tenologías disruptivas también requerían y no tuvieron. El aspecto legal, las normas y las leyes del derecho internacional, y la diplomacia para el mantenimiento de la seguridad y la paz en el Ciberespacio están nuevamente tensionadas con el avance acelerado de la Inteligencia Artificial.

Referencias

1. Agency: United States Special Operations Command. (2020). BROAD AGENCY ANNOUNCEMENT USSOCOM-BAAST-2020.
2. ARG DCTO-2021-457-APN-PTE. (s.f.). Directiva de Política de Defensa Nacional2021.
3. ICRC. (2014). Report of the ICRC Expert Meeting on ‘Autonomous weapon systems.
4. Junta Interamericana de Defensa. Guía de Ciberdefensa.
5. Ministerio de Defensa RESOL-2023-105-APN-MD. (2022). Política de Ciberdefensa. Argentina.
6. Miranda, O. A. OPERACIONES MULTI-DOMINIO: SOLUCIONES TÁCTICAS PARA DESAFÍOS ESTRATÉGICOS Y OPERACIONALES.

7. ONU A/65/201. (2010). Grupo de Expertos Gubernamentales sobre los avances en la información y las telecomunicaciones.
8. ONU A/70/174.
9. ONU. (2021). Informe de las Naciones Unidas A/76/135.
10. República Argentina PEN. (02 de 1 de 2023). https://www.argentina.gob.ar. Recuperado el 28 de 02 de 2023, de https://www.argentina.gob.ar/normativa/nacional/resolución-1-2023-377806/texto
11. Sentencia de 27 de junio de 1986, ICJ Rep. (1986)ONU A/65/201. (2010). Grupo de Expertos Gubernamentales sobre los avances en la información y las telecomunicaciones .
12. ONU A/70/174.
13. ONU. (2021). Informe de las Naciones Unidas A/76/135.
14. República Argentina PEN. (02 de 1 de 2023). https://www.argentina.gob.ar. Recuperado el 28 de 02 de 2023, de https://www.argentina.gob.ar/normativa/nacional/resolución-1-2023-377806/texto
15. Sentencia de 27 de junio de 1986, ICJ Rep. (1986).