



PUBLICADO EN SEPTIEMBRE DE 2021

INTELIGENCIA ARTIFICIAL Y SISTEMAS AUTÓNOMOS DE APLICACIÓN DE LA FUERZA

Generando conceptos que faciliten el
debate y orientación de estrategias

“ DEBE CONSIDERARSE QUE NO HAY
NADA MÁS DIFÍCIL DE LLEVAR A CABO,
NI MÁS DUDOSO DE ÉXITO, NI MÁS
PELIGROSO DE MANEJAR, QUE INICIAR
UN NUEVO ORDEN DE COSAS. ”

NICOLÀS MAQUIAVELO

AUTORES

MG. LUCIANA MICHA, DRA. VANINA MARTÍNEZ,
COMODORO PABLO FARÍAS Y DR. RICARDO RODRÍGUEZ



TABLA DE CONTENIDOS

CAPÍTULO 1

| | |
|---------------------------|---|
| PRECISIONES METODOLÓGICAS | 4 |
|---------------------------|---|

CAPÍTULO 2

| | |
|----------------------------------|---|
| ANÁLISIS DE LAS AWS DESDE EL DIH | 8 |
|----------------------------------|---|

CAPÍTULO 3

| | |
|-----------------------------------|----|
| MODELIZACIÓN A TRAVÉS DE MATRICES | 17 |
|-----------------------------------|----|

CAPÍTULO 4

| | |
|-----------------------------------|----|
| DEFINICIONES CONCEPTUALES FINALES | 34 |
|-----------------------------------|----|

| | |
|-----------------------|----|
| FUENTES Y BIBLIGRAFÍA | 41 |
|-----------------------|----|

| | |
|------------------------------------------|----|
| SOBRE LOS AUTORES DEL PRESENTE DOCUMENTO | 44 |
|------------------------------------------|----|

CAPITULO 1

PRECISIONES METODOLÓGICAS

El objetivo central del presente trabajo es proporcionar una aproximación conceptual que contribuya al debate sobre la realidad compleja que implica el diseño, desarrollo, empleo e impacto de dispositivos militares autónomos con inteligencia artificial (IA) en el marco del análisis regional e internacional sobre su regulación, limitación y /o prohibición.

Este trabajo presenta un abordaje diferente al actual enfoque discursivo en relación a los **Sistemas de Armas Autónomos** (en adelante, AWS por sus siglas en inglés) el cual coloca como centro de gravedad de los debates sobre el uso, regulación y prohibición enfocados en un **producto** (“máquina o robot”) y valoraciones que surgen de su empleo **meramente táctico** (empañamiento).

Entendiendo la complejidad y desafíos que presentan los sistemas y subsistemas autónomos de aplicación de la fuerza, la presente propuesta amplía el campo de análisis conceptual a fin de brindar **profundidad al análisis de las AWS -sistemas y procesos- y su impacto a lo largo de todo el ciclo de toma de decisión (Político-Operacional-Táctico)**.

El enfoque precedente permite identificar acciones/ estrategias preventivas y/o anticipativas -por sobre la reactivas- visibilizando objetivos y estrategias intermedias (subóptimos) que contribuyan tanto al desarrollo de elementos de juicio desde una aproximación integral e integrada como su contribución a la continuidad de instancias de diálogo y la construcción progresiva de consensos para su

uso, regulación, no proliferación y/o prohibición.

El producto del presente trabajo consiste en el desarrollo de **matrices conceptuales que relacionan las principales dimensiones y procesos asociados a los Dispositivos Autónomos de aplicación de la fuerza o AWS** y el establecimiento de pautas que se desprenden de los principios y espíritu del Derecho Internacional Humanitario, el Derecho Internacional de los Derechos Humanos (DDHH) y los principios rectores para una IA fiable y responsable.

Dichas matrices pretenden sumarse y articularse al desarrollo progresivo de los diálogos internacionales y regionales que actualmente se encuentran en vigencia, especialmente respondiendo a interrogantes claves expresados durante el desarrollo de las actividades de las Naciones Unidas dentro del Grupo de Expertos Gubernamentales 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, con fecha Agosto - Septiembre 2021 (GGE 2021).

Asimismo la presente contribución académica se encuentra en el marco de la profundización de una de las líneas de análisis interdisciplinarias que se desprenden de investigaciones anteriores (Micha- Farías 2021 y Martínez-Rodríguez 2021) por parte de los autores, las cuales buscan establecer criterios objetivos con respecto al desarrollo, regulación, no prolifera-

ción y/o prohibición de esta tecnología dual (civil-militar), en particular aquellos correspondientes al módulo de aplicación de la fuerza.

Resulta pertinente destacar que los distintos componentes y procesos asociados al diseño y desarrollo de las AWS se encuentran en diversos estadios evolutivos: conceptual, embrionario, en evolución, entre otros. Por lo tanto, los estándares que uno puede proponer en este momento, sobre una tecnología en pleno proceso de incipiente evolución, pueden quedar obsoletos frente a desafíos y riesgos futuros. Asimismo, su eventual comercialización y utilización implican un doble riesgo dados los crecientes procesos de privatización del uso de la fuerza por parte de los Estados como de actores no estatales de naturaleza diversa

En cuanto a la **metodología empleada**, ella consistió en el análisis profundo de los diferentes principios del DIH, dimensiones conceptuales de la IA y las características tecnológicas que poseen a la fecha y se les adjudica a las AWS, conjuntamente con algunos de los principios rectores para una IA fiable y responsable (Comisión Europea 2019 y Dignum, V. 2019).

A través de diferentes matrices se abordan e interrelacionan cuestiones claves desde la IA y el DIH a fin de analizar las condiciones de uso de los AWS. Dicho análisis hace énfasis en dos cuestiones críticas, la interacción humana/máquina (human/machine interaction (IHM)) y los principios que debe satisfacer el AWS.

LAS MATRICES CONFECCIONADAS SON:

MATRIZ 1

Interrelaciona **las Funciones/tareas de las AWS** (Búsqueda, Localización, Seguimiento, Chequeo, Enfrentamiento, Evaluación, entre otras) con los Niveles de toma de decisión (político, estratégico, operacional, táctico). Esta matriz pretende focalizar de manera preventiva la IHM en los distintos niveles del proceso de toma de decisión político militar que proporcione una adecuada IHM de acuerdo a los principios y obligaciones del DIH.

MATRIZ 2.1

En el marco del **Dominio Operacional de empleo de las AWS**- en general- interrelaciona las exigencias del ambiente operacional y los niveles de aplicación de la fuerza (capacidad no letal / letal) de las AWS. La matriz permite identificar los tipos de interacción humano/dispositivos (IHM) mínimos exigibles.

MATRIZ 2.2

En el marco específico del Dominio aéreo/ espacial (a modo de ejemplo de un dominio), la matriz interrelaciona las exigencias de las distintas variables del ambiente operacional y la aplicación de la fuerza no letal / letal. La matriz permite identificar los tipos de interacción humano/dispositivos (IHM) mínimos exigibles.

MATRIZ 3

Interrelacionan los Objetivos (tipos de blanco y condición en función a la situación estratégica) y la capacidad Letal/ No letal de las AWS. La matriz permite identificar los tipos de interacción humano/dispositivos (IHM) mínimo aceptable en el sistema.

MATRIZ 4

Interrelaciona las Funciones/tareas de las AWS con principios de la IA (Previsibilidad, Comprensión, Justificación / Accountability y Explicación). Esta matriz identifica el grado de exigencia requerido para cada uno de los principios críticos de la IA en función a las tareas/ funciones de los AWS (ALTO-MEDIO-BAJO). Dicha metodología permite vincular el nivel requerido apropiado de un determinado principio para garantizar el control humano significativo en una determinada etapa, en relación a los principios de la IA.

“Resulta imperativo indagar en el debate conceptual del control humano sobre las funciones críticas de los sistemas autónomos en relación con el proceso de comando y control, selección de blancos y uso de la fuerza” (Micha-Farías

2020). Una vez determinada la secuencia de proceso de decisión se debe determinar el grado de control humano significativo a través de la IHM (Matriz 1), seguido por considerar las exigencias del ambiente operacional y características del objetivo contexto (Matrices 2.1, 2.2 y 3). Finalmente, se determinan las exigencias al AWS en relación a los principios de confiabilidad para la IA (matriz M4) con el fin de garantizar el grado de control humano significativo previamente determinado.

Por lo expuesto, las matrices no son independientes entre sí. A fin de evaluar integralmente un AWS se lo debe analizar al mismo y junto con su contexto de diseño y aplicación a través de todas las matrices presentadas para verificar su posible desarrollo, uso o necesidad de limitación, regulación y/o prohibición, pudiendo existir más matrices para un análisis detallado sobre cada sistema/ambiente operacional. En cuanto a la tecnología de la IA, las matrices confeccionadas aplican independientemente de los modelos específicos de IA subyacentes en los AWS, el foco está en la autonomía (human-machine interaction) y los principios de confiabilidad que intrínsecamente deben verificar.

CAPITULO 2

ANÁLISIS DE LAS AWS DESDE EL DIH¹

Para efectuar el análisis sobre las consideraciones desde el DIH y los DD.HH., para el presente trabajo se tuvieron en cuenta el estado del Estado de Arte de expertos juristas, científicos y filósofos que estudiaron el cumplimiento o incumplimiento de las normas del DIH por parte de los AWS², así como también las diversas

posiciones y opiniones consultivas emanadas por diversas organizaciones internacionales y no gubernamentales como el Comité Internacional de la Cruz Roja³ (ICRC), Amnistía Internacional, International Human Rights Clinic y la Campaña Stop Killer Robots (SKR), Human Rights Watch, Article 36, entre otras⁴.

1 Por DIH se entiende al “cuerpo de normas internacionales, de origen convencional o consuetudinario, específicamente destinado a ser aplicado en los conflictos armados, internacionales o no internacionales, y que limita, por razones humanitarias, el derecho de las Partes en conflicto a elegir libremente los métodos y los medios utilizados en la guerra, o que protege a las personas y a los bienes afectados, o que pueden estar afectados, por el conflicto. Definido de esta manera, el derecho internacional humanitario es parte integrante del derecho internacional público positivo, ocupando el lugar del cuerpo de reglas que antes se conocía con la denominación de derecho de la guerra” en CICR: <https://www.icrc.org/es/doc/resources/documents/misc/5tdl7w.htm>

2 PSe destacan: Christof Heyns (2013) Relator Especial del Consejo de Derechos Humanos, Ronald Arkin (2009), Roser Martínez Quirante y Joaquín Rodríguez Álvarez (2018), Carlos Fernando Álvarez González (2017), Noel Sharkey (2012), Martha Isabel Hurtado Granada (2017), Juan Ramón Martínez Vargas, Andrés Felipe López Latorre, Steffany Serebreni Beltrán y Natalia Fernández Alba (2019), Peter Bergen (2014), Gómez de Águeda (2019), Toby Walsh (2020), Stuart Russell, Elon Musk, Stephen Hawking, Noam Chomsky y Peter Norving (2015) entre otros.

3 Posición Oficial del CICR Extraído file:///C:/Users/administrador/Downloads/4550.01_003-ebook.pdf

4 Se realizaron entrevistas en profundidad con diversos especialistas, miembros de la Campaña internacional de SKR y con expertos en DIH de las Fuerzas Armadas de la Argentina, como el Brigadier (R) Armando Bonadío, director del Instituto de Derecho Aeronáutico y Espacial, director y profesor titular del Curso de DIH del INDAE.

Asimismo, se analizaron en profundidad y se realizó un seguimiento de los debates actuales dentro del Grupo de Expertos Gubernamentales - GEG 2021 de la CCW de la ONU, en especial y a requerimiento de su presidente las preguntas centrales sobre: 1.1) ¿Qué características específicas harían inherentemente un sistema de armas para que fuera imposible de utilizar en el marco del DIH? y 1.2) ¿Dónde y cómo trazar el límite entre la legalidad e ilegalidad del uso de los AWS?

A fin de responder a dichos interrogantes y para el desarrollo del presente trabajo se abordan para el primer interrogante en especial los principios de Humanidad, de Responsabilidad y Limitación, sin dejar de lado los otros principios del DIH, pero encontrando en estos, los más críticos frente a los AWS, entendiendo que las mismas tendrían la capacidad de procesar información, buscar, identificar objetivos de acuerdo a patrones de datos a fin de aplicar la fuerza de manera autónoma conforme a criterios de diseño contemplados en algoritmos. Según la definición del CICR, las AWS son consideradas como

“cualquier sistema de armas con autonomía en sus funciones críticas (búsqueda, detección, identificación y rastreo para la selección de un blanco) y atacar (utilizar la fuerza para neutralizar, dañar o destruir) objetivos sin intervención humana”

(Davidson, N., 2017)

En relación a los principios del DIH sobre distinción, proporcionalidad y precaución es evidente y legalmente imperioso que los AWS deberán ser diseñados y utilizados a fin de operar siempre de conformidad con las normas de distinción de combatientes y no combatientes, proporcionalidad, conforme a la cual el número de víctimas civiles que incidentalmente podría provocar un ataque contra un objetivo militar no debe ser excesivo en relación con la ventaja

militar concreta y directa prevista; y operar de modo que permita tomar las precauciones necesarias en el ataque a fin de reducir al mínimo el número de víctimas civiles.

Dichos principios fundamentales del DIH (distinción, proporcionalidad y precaución) son muy importantes para la ejecución, evaluación y legalidad de una acción militar y/o de aplicación de la fuerza, y se entiende que serían potencialmente parametrizables en un AWS desde un punto de vista tecnológico. Aunque se reconoce que estos principios requieren realizar evaluaciones complejas basadas en las circunstancias imperantes, tanto al momento de la toma de decisión como durante el despliegue de una acción militar, consideramos que los mismos pueden ser capturados por reglas lógicas y métricas de performance adecuadas y los principios de la IA fiable y responsable.

En cuanto a los tres principios del DIH que consideramos críticos se encuentra el Principio de Humanidad, que de acuerdo al CICR implica que “Se debe tratar con humanidad a todas aquellas personas que no participen en las hostilidades (incluso miembros de las Fuerzas Armadas que hayan depuesto las armas y las personas que hayan quedado fuera de combate por enfermedad, herida, detención o cualquier otra causa.” En palabras de Robin M. Coupland (2001)

“el principio de humanidad incluye la posibilidad y la voluntad de reducir la capacidad de emplear la violencia armada, así como de limitar los efectos de ésta sobre la seguridad y la salud. Así entendido, el principio de humanidad engloba el humanitarismo, la moralidad, el desarrollo, los derechos humanos y la seguridad de las personas. Por ello, constituye una de las principales fuentes del derecho internacional en general y del derecho internacional humanitario en particular.”

Asimismo, el principio de humanidad incluye que “(...) toda persona que no participa o que ha dejado de participar en las hostilidades debe ser tratada humanamente y no puede ser objeto de discriminación en razón de su sexo, nacionalidad, raza, religión o pensamiento político” (López Díaz, 2009:233). Este principio, se encuentra vinculado a los Derechos Humanos, y se verían en grave contradicción con las AWS, ya que “sería muy difícil garantizar una toma de decisiones por parte de estas máquinas que incluya la no discriminación en el accionar autónomo; incluso, rasgos tan humanos como la empatía” (López Díaz, 2009; SKR, 2021; SE-HLAC, 2021).

Más aún, el Principio de Humanidad se aplica cuando no existiera una norma explícita - aplicación de la “cláusula de Martens”-, de esta manera y en cuanto a las prohibiciones o restricciones basados en el principio de humanidad y los dictados de la conciencia pública, los AWS no podría obrar bajo preceptos humanitarios ni los mandatos de la conciencia pública, debido a que carecen de emociones humanas; los sistemas no posee empatía ni sentimientos de compasión, ni emociones, ni principios éticos relacionado al principio de humanidad.

Tal cual lo expresa el documento del CICR (2006) Guía para la Revisión Legal de Nuevas armas, medios y métodos de guerra - Medidas para la implementación del Art 36 del Protocolo Adicional I de 1977, en su punto 1.2.2.3 (Prohibiciones o restricciones basadas en los principios de humanidad y los dictados de la conciencia pública -cláusula de Martens-), se debe prestar atención al momento de incorporar una nueva, si está respeta los principios de humanidad y los dictados de la conciencia pública, tal como establecen el artículo 1(2) del Protocolo adicional I, el preámbulo de la Convención (IV) de La Haya de 1907, y el preámbulo de la Convención de La Haya (II) de 1899. Allí se hace referencia a la llamada “cláusula de Martens”, que el artículo 1(2) del Protocolo

adicional I enuncia de la siguiente manera

“En los casos no previstos en el presente Protocolo o en otros acuerdos internacionales, las personas civiles y los combatientes quedan bajo la protección y el imperio de los principios del derecho de gentes derivados de los usos establecidos, de los principios de humanidad y de los dictados de la conciencia pública.” (Lawand, K., Coupland, R. y Herbie P. 2006)

En el pasado, la cláusula de Martens fue invocada para justificar la prohibición preventiva de las armas láser cegadoras que tengan como finalidad total o parcial producir ceguera en forma definitiva. Por su parte, la Corte Internacional de Justicia (1996) en el caso de legalidad sobre la amenaza o uso de armas nucleares afirmó la importancia de la cláusula de Martens y su aplicación como derecho internacional consuetudinario, la cual “su continua existencia y aplicación es indudable”

Entendiendo que la evolución de la IA tiene una característica básica de ser una tecnología de blanco móvil en continuo avance y evolución, lo cual implica un elevado dinamismo y velocidad de innovación; en los casos de aplicación de la fuerza, no estaríamos en condiciones de dar una definición exhaustiva de su conformación, alcance, sistemas y subsistemas asociados.

Para el Siglo XXI, teniendo en cuenta los avances tecnológicos en cuanto al desarrollo de la IA y su aplicación a AWS, el principio de humanidad no estaría basado sólo en los efectos y consecuencias que la aplicación de AWS implicaría, sino, que además debe considerar el sistema de armas mismo y procesos de decisión. La humanidad debe estar presente en todo el proceso, no solo en los resultados de la acción violenta, sino también en los medios (instrumentos de aplicación de la fuerza) que se utilizan y los procesos de decisión intrínsecos.

“(…)sin desconocer las eventuales ventajas militares que para determinados escenarios implican el uso de estos ingenios tecnológicos, toda pérdida progresiva de humanidad responsable y trazable en el proceso estratégico-operativo-táctico de la guerra transformará al humano -actor del proceso- en un mero espectador inmerso en un círculo vicioso -trampa estratégica por concepción y diseño- en el que corre el riesgo de perder en forma progresiva el control de su propia creación. Es imperativo evitar las trampas semánticas, ya que la guerra representa los peores aspectos de la naturaleza humana; estas tecnologías disruptivas podrían deshumanizar aún más la guerra” (Micha-Farías 2021).

Es por ello que el principio de la humanidad debe estar presente tanto en los medios, sistemas a emplear y sus consecuencias, entendiendo que debe existir un control humano significativo en todo sistema de armamento, independientemente del avance tecnológico que exista al presente, máxime cuando se trate de armamento anti personal.

En respuesta al segundo interrogante: ¿Dónde y cómo trazar el límite entre la legalidad e ilegalidad del uso de los AWS? En la actualidad no existe una norma específica sobre los AWS, sin embargo, de conformidad con el derecho, los Estados tienen la obligación de determinar si el empleo de un arma nueva, medio o métodos de guerra nuevos que desarrolle o adquiera estaría prohibido por el derecho internacional en ciertas condiciones o en todas las circunstancias, en virtud de lo establecido por el Artículo 36 del Protocolo Adicional I a los Convenios de Ginebra del año 1977.

El Artículo 36 está asociado directamente a otros dos principios rectores del DIH que es el de la Responsabilidad y Limitación. Estos imperativos entienden que, por un lado, los Estados Nacionales reconocen la obligatoriedad del

Art 36 del Protocolo Adicional I y por el otro que el control humano es determinante para la atribución de responsabilidad personal en todas las diferentes etapas de diseño, desarrollo y utilización.

A nivel intergubernamental se reconoce que “debe mantenerse la responsabilidad humana en las decisiones sobre el uso de la fuerza letal” (CCW, 2019). Asimismo, desde el CICR (2021), expresa como una grave limitación “la pérdida de control y juicio humano en el uso de la fuerza y las armas”. Incluso, autores como Daniele Amoroso and Guglielmo Tamburrini (2021) en la publicación titulada “Toward a Normative Model of Meaningful Human Control over Weapons Systems” expresan

“A medida que los sistemas de armas se vuelven más autónomos en sus decisiones de selección de objetivos, el papel de la responsabilidad penal individual (y del Derecho Penal Internacional (DPI) en general) se reduce a la hora de regular el uso de la violencia armada (...) Las limitaciones epistémicas de la capacidad del comandante para hacer predicciones sólidas sobre los comportamientos de los sistemas de alerta temprana son las que impulsan el problema de la accountability gap, especialmente desde la perspectiva del ICL”

El DIH impone obligaciones a los Estados, a los actores presentes en los conflictos armados e individuos, no a sistemas ni máquinas. La responsabilidad humana y la justificación o rendición de cuentas (accountability) deben mantenerse a través de todo el proceso de diseño, desarrollo y uso de un sistema de armas, incluyendo la decisión final en el uso de la fuerza, dado que ni la responsabilidad y ni la justificación (accountability) pueden ser transferidos a un sistema de armas ni a una máquina. En síntesis, “siempre debe haber un RESPONSABLE de la acción militar sobre los humanos” (Brigadier, Bonadío A. 2021). Es imperiosa la

necesidad de asignación de la responsabilidad penal individual en el uso de la fuerza, no solo la estatal (Matthias, A. 2004), (Sparrow, R. 2007).

En cuanto a este principio crítico, existe un doble riesgo, el cual implicaría el desarrollo y utilización de AWS por actores no estatales y la imposibilidad de atribución individual al uso de la fuerza y sanciones posteriores.

En cuanto al Principio de Limitación, implica que las armas y métodos que puedan ser utilizados en los conflictos no son ilimitados. En efecto, el Protocolo Adicional I a los Convenios de Ginebra recoge este principio en el Artículo 36 en los siguientes términos

“En todo conflicto armado, el derecho de las Partes en conflicto a elegir los métodos o medios de hacer la guerra no es ilimitado. Queda prohibido el empleo de armas, proyectiles, materias y métodos de hacer la guerra de tal índole que causen males superfluos o sufrimientos innecesarios”.

En igual sentido la Corte Internacional de Justicia en su Opinión Consultiva sobre la Legalidad o el Empleo de Armas Nucleares indicó que “los Estados no poseen un derecho ilimitado a escoger los medios ni los métodos de combate dentro de un conflicto armado, sino que aquellos se encuentran vinculados a razones humanitarias que efectivamente los limitan”

Lo que no esté expresamente prohibido o restringido, NO ESTÁ PERMITIDO, deberá evaluarse bajo las normas generales del DIH:

La Guía para la Revisión Legal de Nuevas armas, medios y métodos de guerra - Medidas para la implementación del Art 36 del Protocolo Adicional I de 1977 tiene como propósito ayudar a los Estados a establecer o mejorar los procedimientos para determinar la licitud de las armas, los medios y los métodos de guerra

nuevos de conformidad con el artículo 36 del Protocolo adicional I a los Convenios de Ginebra de 1949.

De acuerdo a dicho documento, a fin de determinar la legalidad/ilegalidad de la introducción de un nuevo sistema de armamento por parte de los Estados, se especifica que en su texto en español en el punto 1.2.2:

“Prohibiciones o restricciones generales relativas a las armas, los medios y los métodos de guerra: Si no existe ninguna prohibición o restricción específica aplicable, el arma o medio de guerra bajo examen y los métodos normales o previstos con que se utilizarán deben evaluarse a la luz de las prohibiciones o restricciones generales establecidas por los tratados y por el derecho internacional consuetudinario que se aplican a todas las armas, los medios y los métodos de guerra. (...) esas normas también son pertinentes para la evaluación de la licitud de un arma nueva antes de que haya sido utilizada en el campo de batalla, en la medida en que las características, el uso previsto y los efectos previsibles del arma permiten a la autoridad examinadora determinar si el arma podrá ser empleada o no de manera lícita en ciertas situaciones previstas y bajo ciertas condiciones”.

Por consiguiente, como se establece en dicho documento, si no se encontrara una prohibición específica o restricción, la evaluación sobre la legalidad de un nuevo sistema de armas debe realizarse bajo la luz de las prohibiciones generales y de acuerdo a los principios básicos del DIH y del derecho consuetudinario, los cuales al menos tres principios analizados anteriormente estarían contradiciendo el desarrollo y uso de las AWS, independientemente del avance tecnológico que este tipo de sistemas logre alcanzar a futuro.

Asimismo, en el mencionado documento se

listan los datos empíricos necesarios para ser considerados al momento de la revisión de un nuevo sistema de armas, los cuales determinarían su legalidad o ilegalidad. En su punto 1.3 Datos empíricos que deben considerarse en el examen, expresa que

“A fin de determinar si un arma bajo examen está sujeta a prohibiciones o restricciones específicas o si contradice una o más normas generales del DIH aplicables a las armas, los medios y los métodos de guerra (enumeradas en la sección 1.2.2), la autoridad examinadora tendrá que tomar en consideración varios factores de índole militar, técnica y ambiental, así como factores relacionados con la salud. Por ello, es necesario que participen expertos de varias disciplinas en el proceso de examen” (Lawand, K. Coupland R. y Herby, P. 2006).

Reconociendo, además, que es necesario un enfoque multidimensional a efectos llevar a cabo un examen integral y la importancia de adoptar un enfoque multidisciplinario para efectuar un adecuado examen jurídico de las armas.

El listado, al presente, de los Datos Empíricos requeridos a ser analizados para la eventual legalidad de una nueva arma son: 1) Descripción técnica del arma, 2) Funcionamiento técnico del arma, 3) Consideraciones relativas a la salud y 4) Consideraciones relativas al ambiente.

En cuanto 1.2 el funcionamiento técnico del arma examinada es de particular importancia para determinar si su empleo puede tener efectos indiscriminados. Los factores relevantes serían:

”la precisión y la fiabilidad del mecanismo de ataque (por ejemplo, tasa de fallas, sensibilidad de explosivos sin estallar, etc.); el área cubierta por el arma; si los efectos previsibles del arma pueden limitarse al objetivo o pueden ser controlados en tiempo o

espacio (incluido el grado en que el arma representará un riesgo para la población civil después de haber cumplido su fin militar)”

Dado el avance de la tecnología y la aplicación de la IA a las armas se debería considerar dentro de este listado otros factores de importancia, a fin de medir su performance y garantizar la fiabilidad técnica específica de los AWS, por lo tanto, podríamos agregar a la lista del CICR las siguientes consideraciones sobre los riesgos inherentes y asociado de las AWS, las cuales comprenden entre otras:

- 1) Problemas de diseño, programación y producción.
- 2) Posibilidad de sesgos en los algoritmos (algorithmic bias) debido a subjetividades propias del humano (o grupos) sobre apreciaciones personales/ culturales /éticas (UNIDIR 2018). Dichos sesgos serían un agravante en las armas dirigidas a humanos,
- 3) Inconsistencias o vicios ocultos,
- 4) Datos e información incompleta, ambigua, contradictoria, irrelevante y/o exceso de información (Holland, 2021),
- 5) Posibilidad de ser factible de hackeo, fundamentalmente los que trabajan en sistemas nodales o en red,
- 6) Engaño o interferencia a los sensores y/o fuentes. Acciones adversales (adversail actions) o interferencias a los sensores y / o fuentes de datos
- 7) Problemas para identificar la trazabilidad en la acción para otorgar responsabilidades y la rendición de cuentas donde el uso de la autonomía y los desafíos de la caja negra resultan un obstáculo de importancia crítica,

8) Estándares de funcionamiento y targeting no encuadrados dentro del DIH,

9) Modificación de su accionar a las condiciones de diseño y comportamiento/ desempeños diferentes al que fueron concebidos, a la par de la imposibilidad de resolver situaciones complejas que exigen una adecuada conciencia situacional. (Micha-Farías 2021),

10) Consideraciones asociadas sobre espacio, tiempo, entre otras.

El desafío está en identificar a lo largo de todo el proceso, desde los niveles de decisión como de ejecución, un adecuado balance entre limitación, riesgo y aceptabilidad y cómo transformar este riesgo inherente (incertidumbre) en un escenario de riesgo residual que haga política y militarmente aceptable su empleo en el ámbito de aplicación de los principios del DIH (Micha Farías 2021), lo cual al presente resulta imposible dado su desarrollo inmaduro y a futuro no se subsanaron las limitaciones previstas por los principios de Humanidad, Limitación ni Responsabilidad.

Sumado a todo lo anteriormente mencionado, el CICR (2021B) expresa la necesidad que los Estados establezcan de manera internacional límites consensuados a fin de asegurar la protección de civiles en cumplimiento del DIH y aceptabilidad ética

“instó a los Estados a establecer límites acordados internacionalmente para los sistemas de armas autónomas, a fin de garantizar la protección de los civiles, el cumplimiento del derecho internacional humanitario y la aceptabilidad ética (...) Los sistemas de armas autónomas plantean problemas éticos fundamentales para la humanidad, ya que, en efecto, sustituyen las decisiones humanas sobre la vida y la muerte por procesos sensoriales o de software y máquinas El ICRC recomienda que

los Estados adopten nuevas normas jurídicamente vinculantes”.

En particular el documento hace hincapié en los efectos indiscriminados que dichos sistemas podrían ocasionar si no son previsibles, entendibles y explicables (predictable, understandable, explainable).

Es por ello que el CICR expresa la necesidad de Prohibición y Regulación:

1. Prohibición total de aquellos sistemas diseñados para la utilización de la fuerza contra personas.

2. Regulación en el diseño y uso de las AWS en su versión de uso antisistema de acuerdo a la combinación de:

- Límites en los tipos de blancos, solo permitiéndose objetivos militares por naturaleza.
- Límites en la duración, alcance geográfico y escala de utilización incluyendo la posibilidad de control humano significativo en relación a un ataque específico.
- Límites en las situaciones de uso, como la posibilidad de utilización en escenarios sin presencia de población civil o bienes protegidos.
- Requerimientos para la interacción máquina-humano a fin de asegurar una supervisión humana efectiva, intervención oportuna y desactivación.

En síntesis, a los primeros dos interrogantes del director del GGE (2021) se establece:

1.1 ¿Qué características específicas harían inherentemente un sistema de armas para que fuera imposible de utilizar en el marco del DIH?

En el marco del DIH y en consonancia con los

DDHH y la ética, los tres principios críticos que harían imposible de utilizar los AWS, dada su capacidad de autonomía, serían los principios de **Humanidad, Responsabilidad y Limitación**, y en aquellos sistemas letales antipersonales y armamento antisistema que pongan en riesgo civiles y bienes civiles en un sentido amplio, deberían ser prohibidos preventivamente.

El principio de humanidad implica que, frente a la existencia de la capacidad de **autonomía absoluta de los sistemas de aplicación de la fuerza, este principio crítico no podría ser resuelto, ni hoy ni en el futuro, independientemente del avance de la tecnología**. De acuerdo al principio de Limitación, y a fin de dar inicio al proceso de convalidación legal sobre la incorporación de una nueva arma, es imperativo que se disponga de todas las revisiones legales, consuetudinarias y técnicas expresadas bajo el Art 36 del Protocolo Adicional de 1977, en especial las prohibiciones o restricciones sobre introducción de nuevas armas bajo tratados internacionales, prohibiciones o restricciones bajo Customary International Law, prohibiciones o restricciones basadas en el Principio de Humanidad (Cláusula de Martens).

En este sentido y en relación a las especificaciones técnicas, el/los sistemas deben ser previsibles, **explicables, con existencia demostrable de control humano significativo y asignación de responsabilidades individuales en todo su proceso de diseño, elaboración y empleo** a fin de garantizar el Principio de Responsabilidad, entendiendo por ello las exigencias de una capacidad de autonomía limitada en aplicación de dicho principio.

La revisión de las autoridades para estas nuevas tecnologías de acuerdo al principio de limitación del Art. 36 deben incluir no solo las 4 consideraciones basadas en datos empíricos que incluyen a la actualidad la descripción

técnica del arma, funcionamiento técnico del arma, consideraciones relativas al daño sobre la salud, y el ambiente, sino un **5to elemento sobre fiabilidad técnica y operativa (trustworthiness)** el cual implique que se pueda demostrar justificadamente - teórica y operativamente- que el/los sistemas son previsibles, explicables, con existencia de control humano significativo y asignación de responsabilidades individuales en todo su proceso de diseño, elaboración, empleo y disposición final. En cuanto a los aspectos técnicos las AWS deberán incluir una caracterización completa del sistema, incluyendo la descripción tanto de los modelos de IA, como de los datos, entradas y salidas, que permitan evaluar los riesgos de su despliegue, y deberán contener las métricas adecuadas para medir el desempeño técnico, etc.

En relación al resto de los principios del DIH, **mientras tanto un AWS no pueda demostrar empírica o teóricamente ser capaz de distinguir efectivamente entre combatientes y no combatientes -Principio de Distinción- y que su accionar permita aplicar los principios de Proporcionalidad y Precaución, su utilización debería estar preventivamente prohibida.**

1.2. ¿Dónde y cómo trazar el límite entre la legalidad e ilegalidad del uso de los Sistemas de Armas Autónomas?

Se entiende que no serían legales si no cuentan con todas las revisiones legales, consuetudinarias y técnicas expresadas bajo el Art 36 del Protocolo Adicional de 1977. Asimismo, no podrían ser legales los AWS **sin control humano significativo ni anti personales**. No serían legales (como demostraremos más adelante a través de las matrices conceptuales) aquellas AWS tácticas en ambientes urbanos o densamente poblados, anti personales o que involucran bienes culturales protegidos a fin de asegurar los principios de distinción y precau-

ción. Tampoco, si están en la cadena de decisión de acciones sobre elementos estratégicos o con armas de destrucción masiva, o si no puede asignarse responsabilidad penal individual por los actos que los sistemas lleven a cabo.

A fin de definir el **grado de autonomía** permitido se debe trabajar en metodologías que permitan medir el **grado de confiabilidad y garantías de fiabilidad** que se puede tener en que los sistemas puedan resolver los riesgos que plantean los diferentes ambientes operacionales.

Es por todo lo anteriormente mencionado, que el presente trabajo propone un abordaje matricial, a efectos poder brindar **un marco conceptual para la caracterización de Dispositivos Autónomos de aplicación de la fuerza.**

CAPITULO 3

MODELIZACIÓN A TRAVÉS DE MATRICES.

El marco metodológico que ofrecemos en el presente trabajo permitirá interrelacionar las principales dimensiones de las AWS en relación a la aplicación de la fuerza y sus riesgos inherentes a este tipo de tecnología y autonomía asociada, que contribuyan a determinar de **manera preventiva un adecuado control humano significativo de acuerdo a los principios y obligaciones del DIH y a la luz de los principios de la IA.**

Este ejercicio se elabora sobre la base de conceptos y líneas de pensamiento iniciadas en trabajos previos (Holland, M. 2020), (Micha-Farías, 2021) y (Persi, Paoli, Spazian y Anand, 2021) intentando profundizar en la construcción metodológica de herramientas que permitan avanzar en la construcción de conocimiento y consenso en relación a las AWS.

A través de matrices conceptuales -no excluyentes ni exhaustivas-, se pretende **focalizar en la interacción Humano-máquina (IHM)**

mínima exigible con las funciones/tareas de las AWS, niveles de toma de decisión, dominios operacionales, exigencias del ambiente operacional y contextos, tipos de objetivos y capacidades de aplicación de la fuerza, lo que permitirá evaluar -de manera teórica- las exigencias prohibitivas, restrictivas y permisivas para el diseño, desarrollo, certificación y empleo de las AWS.

En este trabajo proponemos 4 tipos de matrices que deben ser aplicadas para evaluar el grado de IHM en todo el proceso de aplicación de la fuerza. Cada matriz analiza una dimensión diferente del proceso por lo que deben aplicarse **todas** ellas consecuentemente para arribar a una evaluación completa. Estas matrices están pensadas para ser usadas en los procesos de decisión e interacción de los niveles político estratégico y militar con el objeto de dar profundidad al análisis del problema e **identificar acciones, estrategias y/o posiciones preventivas-anticipativas.**

DESARROLLO DE MATRICES

MATRIZ 1

Interrelaciona las Funciones/tareas de las AWS con los Niveles de toma de decisión. La presente individualiza la IHM aceptable para

las diferentes tareas o ***funciones asociados a los AWS*** que podrían realizar en cada ***nivel de toma de decisión***. Esta matriz pretende focalizar de manera preventiva la IHM en los distintos niveles del proceso de toma de decisión político militar que proporcione una adecuada IHM de acuerdo a los principios y obligaciones del DIH.

5 Marco analítico para determinar las variables para la regulación de sistemas de armas autónomas. El análisis está desarrollado bajo la lógica del uso ofensivo de los sistemas de tal manera de poder establecer un esquema general del tipo de control, lo cual puede llegar en casos concretos en acciones defensivas a la necesidad de flexibilizar los grados de control con el objeto de responder en tiempo y forma a las exigencias de la amenaza que necesitarán un trabajo futuro al respecto.

ESTA MATRIZ COMBINA DOS DIMENSIONES A FIN DE ESTABLECER NIVELES ESPECÍFICOS DE IHM:

1. NIVELES DEL PROCESO DE LA DECISIÓN MILITAR Y SUS FUNCIONES (UNIDIR 2019):

- **Nivel Político Estratégico:** concibe, conforme a los objetivos políticos perseguidos, un efecto final deseado, como así también las condiciones para su consecución (ROE) en el marco de la maniobra general de la guerra.

- **Nivel Operacional:** articula el nivel estratégico (objetivo) con el táctico (ejecución); concibiendo la campaña que contempla una estrategia y maniobra preparatoria, efectos operacionales, Centros de Gravedad, Objetivo-blancos, capacidades, esfuerzos, despliegue que posicionan a las organizaciones y elementos del nivel táctico en las mejores condiciones para ejecutar las misiones, operaciones, tareas, acciones que le fueran asignadas.

- **Nivel Táctico:** conduce el empañamiento de los medios (la batalla/el combate) a partir de la aplicación específica de sus capacidades (aplicación de la fuerza), en el logro de los efectos requeridos por el nivel operacional.

La recolección de insumos, el análisis, la difusión a los usuarios, el planeamiento y la decisión están presentes de manera general en cada una de las etapas mencionadas (etapas del proceso de la acción basados en las OODA: observación, orientación, decisión, acción (Ver

descripción completa de cada etapa en Michá-Farías 2021).

Los Resultados en cada celda están dados por Grados de control - Interacción Humano Máquina (Persi, Paoli, Spazian y Anand, 2021) (IHM).

- **FULL:** Sistemas que no toman ninguna decisión por sí mismo, sino que son completamente dirigidos. Se los conoce como IHM de control total “human full control”.

- **IN:** Sistemas que desarrollan la tarea programada/ordenada con autonomía limitada ya que requiere la intervención del sistema de comando y control estratégico/táctico (humano) para validar o implementar acciones específicas. Se conocen como IHM “human in the loop”.

- **ON:** Sistemas que desarrolla la tarea programada/ordenada bajo la supervisión de elementos del sistema de comando y control (operadores humanos) que tiene la capacidad de modificar, corregir o vetar el funcionamiento de subsistemas críticos y/o aquellos requeridos. Se conocen como IHM “human on the loop”.

- **OFF:** Sistemas de Autonomía total, que desarrolla una tarea programada u ordenada sin supervisión o intervención de ope-

radores del Sistema de Comando y Control, conforme a sus capacidades y condiciones de diseño. Se conocen como IHM “human off the loop”.

La Matriz 1 intenta responder la siguiente pregunta: **¿En qué fase del ciclo vital debe**

ejercerse el control humano? Es por ello que Interrelaciona **las Funciones/tareas de las AWS con los Niveles de toma de decisión**. Individualiza la IHM aceptable para las diferentes tareas o funciones asociadas a los AWS que podrían realizar en cada nivel de toma de decisión.

MATRIZ N° 1

| FUNCIONES Y TAREAS DE LAS AWS | NIVELES DE TOMA DE DECISIÓN | | |
|---------------------------------------|---------------------------------------------------------------|-------------------------------------------------------|-----------------------------------|
| | POLÍTICO ESTRATÉGICO Manejo del conflicto | MILITAR Variable estratégica Operacional | TÁCTICO Batalla/combate |
| Exploración y Reconocimiento | OFF | OFF | OFF |
| Localización | OFF | OFF | OFF |
| Identificación | IN | IN | IN |
| Vigilancia | IN | IN | IN |
| Seguimiento | OFF-IN | ON | IN en función a las ROEs |
| Análisis y Evaluación del Riesgo | IN | ON | OFF |
| Empañamiento de Elementos del Combate | (*) | (*) | (*) |
| Proceso de Valoración de Efectos | IN | IN | (*) |

(*) No es posible en este nivel de análisis fijar restricciones sobre el IHM, la evaluación deberá proseguir con la aplicación de las consideraciones que surjan de las otras matrices.

CONSIDERACIONES DE LA MATRIZ 1:

Se asume que el/los sistemas realizan diferentes funciones de manera combinada y/o complementaria. Asumimos entonces que, para el análisis de dichos sistemas, podemos pensarlos en módulos (subsistemas) que realizan funciones específicas.

La valoración de un sistema complejo se hará en todo su conjunto y quedará determinada -de manera general- por **la valoración más restrictiva de alguno de sus módulos.**

Ejemplo: Consideremos un sistema de inteligencia a nivel político que eventualmente puede, de tratarse de objetivos de defensa, empañar la fuerza. En este caso el componente de inteligencia puede ser completamente autónomo (si las exigencias de tiempo lo requieren), sin embargo, si el módulo de empañamiento de la fuerza está restringido a FULL control (por tratarse de un objetivo anti-personal), la valoración del sistema completo será también de FULL control.

ANÁLISIS DE LA MATRIZ 1:

En términos generales, para las tareas de Exploración, Reconocimiento y localización, los AWS podrían ser sistemas OFF the loop o totalmente autónomos ya que no implican uso de la fuerza. Sin embargo, para las otras funciones, se debería requerir una IHM más cercana y un tipo de control humano más significativo. Es importante notar que esta primera matriz no impone el control total (FULL) bajo ningun-

na circunstancia. Sin embargo, la evaluación completa de un sistema requiere aplicar las consideraciones de las otras matrices donde se determinará el nivel de restricción por la valoración más restrictiva conjunta de todas las matrices.

Análisis de situación (escenario) en base a M1

Analizaremos ahora un escenario propuesto en una serie de ejercicios de mesa (TTX) (Parsi, Paoli, Spazian y Anand, 2021) realizado por el Instituto de las Naciones Unidas de Investigación sobre el Desarme (UNIDIR), para estudiar la interacción entre la introducción de grados de autonomía en los AWS y la retención del control humano.

Escenario: Esta misión se desarrolla en medio de un conflicto armado internacional entre el estado A y el estado B. El dominio de la misma es terrestre.

Objetivo: Destruir un lanzador de misiles no tripulado (que ha sido usado previamente por el estado B para derribar aviones BFOR Mil). El objetivo es fijo.

Ubicación del objetivo: pequeña isla en un lago. El acceso a la isla es por medio de un puente. El objetivo está protegido por un alambrado electrificado en el puente y alrededor del mismo. Es observado 24/7 por medio de sensores. Cualquier intento de cruzar el alambrado dispara una alarma. El personal de mantenimiento se encuentra a 1 km en un búnker.

Última Intel del objetivo: Por medio de imágenes de un drone y análisis realizado por medio de técnicas de IA se detectó lo siguiente:

- No hay civiles o personal militar en el objetivo.

- Presencia de civiles en las cercanías buscando agua del lago para regar la huerta.

- Lanzador de misiles armado y listo para disparar.

Estimación de daños colaterales (CDE): Mk82 (bomba de propósito general) – CDE 4 (no hay riesgo para civiles o bienes personales).

Restricción Legal: No hay restricciones legales para perseguir el objetivo

Restricciones de ataque: Atacar solo de noche – Escaneo transitorio antes de atacar.

Supongamos que se cuenta con AWSs capaces de realizar cada una de las tareas de la misión táctica. Aplicaremos la Matriz M1 a este escenario para determinar el grado de IHM requerido en cada caso para este escenario particular.

Resultados en base a la Matriz 1:

- En el caso de las tareas de exploración, reconocimiento, localización y vigilancia: nuestra matriz nos dice que esta tarea puede ser llevada a cabo con un dispositivo autónomo sin supervisión o intervención de operadores **(OFF)**.

- En el caso de las tareas de Identificación y seguimiento: la matriz M1 nos dice que de utilizarse un AWS para estas tareas el mismo debe responder a un esquema de IHM con intervención de un operador humano para la validación, en este caso de la identificación positiva correcta **(IN)**.

- El grado de IHM necesario para las tareas de análisis y evaluación del riesgo, empañamiento de los elementos de combate y el proceso de valoración de efectos no puede ser determinado sólo en función de la matriz M1. Volveremos a este mismo escenario para ejemplificar la utilización del resto de las matrices y conseguir una valoración completa del IHM requerido.

MATRIZ N° 2

Desarrollo Conceptual de la Matriz 2:

La Matriz 2 intenta abordar la siguiente pregunta **¿De qué manera elementos del contexto- ambiente operacional- influyen en la calidad y el alcance de la IHM?** Por ello, la Matriz 2 permite a modo general, desde el marco del **Dominio Operacional de Empleo de las AWS**, interrelacionar las exigencias del **ambiente operacional con los niveles de aplicación de la fuerza** (capacidad no letal / letal) de las AWS. Las celdas indican el tipo de interacción humano/dispositivos (IHM).

Desde una aproximación clásica, **los dominios operacionales** eran 3: tierra, mar y aire. Ante la complejidad de los escenarios y el avance de la tecnología se incorporan dominios emergentes como el espacio exterior y el ciberespacio. Por ello, hoy se conciben las operaciones multidominio, en el que la resolución de problemas militares no es de manera lineal y sino a partir de la integración de dominios centradas en efectos, situación que dará lugar a la aplicación de la matriz 3 para cada unas de las distintas dimensiones, correspondiendo aplicar la exigencia a las AWS más restrictivas en marco del DIH.

Por otro lado, como **ambiente operacional** se entiende al conjunto de factores de diversa naturaleza que existen en forma estable y semi-estable en una determinada región. Ellos influirán en la determinación de la composición, magnitud, equipamiento, y aptitud de las fuerzas que en él deban intervenir, como así también en la aplicación de su poder de combate (Ejercito argentino 1999).

En este documento desarrollaremos la Matriz 2 para el caso particular del Dominio Opera-

cional (Matriz 2.1) y el dominio Aéreo espacial (Matriz 2.2). Esta metodología de análisis deberá ser aplicada para cada ambiente operacional donde el AWS pueda participar.

Matriz 2.1:

Dominio Operacional (en general) - Exigencias de las distintas variables del ambiente operacional/aplicación de la fuerza no letal/letal. Individualiza el grado de control humano mínimo aceptable para la aplicación de la fuerza en este contexto.

EXIGENCIAS IHM EN EL AMBIENTE OPERACIONAL

NIVEL DE APLICACIÓN DE LA FUERZA

| Características del Dominio | | NO LETAL | LETAL |
|--------------------------------------------------------------------------------------------------------------------------------|--------------|----------|-----------------|
| Densidad Poblacional | Urbano | IN | FULL |
| | No Urbano | OFF | ON |
| | Inhabilitado | OFF | ON |
| Presencia de bienes protegidos (culturales o naturales) o infraestructura crítica (art. 56 protocolo 1) | SI | IN | FULL |
| | NO | OFF | ON |
| Exigencias de espacios comprometidos: confinado, inaccesible/contaminado/sin aire, condiciones ambientales adversas o críticas | SI | OFF | ON (Ver nota 1) |
| | NO | ON | IN |

| | | | |
|------------------------------------------------------------|-------------------------|-------------|------------------------|
| Tipo de armamento presente en el escenario | Convencional | No incide | No incide |
| | ADM (QBNRe) | FULL | FULL |
| | Híbrido | FULL | FULL |
| | AWS no propias | ON | ON |
| Características de la lucha en relación al empañamiento | Empañamiento simétrico | OFF | ON |
| | Empañamiento asimétrico | ON | ON |
| | Condición irregular | IN | IN |
| Exigencias Temporales asociada al ciclo de acción-reacción | SI | OFF | ON (Ver nota 1) |
| | NO | NO Incide | ON |

NOTA 1: La lógica de la tabla contempla el uso ofensivo de los sistemas de armas autónomas. Así mismo, la combinación de las variables que conforman el ambiente operacional puede exigir la genuina implementación de la opción de la legítima defensa por lo tanto permitiría una flexibilización en los tipos de control, a efectos de satisfacer -en tiempo y forma- a las exigencias, impactos o riesgos. Este tratamiento requiere un estudio posterior que complemente y profundice el presente documento.

Observemos que la condición “FULL” debería asegurar -en forma permanente- la capacidad de poder asumir el control -por diseño y características del Sistema de C2- del módulo de aplicación de la fuerza -en general- en el marco del manejo de la crisis-conflicto (el hombre como actor -objetivo y subjetivo- de su creación).

Análisis de situación en base a M2.1 (Dominio Operacional - “Terrestre”)

Tomando como ejemplo nuevamente el escenario 1, el análisis que proponemos es el siguiente. Primero caracterizamos el escenario en términos de las variables de la exigencia del ámbito operacional.

1. ensidad poblacional: NO URBANO.
2. Presencia de bienes protegidos (culturales o naturales) o infraestructura crítica (art. 56 protocolo 1): NO.
3. Exigencias de espacios comprometidos: NO.
4. Tipo de armamento presente en el escenario: CONVENCIONAL.
5. Características de la lucha en relación al empañamiento: SIMÉTRICO.
6. Exigencias Temporales asociadas al

ciclo de acción-reacción: NO.

Además, la descripción del escenario implica una acción LETAL de acuerdo al explosivo que se pretende utilizar. Bajo esa caracterización del ambiente operacional, la aplicación de M2.1 nos **indicaría que el mínimo control humano aceptable en este caso es IN** (el mínimo es tomado entre las indicaciones que M2.1 ofrece para cada una de las características de 1 a 6, en este caso 3). Vale la pena remarcar que si la condición fuera no letal, **el resultado final sería ON**.

Matriz 2.2: Dominio Aeroespacial (ejemplo de aplicación para un Dominio Operacional en particular) - exigencias de las distintas variables del ambiente operacional versus aplicación de la fuerza -no letal / letal. Individualiza

el grado de control humano mínimo aceptable para la aplicación de la fuerza en este contexto.

Para desarrollar esta Matriz se han tenido en cuenta las exigencias de las tareas asociadas a las Operaciones Aeroespaciales (factor crítico “el tiempo”), y su incidencia -a la par de la naturaleza y exigencias del ambiente operacional- en el manejo de la crisis/conflicto en el marco de las ROE⁶ establecidas.

En ese sentido, resulta pertinente contemplar los riesgos de escalada por una falla AWS y, por ello, la posibilidad del operador para asumir el control -capacidad reactiva- conforme a la velocidad los eventos tácticos y la complejidad de la etapa en desarrollo con el objeto de anular y/o minimizar el impacto de la eventual contingencia.

EXIGENCIAS IHM EN EL DOMINIO AEROESPACIAL

NIVEL DE APLICACIÓN DE LA FUERZA

| Características del Dominio | | No LETAL | LETAL |
|---------------------------------------------------------------------------------------------------------|--------------|----------|-------|
| Densidad Poblacional | Urbano | IN | FULL |
| | No Urbano | OFF | ON |
| | Inhabilitado | OFF | ON |
| Presencia de bienes protegidos (culturales o naturales) o infraestructura crítica (art. 56 protocolo 1) | SI | IN | FULL |
| | NO | OFF | ON |

6 ROE: Reglas de Empañamiento.

| | | | |
|----------------------------------------------------------------------------------------------------------------------------------|-------------------------|-----------|-----------------|
| Exigencias de espacios comprometidos: confinado, inaccesible/contaminado/sin aire, condiciones ambientales adversas o críticas | SI | OFF | ON |
| | NO | IN | FULL |
| Tipo de armamento presente en el escenario | Convencional | No incide | No incide |
| | QBNRe | FULL | FULL |
| | Híbrido | FULL | FULL |
| | AWS no propias | ON | ON |
| Características de la lucha en relación al empañamiento | Empañamiento simétrico | OFF | IN |
| | Empañamiento asimétrico | ON | IN |
| | Condición irregular | IN | FULL |
| Exigencias Temporales por el ciclo de reacción del tiempo y escasa permanencia en las posiciones relativas favorables de interés | SI | OFF | ON (Ver nota 1) |
| | NO | ON | IN |

En el marco del tipo de control y restricciones en el uso de los sistemas debe analizarse de manera integral y amplia en relación a su impacto tanto sobre la vida como de aquellos bienes

protegidos por el DIH (medio ambiente, atmósfera y especies protegidas), que se pueden verse afectados en circunstancia del desarrollo de operaciones militares y su evolución.

Nota 1: La lógica de la Matriz contempla el uso **ofensivo** de los sistemas de armas autónomas. Así mismo, la combinación de las variables que conforman el ambiente operacional puede exigir la genuina **implementación de la opción de la legítima defensa** por lo tanto permitiría una flexibilización en los tipos de control, a efectos de satisfacer -en tiempo y forma- a las exigencias, impactos o riesgos. Este tratamiento requiere un estudio posterior que complemente y profundice el presente documento.

MATRIZ N° 3

Interrelaciona los **Objetivos** (UNIDIR, 2019) (tipos de blanco y condición en función a la situación estratégica) y la **capacidad No letal / Letal de las AWS**. Las celdas indican el nivel de control humano/autonomía mínima aceptable en el sistema. Interacción entre el humano y el dispositivo en relación a los objetivos y su capacidad letal.

La Matriz 3 intenta abordar la siguiente pregunta **¿Qué calidad y alcance debería tener el IHM en relación a los niveles de aplicación de la fuerza, las características de los objetivos y diferentes etapas del manejo del conflicto para mantenerse en el marco del DIH y ROEs establecidas?**

| MANEJO DEL CONFLICTO | APLICACIÓN DE LA FUERZA | | | | |
|----------------------------|---------------------------------|--------------|---------------|--------------|---------------|
| | CARÁCTER | No LETAL | | LETAL | |
| Objetivo | Objetivo Militar Válido | Anti sistema | Anti personal | Anti sistema | Anti personal |
| Tipo | Combatiente (obj. mili. válido) | OFF | ON | ON | IN |
| | No Combatiente | ON | FULL | FULL | FULL |
| Determinación del Blanco | Proceso deliberado | ON | IN | ON | FULL |
| | Acciones en Desarrollo | ON | IN | IN | FULL |
| | Manejo de Crisis | IN | FULL | IN | FULL |

Con respecto al proceso de determinación del blanco, el proceso **DELIBERADO** es una actividad preliminar detallada de análisis y determinación de blancos y riesgos -certidumbre y calidad de información-, durante el **DESARROLLO** de las acciones -condiciones dinámicas y cambiantes- y en oportunidad del manejo de la **CRISIS** -incertidumbre en relación a los intereses y evolución de las estrategias de los actores- que requieren de una rápida respuesta.

CLASIFICACIONES UTILIZADAS PARA LA MATRIZ 3 :

La clasificación en relación al Objetivo Material corresponde al concepto de:

- Antipersonal: acción sobre el combatiente -fundamentalmente en el empañamiento táctico- en el que las condiciones y los métodos empleados en la lucha reducen sensiblemente e imposibilitan toda acción y/o expectativa de supervivencia, sobretodo en situaciones de autodefensa, ante la efectividad y/o eventual discrecionalidad de los AWS.

- Anti sistemas: acción sobre infraestructura y servicios críticos asociado a las operaciones militares y/o esfuerzo de guerra; como así también a los Sistema de Armas y/o Armas (tripulados o no), sistema de comando y control, de defensa en los dominios operacionales, buques, submarinos, medios aéreos, vehículos terrestres, armamento dirigido, munición/obuses, entre otros), orientados a su destrucción modular y/o neutralización sistémica con elevadas expectativas de ventaja militar en función de la efectividad de los AWS.

Las consideraciones en relación a las condiciones para la determinación del blanco hacen referencia a la oportunidad y las circunstancias

(en la variable tiempo oportunidad) que permiten analizar y determinar en detalle las condiciones de empleo y riesgos operacional inherentes asociados a las AWS.

Ello permite, en el marco de las ROEs establecidas, adoptar las acciones de contención/contingencia permiten gestionar dichos **riesgos hacia un riesgo operacional residual aceptable en el marco del/los:**

- Bienes tutelados/protegidos por el DIH.
- Manejo de la crisis/conflicto.
- Niveles y exigencias HMI requeridas en toda la cadena de valor.

La clasificación en relación a la Aplicación de la Fuerza:

- No Letal: métodos y medios de la guerra -en el marco del manejo de la crisis o desarrollo del conflicto- cuya aplicación buscan disuadir, influir, afectar, reducir y/o anular el proceso de decisión y/o efectividad/condición -de manera transitoria y/o parcial- de componentes de los sistemas de armas -sensores de los AWS entre otros- y/o ejecución de operaciones militares.

Si bien una evaluación preliminar desestima impactos de 1er Nivel sobre la vida, medio ambiente y/o biodiversidad, resulta oportuno evaluar efectos de 2 y 3 nivel en relación a los riesgos/impacto en el empleo de las AWS sobre los elementos protegidos precedentes.

- Letal: métodos y medios de la guerra -en el marco del manejo de la crisis o desarrollo del conflicto- cuyo aplicación producen un daño -total/parcial-, neutralización o alteración significativa en el proceso de decisión y/o efectividad/condición -de manera permanente y/o sistémica- de componentes de los sistemas de armas -combatientes entre otros- y/o ejecución de operaciones militares;

como así también, afectación significativa y/o permanente del ecosistema, medios ambiente y/o atmósfera en oportunidad de aplicación de la fuerza.

El grado de control sobre los AWS eventualmente puede no estar necesariamente vinculada a la letalidad de sus armas sino al impacto/ riesgos políticos, estratégicos y/o tácticos de su empleo.

Análisis de situación en base a M3

Tomando como ejemplo nuevamente el Escenario 1, el análisis que proponemos es el siguiente. Primero describiremos el escenario en términos de las variables que caracterizan el objetivo.

1. **Objetivo Militar Válido:** ANTISIS-TEMA.
2. **Tipo de Personal Protegido:**

COMBATIENTE (Objetivo militar válido).

3. Ciclo de Determinación del Blanco: DELIBERADO.

La descripción del escenario implica una acción LETAL de acuerdo al explosivo que se pretende utilizar. Bajo esa caracterización del objetivo, la aplicación de M3 nos indicaría que el **mínimo control humano aceptable en este caso es ON**. Finalmente, el hecho de que se trate de un blanco fijo deliberado y con poca incertidumbre (3), la **condición 3 habilita un grado de control ON**.

Como reflexión final, **analizando secuencialmente las matrices ofrecidas en el presente trabajo se identifica que la IHM mínima aceptable para el escenario 1 es IN**, es decir con autonomía limitada que requiere intervención del sistema de comando y control para validar o implementar acciones específicas.

MATRIZ N° 4

DESARROLLO CONCEPTUAL DE LA MATRIZ 4.

La Matriz 4 intenta abordar la siguiente pregunta: **¿Qué principios rectores de la IA tiene que tener un AWS de acuerdo a las diferentes etapas del proceso de la aplicación de la fuerza para garantizar un adecuado control humano significativo?** Los principios rectores de la IA que deben cumplir las AWS son: **Previsibilidad, Comprensión, Justificación/Accountability y Explicación**. Para ello, la **Matriz 4** interrelaciona las **diferentes etapas del proceso de la aplicación de la fuerza (funciones**

y tareas donde intervendrían los AWS) y los principios de la IA mencionados. En cada celda se podrá identificar el grado de exigencia requerido para cada uno de los principios críticos de la IA en función a las tareas/ funciones: ALTO-MEDIO-BAJO. Dicha metodología permite definir los niveles de exigencias requeridas en cada etapa del proceso de acción/tareas, en relación a los principios de la IA. A fin práctico, la matriz 4 se realiza para el nivel de decisión táctico, debiéndose realizar la misma metodología para los otros 2 niveles.

A continuación, recordamos brevemente los principios de la IA (Martínez- Rodríguez 2020):

1. Previsibilidad (Holland, M., 2020):

Es la medida en que se pueden anticipar los resultados o efectos de un sistema. Responde la pregunta ¿Qué hará este sistema? En el sentido técnico, refiere a la capacidad para ejecutar una tarea con el mismo rendimiento que exhibió en las pruebas. Es una función de la frecuencia con la que las salidas son correctas (precisión). También se entiende como la medida de su rendimiento a lo largo del tiempo (replicabilidad, reproducibilidad) y la medida en que el sistema puede “generalizar” y adaptarse. En el sentido operacional, da cuenta de la capacidad de adaptación a situaciones dinámicas y altamente complejas. En el sentido de los efectos, se refiere al grado en que se puede anticipar el impacto.

2. Comprensión: Se basa en la interpretabilidad intrínseca del AWS y refiere al grado en que una persona puede comprender el comportamiento del sistema. Responde a la pregunta ¿Por qué el sistema hace lo que hace? En el espectro de los sistemas basados en IA existe una amplia variación con respecto a cuán entendibles son sus fundamentos.

Analizado desde un sentido estricto, esta propiedad parece requerir que un usuario humano (desarrollador u operario) pueda entender las complejidades del modelo de IA. En este sentido, un sistema de IA basado en reglas⁷, o un árbol de decisión⁸, relativamente pequeño, puede ser altamente comprensible para su diseñador, aunque un usuario lego puede requerir cierto nivel de entrenamiento para poder comprender la “lógica” de la estructura de conocimiento y del proceso de inferencia. Por lo contrario, comprender el proceso de aprendizaje en una red neuronal (profunda) debido a la asignación de millones de pesos a características (“features”) que la misma red idéntica de los datos de entrada (por ejemplo, los píxeles de una foto o un video) suele ser difícil. En relación al principio de previsibilidad, la comprensión es complementaria y ambas son necesarias cuanto más compleja es la tarea que el sistema debe resolver y el entorno donde se desenvuelve la acción.

3. Justificación (Rendición de cuentas - Accountability): Los sistemas deben ser auditables: tanto en lo que se refiere a código como a datos. Las decisiones de diseño deben ser explícitas: todos los actores que participan del diseño y desarrollo (en realidad de todo el ciclo de vida del sistema) deben rendir cuen-

7 Un sistema experto está diseñado para trabajar como si fuera una persona experta en algún tema específico. Trata de replicar o de simular el conocimiento que han adquirido estas personas basándose en las herramientas que ofrece la inteligencia artificial. Su función principal es tratar de ofrecer soluciones que puedan resolver problemas prácticos o también la creación de conocimiento nuevo.

<https://sites.google.com/site/sistemasexpertosunah/home/sistemas-expertos-basados-en-reglas>

8 Modelo de predicción utilizado en diversos ámbitos que van desde la inteligencia artificial hasta la economía. Dado un conjunto de datos se fabrican diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, para la resolución de un problema.

ta de suposiciones y decisiones tomadas sobre estos procesos. Deficiencias e impactos negativos de los sistemas deben ser estudiados y reportados con anticipación. En cuanto a su etapa de operación, los sistemas deben ser capaces de justificar si su participación facilitará o dificultará la ejecución de los objetivos de la operación y si podrá hacerlo de manera factible de conformidad con el DIH. Finalmente, después del uso de un AWS, se necesitará contar con la capacidad de determinar exactamente qué sucedió y por qué. Esto incluye determinar por qué el AWS se comportó como lo hizo, y por qué ciertos efectos resultaron de su empleo. La trazabilidad de la operación jugará un rol importante en la tarea de justificar lo acontecido. En particular para determinar si es probable que vuelvan a surgir los efectos adversos del uso de un AWS. (Martínez- Rodríguez, 2020)

4. Explicación: Capacidad del sistema de IA de poder ofrecer/comunicar a su usuario explicaciones para su razonamiento o accionar en términos adecuados. Está relacionado con el interfaz humano/computadora. Este concepto es uno de los más desarrollados en relación a sistemas de IA, de hecho, la XIA (IA explicable) es un área de investigación en expansión dentro del campo de la IA. La explicabilidad es la capacidad del sistema basado en IA de poder justificar sus resultados o decisiones en términos que un humano pueda entender. Para entender el concepto es necesario diferenciarlo de la comprensión que no implica explicabilidad y viceversa. Un sistema de caja negra podría, en principio, dar explicaciones sobre sus resultados, pero esas explicaciones serían igual de opacas e incomprensibles que el resto del modelo. Aun así, podría servir para generar

confianza en el usuario si las mismas apelan a una correlación asequible por el mismo. Por otro lado, la mayoría de los sistemas basados en IA que hoy utilizamos, aún aquellos que son completamente interpretables no están diseñados para ofrecer explicaciones que acompañan sus resultados o sugerencias. La explicabilidad requiere una habilidad específica del sistema, es decir, es el sistema el que actúa para dar explicaciones, mientras que la comprensión deja al sistema en modo pasivo y es el humano quien lo analiza.

Es importante destacar que otros términos se han usados a veces en referencia a los conceptos de comprensión, como **transparencia, interpretabilidad e inteligibilidad**. Para este caso y para este tipo de análisis, se entiende que la transparencia queda cubierta bajo las consideraciones de los principios de Comprensión, Justificación y Explicación. En cuanto el principio de transparencia se podría encuadrar en otro nivel de análisis más abstracto y de manera transversal el cual incluye los otros principios, **por lo cual se decidió utilizar sólo 4 principios más concretos de las IA para la realización de la matriz conceptual**.

La matriz 4 puede presentar distintos tipos de valoración: ALTA, MEDIA o BAJA en las etapas del proceso, según el nivel de decisión que se analice, entendiendo que la previsibilidad, comprensión, justificación y explicación son cuatro principios críticos de la IA para establecer un mínimo control humano significativo para todos los niveles de decisión.

A efectos de comprensión, a continuación, ejemplificamos en el siguiente cuadro, el análisis sobre el nivel operacional.

MATRIZ N° 4: EXIGENCIAS PARTICULARES PARA EL NIVEL OPERACIONAL

| ETAPAS DEL PROCESO DE APLICACIÓN DE LA FUERZA: FUNCIONES/ TAREAS | PRINCIPIOS CRÍTICOS DE LA IA | | | |
|------------------------------------------------------------------------|------------------------------|-------------|--------------------------------|-------------|
| | PREVISIBILIDAD | COMPRENSIÓN | JUSTIFICACIÓN (ACCOUNTABILITY) | EXPLICACIÓN |
| Exploración y Reconocimiento | ALTA | BAJA | BAJA | ALTA |
| Localización | ALTA | BAJA | BAJA | BAJA |
| Identificación | ALTA | MEDIA | ALTA | ALTA |
| Vigilancia | ALTA | BAJA | BAJA | MEDIA |
| Seguimiento | ALTA | BAJA | BAJA | ALTA |
| Análisis y Evaluación del Riesgo | ALTA | ALTA | ALTA | ALTA |
| Empañamiento de Elementos del Combate | ALTA | ALTA | ALTA | BAJA |
| Proceso de Valoración de Efectos | ALTA | ALTA | ALTA | ALTA |

Notas:

- Para completar esta tabla se ha tenido en cuenta aspectos como lo crítico que la etapa es, si hay riesgo de escalada por una falla, y en ese sentido si el operador tendría tiempo para

reaccionar, y la posible velocidad de cambio y complejidad de la etapa.

- Esta metodología de análisis (M4) deberá ser aplicada para cada nivel de decisión (Político/estratégico, operacional y Táctico).

Análisis de M4:

- En cuanto a la exigencia de **Previsibilidad: es ALTA** en todas las etapas del proceso, llegando a la conclusión de que es una exigencia indispensable para todos los aspectos a ser considerados. Este requerimiento es mínimo para que el operador humano desarrolle su tarea con el suficiente **grado de confiabilidad** y aceptables **garantías de fiabilidad**.

- La etapa de **Identificación** es considerada a los efectos de este análisis como de inflexión. En ella se recaba y analiza toda la información que servirá en las etapas posteriores.

- En cuanto a las etapas de **Análisis y Evaluación del Riesgo, Empañamiento de los elementos de combate y procesos de valoración de efectos**, las exigencias de **previsibilidad, comprensión y justificación del sistema de IA son ALTAS** por considerarlas etapas críticas por la vertigocidad de su desarrollo. La explicabilidad de estas etapas también son ALTAS, salvo en relación al empañamiento de elementos del combate donde la explicabilidad no ofrece mayores ventajas dados los tiempos que se manejan.

Análisis de situación en base a M4

Volvamos ahora al escenario de la serie de ejercicios de mesa (TTX) ya analizado. Tengamos en cuenta que, así como las anteriores matrices hacían énfasis en establecer requerimientos sobre el HMI aquí se enfocó en fijar condiciones mínimas que debe satisfacer el AWS para garantizar que dicha interacción es confiable.

Resultados en base la Matriz 4:

- En el caso de las dos tareas de exploración y reconocimiento, y localización: nuestra matriz no requiere una alta comprensión del modelo subyacente al AWS ni que cuente con un alto grado de justificación para que el ope-

rador humano realice la respectiva tarea con un control significativo. Si se requiere que el sistema tenga una alta capacidad de explicabilidad para la tarea de exploración y reconocimiento.

- En el caso de la tarea de Identificación la matriz M4 nos dice que de utilizarse un AWS para estas tareas el mismo debe tener un comportamiento altamente predecible, justificable y explicable. Esto también requerirá la intervención de un operador humano altamente calificado para la validación de la identificación. En este caso es lo que realiza el servicio de inteligencia al analizar las imágenes del dron y los reconocimientos realizados por la IA.

- En relación a las tareas de vigilancia y seguimiento el cuadro requiere utilizar un sistema que sea altamente predecible y explicable. Según la descripción del escenario estas tareas no se desarrollan.

- Para las tareas de análisis y evaluación del riesgo, esta matriz establece el máximo de exigencia. Habrá que evaluar si el escaneo previo al ataque alcanza para alcanzarlos.

- Respecto al empañamiento de los elementos de combate y el proceso de valoración de efectos no puede ser determinado con la matriz M4 debido a la falta de información en la descripción.

CAPITULO 4

DEFINICIONES CONCEPTUALES FINALES.

La gran mayoría de los sistemas de IA implementados en contextos del mundo real son sistemas compuestos. Están conformados por una variedad de subsistemas/módulos de IA interconectados derivados de diferentes fuentes, que trabajan en forma integrada para un propósito. Aún sistemas de armas simples requieren la interconexión de varios submodelos de IA que puedan realizar tareas de reconocimiento, predicción, etc. Esto complejiza aún más el análisis integral de los mismos.

Asimismo, y como advertimos en el desarrollo del presente trabajo, una de las características esenciales de la IA es que se desarrolla bajo una dinámica de *Blanco móvil* de evolución continua, por lo tanto, algunos de los estándares que se proponen en este momento podrían quedar obsoletos frente a desafíos y riesgos futuros.

Es por ello que, a fines conceptuales, es imperioso salir de la dinámica del Blanco móvil y establecer un análisis prospectivo para definir un producto maduro en todas sus etapas. Por lo tanto, resultaría apropiado complementar al esquema reactivo-preventivo, un abordaje preventivo-anticipativo como corazón (Core) de una estrategia que le de una cobertura estructural conforme a las diversas consecuencias po-

líticas, éticas, legales, humanitarias y ambientales que esta tecnología disruptiva implicaría en las condiciones y formas de vida de la humanidad, en particular en lo que respecta al arte de la Guerra.

El control significativo del hombre sobre la IA, desde un punto de vista estratégico implica que: si la tecnología asiste y complementa la acción del hombre, el hombre mantiene su condición de actor; si reemplaza las funciones críticas del hombre, el hombre pasa a ser un espectador de su propia creación y sus efectos.

En el abordaje preventivo-anticipativo entendemos que debe estar el CORE del debate sobre la base de la conceptualización de un producto maduro y sus amenazas, riesgos, desafíos y oportunidades asociados los cuales le otorgarían racionalidad y previsibilidad -en tiempo y forma- a todo proceso de desarrollo de esta tecnología. Hasta tanto se defina y se arribe a un consenso en relación con las características de un producto maduro de AWS, es necesario establecer acciones de contención y limitación en aquellos aspectos críticos del proceso.

En relación a lo expuesto precedentemente, la imprevisibilidad constituye una de los grandes

desafíos sobre de las AWS, tal cual lo expresa el documento del CICR

“Un grado de imprevisibilidad es inherente a los efectos del uso de todos los sistemas de alerta temprana debido al hecho de que el usuario no elige, ni conoce, el objetivo o los objetivos específicos, ni el momento y/o la ubicación precisos de la aplicación o las aplicaciones de fuerza resultantes. Esto conlleva riesgos de daños para los afectados por el conflicto ar-

mado, serios desafíos en la aplicación del DIH y peligros de escalada del conflicto” (ICRC, 2021).

Estos desafíos, implican que **la inacción frente a AWS presenta una lógica exponencial en cuanto a sus consecuencias**, más aún si no contamos con una adecuada capacidad de respuesta oportuna y eficaz (Reglas de Juego) en términos de limitaciones, restricciones o prohibiciones desde un marco legal internacional.

$$\text{Riesgo} = \frac{\text{Amenaza (AWS) x vulnerabilidad (incertidumbre y discrecionalidad)}}{\text{Capacidad de Respuesta (Reglas de Juego)}}$$

Es imperioso establecer internacionalmente las reglas de juego para un regulado desarrollo y uso de las AWS a fin de obtener un **triple impacto a nivel global**:

1. Reducción de riesgos, incertidumbre y discrecionalidad.

2. Institucionalización de un encuadre legal vinculante en el marco de los principios del DIH y los DDHH.

3. Democratización en el funcionamiento del sistema internacional.

Este debate resulta una gran oportunidad como humanidad a fin de establecer preventivamente las amenazas de un futuro incierto; debate que en el siglo pasado no se dio frente a la evolución, diseño, uso y efectos devastadores del armamento nuclear de forma anticipativa.

En el marco del DIH y en consonancia con los Derechos Humanos y la ética, los tres principios críticos que harían imposible de utilizar

las AWS serían los principios de Humanidad, Responsabilidad y Limitación, en aquellos sistemas letales antipersonales y armamento anti sistemas que pongan en riesgo bienes civiles en un sentido amplio (como se ve en las matrices 1, 2.1, 2.2 y 3).

Para aquellas AWS que sean anti-sistema y/o que cumplan otras tareas (como se ve a través de las matrices), mientras tanto un AWS no pueda demostrar teórica o empíricamente ser capaz de distinguir efectivamente entre combatientes y no combatientes y que, en su accionar, permita aplicar los principios de Distinción, Proporcionalidad y Precaución, su utilización debería estar preventivamente prohibida hasta que no haya garantías sobre fiabilidad técnica-operativa del sistema.

Para esta segunda situación es imperativo que se disponga de todas las revisiones legales, consuetudinarias y técnicas expresadas bajo el Art 36 del Protocolo Adicional de 1977, en especial las prohibiciones o restricciones sobre introducción de nuevas armas bajo tratados internacionales, prohibiciones

o restricciones bajo Derecho Internacional Consuetudinario, prohibiciones o restricciones basadas en el principio de humanidad (Cláusula de Martens).

Una posible estrategia sería que bajo el artículo 36, donde se establece el mecanismo para la Revisión de las autoridades a fin de incorporar nuevas armas, se incluya para esta tecnología específica un “5 elemento” que contemple como exigencia la fiabilidad técnica y operativa (trustworthiness) (teórica y/o empírica), lo que implica demostrar justificadamente que el/los sistemas son previsibles, explicables, con existencia demostrable de control humano significativo y asignación de responsabilidades individuales en todo su proceso de diseño, elaboración y empleo.

Se entiende que este “5 elemento” implicaría el desarrollo de procesos robustos de certificación de estándares en toda la cadena de desarrollo, empleo y disposición final del AWS. En este trabajo presentamos -de manera preliminar- una metodología de conceptualización de las AWS que serviría como punto inicial para la evaluación de la fiabilidad técnica y operati-

va de estos sistemas.

Este documento abordó dichos interrogantes desde una propuesta metodológica, integrando los principios críticos del DIH y de la IA, ofreciendo estrategias para considerar su regulación y/o prohibición.

Finalmente entendemos que, sobre la base de los trascendentes riesgos, amenazas e impactos que engendran la IA -en general- y las AWS -en particular- en el futuro de la humanidad (dimensión ética), en la gestión del conflicto (dimensión político-estratégica) y en el empañamiento (dimensión operacional-táctica), ni el más avanzado desarrollo de la tecnología podrá reemplazar aquellos aspectos de la naturaleza hombre que le permite evitar, superar y/o interrumpir la cadena causal de una dinámica del conflicto/violencia.

Por ello, el principio de humanidad constituirá el criterio básico universal de aplicación a la concepción, diseño, desarrollo, certificación y empleo de las AWS, asegurando al hombre -en forma permanente- ser actor de su realidad, de su creación y de su futuro.

FUENTES Y BIBLIOGRAFÍA:

Amoroso, D. and Tamburrini, G., “Toward a Normative Model of Meaningful Human Control over Weapons Systems” Disponible al 22/09/2021 en <File:///C:/Users/administrador/Downloads/toward-a-normative-model-of-meaningful-human-control-over-weapons-systems.pdf>

CCW (2019), “Guiding Principles Affirmed by the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems” Annex Disponible el 21/09/2021 en https://www.ccdcoe.org/uploads/2020/02/UN-191213_CCW-MSP-Final-report-Annex-III_Guiding-Principles-affirmed-by-GGE.pdf

Comisión Europea (2019), “Directrices para una IA fiable”. Disponible al 22/09/2021 en bit.ly/2RYbh8D

Coupland, R. (2001), “Humanity: What is it and how does it influence international law?”, ICRC, Disponible al 21/09/2021, <https://www.icrc.org/en/doc/assets/files/other/irrc-844-coupland.pdf>

Davidson, Neil (2017) “A legal perspective AWS under IHL” UNODA OCCASIONAL PAPERS N°30

Díaz, M. y Muñoz, W. (2020) “Los riesgos de las armas autónomas letales: una perspectiva interseccional latinoamericana”. Red de Seguridad Humana para América Latina y el Caribe. Disponible al 22/09/2021 en <https://img1.wsimg.com/blobby/go/98c6dc90-096f-4389-9309-f1a33c0cad73/downloads/Los%20riesgos%20de%20las%20armas%20aut%C3%B3nomas%20una%20perspec.pdf?ver=1623789686395>

Dignum, Virginia (2019) “Responsible Artificial Intelligence - How to Develop and Use AI in a Responsible Way” in “Artificial Intelligence: Foundations, Theory, and Algorithms”, ISBN 978-3-030-30370-9, pp. 1-120.

Ejército argentino (1999) “Reglamento de conducción para el instrumento militar terrestre” Documento público militar.

Entrevistas en profundidad al Brigadier (R) Armando Bonadío, director del Instituto de Derecho Aeronáutico y Espacial, director y profesor titular del Curso de DIH del INDAE (2021)

G. Persi Paoli, A. Spazian, A. Anand, (2021) “Table-Top Exercises on the Human Element and Autonomous Weapons Systems: Summary Report”, Geneva, Switzerland: UNIDIR.

Gariglio, Damián (2020) “Los Sistemas de Armas Autónomos Letales y el Desbalance de Poder”. Red de Seguridad Humana para América Latina y el Caribe y Centro de Estudios de Política Internacional de la UBA, 2020. Disponible al 22/09/2021 en https://18df0113-5ebe-4413-800c-obfd2e-8147bb.filesusr.com/ugd/7cccdc_b388e2ca492b4f7c841d6eff14a3a50f.pdf

- Holland Michel, Arthur, (2021). “Known Unknowns: Data Issues and Military Autonomous Systems”. Geneva, UNIDIR. Disponible al 22/09/2021 en <https://unidir.org/known-unknowns>
- Holland Michel, Arthur. (2020). ‘The Black Box, Unlocked: Predictability and Understandability in Military AI.’ Geneva, Switzerland: United Nations Institute for Disarmament Research. doi: 10.37559/SecTec/20/AI1
- ICRC (2021) “Position on Autonomous Weapons Systems”. Disponible el 21/09/2021 en https://www.icrc.org/en/download/file/166330/icrc_position_on_aws_and_background_paper.pdf
- ICRC (2021B) “Armas autónomas: el CICR recomienda adoptar nuevas normas” [Declaración]. Disponible al 22/09/2021 en <https://www.icrc.org/en/document/autonomous-weapons-icrc-recommends-new-rul>
- INTERNATIONAL COURT OF JUSTICE (1996) “Advisory Opinion on the Legality of the Threat or Use of Nuclear Weapons” Disponible al 21/09/2021 en <https://www.icj-cij.org/public/files/case-related/95/095-19960708-ADV-01-00-EN.pdf>
- Lawand Kathleen, Coupland Robin y Herby Peter (2006), “Guía para el examen jurídico de las armas, los medios y los métodos de guerra nuevos.” ICRC, Disponible al 21/09/2021 en https://www.icrc.org/es/doc/assets/files/other/icrc_003_0902.pdf
- López Díaz, P. (2009). “Principios fundamentales del Derecho Internacional Humanitario”. REVISMAR. pág. 230 a 238. Disponible al 22/09/2021 en <https://revistamarina.cl/revistas/2009/3/lopez.pdf>
- Martínez Vanina y Rodríguez Ricardo (2020) “Aportes al debate del uso de la IA para aplicaciones armamentísticas” en Red de Seguridad Humana para América Latina y el Caribe”. Disponible al 22/09/2021 en <https://img1.wsimg.com/blobby/go/98c6dc90-096f-4389-9309-f1a33c0cad73/downloads/Contributions%20to%20the%20debate%20on%20the%20use%20of%20Arti.pdf?ver=1630078276981>
- Matthias, Andreas (2004), “The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata” en Ethics and Information Technology
- Micha, L. y Farías, P. (2021) “La evolución de tecnologías disruptivas y los sistemas de armas autónomas letales: consideraciones desde el ámbito militar”. Red de Seguridad Humana para América Latina y el Caribe y Centro de Estudios de Política Internacional de la UBA, (CEPI) marzo, 2021. Disponible al 22/09/2021 en <https://img1.wsimg.com/blobby/go/98c6dc90-096f-4389-9309-f1a33c0cad73/downloads/La%20evoluci%C3%B3n%20de%20tecnolog%C3%ADas%20disruptivas%20Los%20SA.pdf?ver=1627396384155>
- Muggleton, S.H., Schmid, U., Zeller, C. (2018) “Ultra-Strong Machine Learning: comprehensibility of programs learned with ILP”. Mach Learn 107, 1119–1140 Disponible al 22/09/2021 en <https://doi.org/10.1007/s10994-018-5707-3>

Sparrow, Robert (2007) “Killer Robots” en Journal of Applied Philosophy, pág 62 a 77

UNIDIR (2019) “El elemento humano en las decisiones sobre el uso de la fuerza” Disponible al 22/09/2021 en <https://www.unidir.org/publication/el-elemento-humano-en-las-decisiones-sobre-el-uso-de-la-fuerzaG>

UNIDIR (2018). “Algorithmic Bias and the Weaponization of Increasingly Autonomous Technologies” Disponible al 22/09/2021 en <https://unidir.org/publication/algorithmic-bias-and-weaponization-increasingly-autonomous-technologies>

Virginia, Dignum (2019) “Responsible Artificial Intelligence - How to Develop and Use AI in a Responsible Way”. Artificial Intelligence: Foundations, Theory, and Algorithms, Springer, ISBN 978-3-030-30370-9, pp. 1-120.

Luciana Micha es Licenciada en Ciencia Política graduada con diploma de honor de la Universidad de Buenos Aires, Mg en Programación Neurolingüística y candidata a Doctora en Ciencia Política. Es actualmente directora del Centro de Estudios de Política Internacional (CEPI) de la Universidad de Buenos Aires. Docente titular de la UBA y en la Universidad de la Defensa Nacional (UNDEF) como directora y Docente de las Diplomaturas de Asistencia Humanitaria y docente titular en el Instituto Nacional de Derecho Aeronáutico y Espacial (INDAE) en la Diplomatura de Derecho Internacional de los Conflictos Armados.

Es funcionaria de carrera del Ministerio de Defensa de la Argentina desde el año 2001, ejerciendo los cargos de directora nacional de Cooperación para la Paz (2006-2010), Enlace con el Congreso Nacional (2011-2012) Coordinadora de Bienes Culturales (2012- 2015) y Coordinadora de Derecho Internacional Humanitario (2015 -2019). Actualmente presta servicios en la Universidad de la Defensa Nacional, Secretaría de Extensión Universitaria

Pablo Andres Farías es Comodoro de la Fuerza Aérea Argentina, en situación de retiro activo como Art 62. Aviador Militar de la Especialidad de Ataque, que prestó servicios en el país, Antártida Argentina y el exterior; desempeñando funciones en las áreas operativa, educación, docente, instrucción de vuelo, operaciones, inteligencia de combate, planeamiento estratégico institucional/militar, proyectos y relaciones internacionales (Embajada Argentina en Chile), entre otros. Oficial de Estado Mayor (Fuerzas Aéreas de Argentina y Perú), con educación superior (de grado, posgrado, especializaciones y cursos) en áreas de interés vinculadas con la Defensa Nacional, actual-

mente candidato al Doctor en Ciencias Políticas. Desarrolló de actividad de Estado Mayor, análisis, asesoramiento y decisión en Organismos Superiores de la Fuerza Aérea; Estado Mayor Conjunto y Ministerio de Defensa en temas del ámbito político estratégico; estratégico-militar, política de defensa, Directivas Estratégicas Militares, Capacidades Militares y elaboración del Libro Blanco de la Defensa Nacional, entre otros.

María Vanina Martínez es Doctora en Ciencias de la Computación (Universidad de Maryland, USA) con un posdoctorado en la Universidad de Oxford. Es coordinadora del Programa de Ciencia de Datos e Inteligencia Artificial en la Fundación Sadosky, investigadora del Instituto de Ciencias de la Computación (CONICET - UBA) en el área de Inteligencia Artificial y profesora del Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, donde dicta la materia “Ética & IA”. Además es Miembro del Comité Nacional de Ética en la Ciencia y Tecnología del Ministerio de Ciencia, Tecnología e Innovación Productiva y asesora en Inteligencia Artificial de la Dirección Nacional de Promoción de la Política Científica (DNPPC), dependiente de la Secretaría de Planeamiento y Políticas en Ciencia, Tecnología e innovación (SPYPCTEI). Es miembro de la campaña Stop Killer Robots y del SEHLAC.

Ricardo Oscar Rodríguez es Doctor en Ciencias de la Computación con especialización en Inteligencia Artificial. Es Profesor Asociado en el Departamento de Computación, FCEyN-UBA y miembro del Instituto de Ciencias de la Computación (UBA-CONICET). Sus trabajos científicos se inscriben en el desarrollo de modelos lógicos para el razonamiento

bajo incompletitud e incertidumbre. Actualmente dicta la materia “Ética & IA” y es miembro del SEHLAC y de la campaña Stop Killer Robots. Ha sido co-chair & financial chair de IJCAI2015 en Buenos Aires desde donde se lanzó la carta abierta de los científicos contra el uso de la IA en armamento.

APP

Asociación para Políticas Públicas



