IUS REVISTA DEL INSTITUTO DE CIENCIAS JURÍDICAS DE PUEBLA, MÉXICO. E-ISSN 1870-2147. NUEVA ÉPOCA VOL. 15, No. 48. JULIO - DICIEMBRE DE 2021. PP 9-34

# Retos para la regulación jurídica de la Inteligencia Artificial en el ámbito de la Ciberseguridad*

## *Challenges for the legal regulation of Artificial Intelligence in the field of Cybersecurity*

**Anahiby Anyel Becerril Gil***

**RESUMEN**

*La Inteligencia Artificial (IA) y las capacidades de aprendizaje automático están creciendo a un ritmo sin precedentes. La IA se presenta como una herramienta que puede mejorar el nivel de vida de los individuos, incluso resolver algunos de los desafíos más grandes del mundo. También es vista desde el enfoque de la salvaguarda de los ciudadanos, la defensa de las naciones y el mantenimiento de la paz y la estabilidad internacional. La IA se considera una tecnología de doble uso. La mejora en la toma de decisiones que se busca lograr con el empleo de IA se considera también en el ámbito de la ciberseguridad. A medida que la IA está cada vez más integrada en los sistemas críticos, deben protegerse ante posibles ciberataques. Los targeting killings, la desinformación y el empleo de sistemas inteligentes para buscar vulnerabilidades en infraestructuras se presentan como uno de los grandes desafíos de estos tiempos. En este trabajo analizamos diversas interrogantes sobre la regulación jurídica de la IA y sus impactos en el ámbito de la ciberseguridad.*

**PALABRAS CLAVE**

*Inteligencia, artificial, algoritmos, ciberseguridad.*

**ABSTRACT**

*Artificial Intelligence (AI) and machine learning capabilities are growing at an unprecedented rate. AI is presented as a tool that can improve the living standards of individuals, even solve some of the biggest challenges in the world. It is also seen from the approach of safeguarding citizens, defending nations and maintaining peace and international stability. AI is considered a dual use technology. The improvement in decision- making that is sought with the use of AI is also considered in the field of cybersecurity. As AI is increasingly integrated into critical systems, they should be protected against possible cyberattacks. Targeting killings, misinformation and the use of intelligent systems to search for vulnerabilities in infrastructure are presented as one of the great challenges of these times. In this paper we analyze various questions about the legal regulation of AI and its impacts in the field of cybersecurity.*

**KEY WORDS**

*Intelligence, artificial, algorithms, cybersecurity.*

---

*Artículo de Investigación postulado el 17 de febrero de 2020 y aceptado el 8 de octubre de 2020

**Especialista en Derecho y Tecnología. (anahiby@hotmail.com) orcid.org/0000-0002-5726-5400

**SUMARIO**



## 1. Introducción

La Inteligencia Artificial (en adelante IA) ha salido de los laboratorios a nuestra vida cotidiana. Ya sea que decidamos cuáles libros comprar, a cuál nuevo "amigo" incorporar a nuestra "red", con quién "emparejarnos" o incluso votar, un algoritmo se encuentra inmerso en la decisión (*algorithmic decision-making*). La encontramos inmersa en la toma de decisiones automatizadas en el ámbito administrativo o judicial, en el monitoreo de nuestra salud,[1] en sistemas de asistencia vehicular, de producción ciberfísica (Industria 4.0), en las finanzas y la banca,[2] así como en el campo militar. Los algoritmos también pueden predecir la orientación sexual de un individuo.[3] La realidad y "verdad" que se nos presenta también es desarrollada por máquinas que actualmente son controladas y programadas por personas. Nos encontramos en la era de la implementación.[4]

Para alimentar la toma de decisiones algorítmicas hacen falta dos insumos importantes: datos y energía. Los datos son el núcleo que alimenta a la IA, sin estos su desarrollo, entrenamiento y procesamiento para la toma de decisiones no sería posible. Si bien el debate internacional[5] sobre las aplicaciones militares



[1] Evidencia de ello es el monitoreo que se raliza para detectar, ofrecer recomendaciones sobre tratamiento y contener la pandemiade COVID-19 que en el año 2020 ha inmerso al mundo. Con la ayuda de la información que recolectan los dispositivos móviles (s*martphones*) se estan creando y analizando más datos para comprender con mayor precisión los síntomas y a la enfermedad.

[2] En donde puede ser empleada tanto para ofrecer un servicio personalizado de asistencia financiera virtual, analizar sus finanzas, hasta determinar si es sujeto a crédito, detectar fraude o lavado de dinero.

[3] Wang, Yilun, Kosinski, Michal, "Deep neural networks are more accurate than humans at detecting sexual orientation from facial image", Journal of Personality and Social Psychology, febrero 2018, Vol. 114, Issue 2, pp. 246-257.

[4] Lee, Kai-Fu, AI SUPERPOWERS. CHINA, SILICON VALLEY, AND THE NEW WORLD ORDER, Houghton Mifflin Harcourt, 2018, p. 13.

[5] *Cfr.* United Nations Institute for Disarmament Research (UNIDIR), Algorithmic Bias and the Weaponization of Increasingly Autonomous Technologies. A Primer, UNIDIR, 2018.

que emplean IA se ha centrado más en la toma de decisiones automatizada, no deberíamos pasar por alto que existe un elemento de debida diligencia tanto en la elección de la forma en que se obtienen los datos como en los que se emplean para entrenar e implementar los algoritmos. Los datos no siempre son objetivos ni tienen los impactos o la funcionalidad que el diseñador o usuario deseaba o esperaba. Estas "fallas en el sistema" pueden producir efectos nocivos. El "sesgo algorítmico" o fallas algorítmicas, nos recuerdan que los sistemas se encuentran desarrollados por humanos, para satisfacer necesidades y objetivos humanos, en consecuencia, reflejan determinados valores y decisiones.

El aumento exponencial en las velocidades de procesamiento de datos, junto con un incremento en las redes y avances en el hardware se encuentran remodelando las industrias, permitiendo llevar a la IA a un nivel comercial y más accesible. Es decir, no son solo los Estados y/o grandes empresas quienes se encuentran en posibilidad de desarrollar e implementar esta tecnología, sino que resultan posibles a cualquier individuo, incluidos los actores no estatales, delincuentes, grupos extremistas o terroristas.

La IA es una integración de atributos técnicos y sociales. Por un lado, la IA promueve la transformación de la economía y la sociedad. Se ha convertido en una nueva competitividad central y los países que obtendrán mayores beneficios serán quienes la han adoptado como parte de una estrategia nacional; Estados Unidos, la Unión Europea, China, Francia, han publicado en los últimos años estrategias para optimizar los beneficios de la IA, en la búsqueda por liderar su desarrollo.

Como sucede con otras tecnologías, la IA trae consigo riesgos y desafíos, la expansión de los sistemas analíticos y de toma de decisiones basados en algoritmos que funcionan con IA facilitan nuevas formas de vigilancia, otro ejemplo lo tenemos en los drones autónomos, los cuales pueden dejar caer medicinas en lugares de difícil acceso o bombas, localizar personas para salvarlas o matarlas. Fuera de control podrán causar graves daños.

Todos estos cambios han venido acompañados de gran agitación e incertidumbre. Gracias a su versatilidad la IA tiene aplicaciones en múltiples ámbitos, incluido el entorno de la seguridad nacional, lo que en algunos casos ha resultado controversial.



Las tecnologías que ha traído consigo este nuevo siglo, habilitadoras de la Cuarta Revolución Industrial (4IR) – genética, nanotecnología y la robótica (GNR)[6]- se han vuelto cada vez más accesibles. A diferencia de la energía

[6] Acrónimo para "genetics, nanotechnology and robotics".

eléctrica o nuclear (el desarrollo de armas de esta última), por ejemplo, no requieren grandes instalaciones ni materias primas, sino el conocimiento para su desarrollo e implementación. Pensemos en un escenario donde un usuario, gracias a la accesibilidad y conectividad, pueda aprender a desarrollar software y a través de éste realice intromisiones a sistemas de cómputo, de una empresa, de un banco, de una infraestructura de información crítica. Estamos ante la presencia de armas de destrucción masiva habilitadas por la información y el conocimiento (*knowledge-enabled mass destruction,* KMD).[7]

El empleo de sistemas autónomos y robóticos de doble uso está a punto de aumentar dramáticamente, esto como consecuencia de su proliferación y bajo costo. Las mejoras en la utilización de técnicas de aprendizaje automático y su capacidad sin duda han incrementado el interés tanto de los actores no estatales como de los estados nacionales para su empleo con diversos fines.

La introducción al ciberespacio ha tenido beneficios para todo tipo de actores. Los principales estados construyeron poderosas ciberarmas, realizaron un extenso espionaje cibernético y mejoraron las operaciones militares existentes con redes digitales. El bajo costo de las capacidades cibernéticas ha resultado mucho más accesible que sus equivalentes no cibernéticos, estados más pequeños con ejércitos menos poderosos también han hecho uso de la tecnología. Asimismo, los actores no estatales hostiles, incluidos criminales y terroristas, han hecho un uso efectivo de las herramientas cibernéticas para actividades geográficamente dispersas que serían mucho más difícil de llevarse a cabo. Esta situación puede resultar exponencial con el uso de la IA.

Desde su empleo ya sea en el desarrollo de inteligencia y análisis para hacer frente a las amenazas cibernéticas, o en el desarrollo de herramientas para mitigar actividades maliciosas y detectar vulnerabilidades en los sistemas (ciberdefensa), el avance en el desarrollo de la IA enfocada a la ciberseguridad nacional podría ser tal vez igualable en impacto al desarrollo nuclear, el informático y la biotecnología; lo que significa que estamos ante el desarrollo de una carrera armamentista.



Esto trae consigo cambios significativos en la estrategia, organización, prioridades y recursos a desplegar. Nos encontramos ante una nueva era en la búsqueda por la seguridad y estabilidad internacionales. El empleo de esta tecnología, sus riesgos, amenazas y retos constituyen el objeto de el presente artículo.

---

[7] Joy, Bill, "Why the Future Doesn´t Need Us", *Wired*, 4 de enero de 2000, [Consultada 06 de octubre de 2020], Disponible en: https://www.wired.com/2000/04/joy-2/

## 2. La inteligencia creada por humanos

*Until they become conscious they will never rebel,*
*and until after they have rebelled they cannot*
*become conscious* (George Orwell)

La inteligencia o el acto de entender como lo realiza la mente humana, artificial, o sea, hecha por el hombre,[8] implica que como humanos estamos desarrollando máquinas que tengan la capacidad de adquirir conocimientos y habilidades, pero más importante, implementar la racionalidad. Significa que una entidad que se suponga inteligente debe ser capaz de adquirir conocimiento a través de diversas formas, ya sea leyendo información (datos), procesando textos, argumentando con otros entes o a través de la experiencia. A partir de ahí, el conocimiento adquirido deberá ser razonado y comprendido, para ser aplicado en el contexto o situación correspondiente. En este sentido, una máquina se considerará inteligente si puede razonar. Un sistema de IA debe aprender, razonar y aplicar el conocimiento adquirido en la solución de problemas, debe ser capaz de elegir la mejor acción a tomar para el logro de un objetivo (o varios), con relación a los recursos disponibles y criterios.

> "La inteligencia artificial es la aclaración del proceso de aprendizaje humano, la cuantificación del proceso de pensamiento humano, la explicación del comportamiento humano y la comprensión de lo que hace posible la inteligencia. Es el último paso de los hombres para comprenderse a sí mismos".[9]

*Una ciencia nueva y prometedora.*

Una definición más amplia sobre la IA, es la brindada por el *Independent High-Level Expert Group on Artificial Intelligence* (HLEG), señalada como:

> "... sistemas de software (posiblemente también hardware) diseñados por humanos que, dando un objetivo complejo, actúan en la dimensión



[8] Diccionario Real Academia Española [Consultado el 10 de diciembre de 2019] Disponible en: https://dle.rae.es/artificial?m=form

[9] Lee, Kai-Fu, AI SUPERPOWERS. CHINA, SILICON VALLEY, AND THE NEW WORLD ORDER, Houghton Mifflin Harcourt, 2018, p. 7.

> física o digital percibiendo su entorno a través de la adquisición de datos, interpretando los datos estructurados o no estructurados recopilados, el conocimiento, o el procesamiento de la información, derivado de estos datos y decidir la mejor acción(es) a tomar para lograr el objetivo dado. Los sistemas de inteligencia artificial pueden usar reglas simbólicas o aprender un modelo numérico, y también pueden adaptar su comportamiento analizando cómo el entorno se ve afectado por sus acciones anteriores".[10]

Como disciplina científica incluye varios enfoques y técnicas, dentro de las que se encuentran el aprendizaje automático (*machine learning*), el razonamiento automático (*machine reasoning*) y la robótica (donde se integran las técnicas de sistemas ciberfísicos).

El aprendizaje automático o *machine learning* (ML) es el subconjunto de la IA que se centra en enseñar a las máquinas cómo aprender cosas nuevas y tomar decisiones aplicando algoritmos a los datos. Es decir, se ocupa de los algoritmos que permiten que los sistemas de la IA aprendan. El ML es aprendizaje supervisado en el que se instruye al sistema utilizando datos que le permiten capacitarse.[11] Sin embargo, también existen algoritmos no supervisados, semisupervisados y de refuerzo. Un ejemplo del aprendizaje por reforzamiento constituye en Google DeepMind´s Alpha Go.

Estas características convierten al aprendizaje automático no solo en un subconjunto importante de IA, sino también en un método para su implementación. Al aprender el conocimiento y las reglas de los datos de muestra, el aprendizaje automático los usará para hacer inferencias y decisiones prácticas.

El *Deep Learning* (DL) o aprendizaje profundo constituye a su vez un subcampo del ML que emplea datos para enseñar a las computadoras a cómo hacer las cosas que antes solo los humanos éramos capaces de hacer. Este subcampo es el responsable de imitar el mecanismo del cerebro humano para interpretar los datos, mediante la construcción de redes neuronales. Dentro del DL se encuentra el *Natural Language Processing* (NLP), el cual decodifica el significado detrás de las palabras, constituyendo así la conexión entre las computadoras y cualquier lenguaje humano natural, cuyo máximo exponente constituye Watson de IBM.



[10] High-Level Expert Group on Artificial Intelligence (HLEG), "A definition of AI: Main Capabilities and Disciplines", European Comission, 2019, p. 6 [Consultada 06 de octubre de 2020], Disponible en: https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

[11] El aprendizaje por refuerzo, realizado por recompensas artificiales, se emplea para aprender nuevas tareas.

El núcleo del aprendizaje automático es generar modelos a partir de datos con el empleo de algoritmos. Para la Secretaría de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) los algorítmos constituyen una "secuencia de reglas que deben realizarse en un orden exacto para llevar a cabo una determinada tarea".[12] Mientras que el NIST[13] refiere que un algoritmo constituye un "conjunto de pasos computables para lograr un resultado deseado". Constituyen una secuencia de pasos lógicos que genera un *output* a partir de un *input* dado.

A los algoritmos de aprendizaje disponibles, podemos "alimentarlos" con datos empíricos y luego "cosechar" modelos estadísticos, que revelan las reglas de las cosas y nos proporcionan juicios correspondientes para los eventos futuros (predicción). Se puede decir que el aprendizaje automático es el estudio de "algoritmos de aprendizaje", que son esencialmente versiones avanzadas de algoritmos ordinarios que hacen que los programas de computadora sean más inteligentes al descubrir y aprender automáticamente las reglas de datos.

Aunque hemos llegado a construir IA que a su vez crea otras inteligencias artificiales mejor que los humanos,[14] las personas que crean estas inteligencias a menudo no entienden cómo la IA puede llegar a comportarse. Por tanto, no se deben exagerar las ventajas de la IA, ni se debe prohibir su uso. Es viable que busquemos ventajas y evitemos desventajas maximizando el beneficio de la IA.

## 3. ¿Más inteligente, más veloz y más resiliente?: el empleo de la IA en la ciberseguridad nacional

Nos encontramos en una carrera armamentista, con sectores y sistemas, (bancario, de salud, industria, manufactura, gobierno), que se encuentran bajo constantes y complejos ciberataques.[15] El acceso no autorizado a sistemas o computadoras, la destrucción de sitios web, los *DDoS*, el robo de datos, *ransomware* (a través de botnets), así como la interrupción de servicios -entre otros- lleva a desarrollar nuevos mecanismos de protección, mientras que los



[12] OCDE, Algorithms and Collusion. Competition policy in the digital age, OCDE, 2017, Disponible en: https://www.oecd.org/daf/competition/Algorithms-and-colllusion-competition-policy-in-the-digital-age.pdf

[13] National Institute of Standards and Technology (NIST), "Dictionary of Algorithms and Data Structures", 1998 [Consultado el 10 de febrero de 2020] Disponible en: https://xlinux.nist.gov/dads/HTML/algorithm.html

[14] Según refiere el CEO de Google, su AutoML (AI inception) mejora a los humanos en la creación de IA; *Cfr.* Gard, Tom, Gooble´s new AI is better at Creating AI Than rhe Company´s Engineers, Futurism, 17 de mayo 2017 [Consultado el 10 de febrero de 2020], Disponible en: https://www.cnet.com/news/its-happening-googles-ai-is-building-more-ais/

[15] Cuestión distinta es que estos ataques sean exitosos.

atacantes a su vez desarrollan otros de ataque que cada vez resultan ser más sofisticados. La infraestructura de información crítica y los sistemas de los gobiernos no son la excepción, en la Comunidad Internacional existen crecientes preocupaciones por el despliegue de ciberoperaciones militares con el empleo de IA, tanto en época de conflicto o en tiempo de paz.

La ciberseguridad se encuentra "cada vez más afectada por los desarrollos en IA, principalmente porque muchas partes de la esfera digital están siendo transformadas por IA",[16] lo que amplía la superficie de ataque. Es una realidad que los Estados se encuentran desarrollando capacidades cibernéticas basadas en IA para la defensa y embate, entre las que se encuentran mayores capacidades de ciberdefensa, análisis forenses, nuevas vulnerabilidades de software en los mismos sistemas de IA, así como su uso deliberado y malicioso. Estas ciberoperaciones, reflejadas en ciberataques no sólo impactan de forma económica a los países, sino que causan daño a la población civil.

Para el Comité Internacional de la Cruz Roja (CICR), las operaciones cibernéticas pueden dañar la infraestructura de los países, al menos de dos maneras. Primero, pueden afectar la prestación de servicios esenciales a los civiles, como ha sucedido con los ataques cibernéticos contra las redes eléctricas y el sector de la salud. En segundo lugar, pueden causar daños físicos como fue el caso del ataque de Stuxnet contra una instalación de enriquecimiento nuclear en Irán en 2010 y un ataque contra una acería alemana en 2014. Mientras exista la necesidad de atacar existirá la necesidad de defender. Es un equilibrio de sistemas.

Nos encontramos en una carrera armamentista, por lo que deberíamos preocuparnos y atender la proliferación de herramientas cibernéticas que emplean IA, tema del cual la Comunidad Internacional aún debate si debería ser considerado del mismo modo que la proliferación de armas o de tecnología de doble uso. Otra realidad también es que no todos los países cuentan con el nivel de madurez necesario para la creación de capacidades. En este sentido, existen empresas que se encuentran ofertando la ciberseguridad como un servicio (*Cybersecurity as a Service*, CSaaS), a través de una diversidad de sistemas de ciberdefensa y ciberataque al mejor postor. Llevar a cabo un ataque cibernético contra un sistema de control industrial requiere cierta experiencia y sofisticación, a menudo malware (*custom-made malware*). Es aquí donde aumenta el riesgo de que las herramientas desarrolladas por actores con mejores recursos puedan ser reutilizadas o compradas por otros actores que carecen de la



[16] European Commission, Cybersecurity Our Digital Anchor. A European Perspective, European Comission, Joint Research Centre, Italy, p. 51.

experiencia necesaria para desarrollarlas desde cero. Lo cierto es que, mientras exista demanda por estos servicios la oferta persistirá.

La cuestión no es ya si la inteligencia artificial y el aprendizaje automático, serán fundamentales para el futuro de la ciberseguridad, sino el cuándo y cómo vamos a lidiar con su empleo en esta área. ¿Estamos preparados?

En la ceremonia inaugural de las Olimpiadas de Invierno celebradas en Pyeongchang (2018), pudimos observar el cielo dibujado de animaciones realizadas por 1218 drones trabajando al unísono. Estos "enjambres de drones[17]" (*drone swarms*), controlados por desarrolladores de Intel, operaban de manera independiente, comunicándose con una computadora central y no con los drones que se encontraban a su alrededor. Estas herramientas tecnológicas también pueden ser empleadas con fines muy distintos al de unión que conjugan los Juegos Olímpicos. Imaginemos por un momento que las instrucciones otorgadas a esos enjambres de drones sean con fines de atacar a la población civil.

Las capacidades de automatización y análisis de datos de la IA pueden ser empleados para analizar grandes cantidades de estos de forma eficiente, precisa y veloz. En un entorno de ciberdefensa, este sistema puede aprovechar lo que sabe, comprendiendo las amenazas pasadas para identificar futuros ataques similares incluso si los patrones cambian. Pero de la misma forma en que estas pueden ayudar a detectar ataques, pueden ser empleados con fines contrarios, buscando constantes vulnerabilidades de sistemas, cambiando patrones en cada intento, haciéndolos más difíciles de predecir y prevenir.

Los intentos de atacantes para comprometer las infraestructuras críticas son inevitables. Por ello los sistemas de defensa, así como de ataques se encuentran en constante evolución. Para avanzar en el combate contra actores maliciosos y permitir una mejor gestión de riesgos, se requieren enfoques avanzados e innovadores con aplicaciones técnicas basadas en IA que brinden flexibilidad y capacidad de aprendizaje, ayudando a detectar "puntos ciegos", brindando información sobre la situación o contexto cibernético, en tiempo real, volviéndose auxiliares en el combate a las amenazas del ciberespacio.

Entre las ventajas de la IA en la ciberseguridad se encuentran:

a. Descubrir nuevos y sofisticados cambios en la flexibilidad de los ataques;



[17] También denominados "enjambres robóticos" (*robotic swarms*).

b. Manejar de forma veloz y más eficiente mayores volúmenes de datos;
c. Un sistema de seguridad que pueda aprender a responder de manera más eficiente a las amenazas.

Uno de los principales retos para la defensa contra ciberataques, consiste en reconocer el punto de ataque, así como las vulnerabilidades del sistema que pueden ser explotadas por los atacantes. La detección, de forma automática, de amenazas, no solo ayuda a reducir el trabajo de los expertos, sino que puede implicar una forma más efectiva en su detección.

El tráfico constante de grandes cantidades de datos de seguridad en la red implica un monitoreo al mismo nivel, este trabajo gradualmente traerá consigo el desarrollar esfuerzos extraordinarios (mayor personal y tiempo) para su monitoreo, además de la detección de comportamiento malicioso. En este sentido, la IA puede ayudar a expandir el monitoreo y detectar comportamientos sospechosos, permitiendo enfocar sus capacidades en la respuesta a estos incidentes.

Los sistemas de IA tienen la oportunidad de conocer el comportamiento de las aplicaciones y actividades de la red, lo que le permitirá detectar el comportamiento usual o inusual que se esté llevando a cabo, es decir, distinguir el patrón normal del sistema o red.

En algunos casos, los atacantes optan por acometer contra el eslabón más débil, con métodos que van desde software malicioso (*malware*) en diversos formatos al dispositivo de la víctima, explotar vulneraciones día cero, entre otras. En ese sentido seguiremos siendo los humanos los más vulnerables.

## 4. Los impactos de la "realidad" creada por IA

Las tendencias y los temas descritos anteriormente podrían combinarse para crear un panorama de poder militar muy diferente al que existe hoy. A continuación, ofrecemos algunos escenarios a través de los cuales las capacidades crecientes de la IA podrían transformar el poder militar. No se trata de predicciones firmes. Más bien, pretenden ser provocativos y demostrar cuán amplio es el rango de resultados posibles, dadas las tendencias actuales. Además, no son alternativas mutuamente excluyentes. Más de uno o varios podrían suceder simultáneamente.



*a. Robotic swarms.*

Los grupos de robots, como los drones o micro-drones, preprogramados para realizar tareas complejas, de forma flexible y robusta, funcionando como un microorganismo, pueden llegar a "compartir un cerebro distribuido para la toma de decisiones y se adaptan entre sí como enjambres en la naturaleza"[18]. Estos conjuntos de unidades robóticas (idénticas o diferentes), controladas por pocas o una sola persona, se comportan como un enjambre. Son miembros autónomos basados en algoritmos distribuidos, es decir, el algoritmo del enjambre se ejecuta por separado en cada robot que lo conforma. A diferencia de un grupo de robots, que reaccionan con base en reglas internas y el estado del medio ambiente, por lo que exhiben un comportamiento colectivo a través de la colaboración entre sus unidades individuales. Para lo anterior requieren de comunicación que les permita el intercambio de información entre ellos, lo que puede llevarse a cabo a través de Bluetooth o Wi-Fi.

Estos sistemas multi-robot que coordinan sus acciones para trabajar colectivamente hacia la ejecución de un objetivo,[19] que responden al mando humano (hasta ahora) son una realidad. Para su formación, supervisión, separación, trayectoria de vuelo, distribución de tareas e identificación de objetivos, dependen de algoritmos, es decir, de la interacción máquina-máquina (*machine-to-machine*). Estas arquitecturas de control (algoritmos) son las que definirán las tareas entre las unidades robóticas que conforman el enjambre.

*¿Cómo trabajan? Existen diversos tipos de arquitecturas de control que pueden implementarse en estos swarms*: (i) control centralizado, en donde las órdenes previstas por un humano se dirigen a un robot que actúa como el controlador central del enjambre, en estos casos no existe una colaboración directa entre las unidades, sino que dependerán de su control central; (ii) control jerárquico, en donde se eligen a diversos "líderes" dentro del enjambre, de esta forma el control de cada robot podrá controlarse por otros agentes de diversos niveles; (iii) control a nivel de conjunto, método descentralizado que permite transmitir las instrucciones al enjambre como un todo (no de forma individual), permitiendo a los robots decidir la forma de ejecutar el comando; (iv) control de comportamiento, a cada robot se le provee previamente de una biblioteca de comportamiento y los operadores indican cuál deben ejecutar.[20]



[18] Will Roper en McCullough, Amy, "The Looming Swarm", 22 de marzo 2019, [Consultada 06 de octubre de 2020], Disponible en: https://www.airforcemag.com/article/the-looming-swarm/

[19] United Nations Institute for Disarmament Research, UNIDIR, Swarm Robotics: Technical and Operational Overview of the Next Generation of Autonomous Systems, UNIDIR, *2020,* [Consultada el 06 de octubre de 2020] Disponible en: https://unidir.org/publication/swarm-robotics-technical-and-operational-overview-next-generation-autonomous-systems

[20] Ibíd.

Algunas de las áreas de oportunidad que destaca la UNIDIR para el empleo de estos enjambres en el ámbito militar son: operaciones de inteligencia, vigilancia y reconocimiento (ISR, por las siglas en inglés para *Intelligence, Surveillance* y *Reconnaissance*)[21]; vigilancia y protección perimetrales,[22] sistemas de armas distribuidas (que distribuyen objetivos de manera autónoma entre las unidades robóticas del grupo),[23] para la defensa y protección contra otros enjambres; la detección de minas, señuelos, entre otros.

*b. Datos "envenenados".*

La participación del aprendizaje automático en sistemas militares creará nuevos tipos de vulnerabilidades y nuevos tipos de ataques cibernéticos que se dirigen a los datos de entrenamiento deestos sistemas. Gartner estima que hasta el año 2022 el 30% de los ataques cibernéticos de IA aprovecharán el envenenamiento de datos de entrenamiento, el robo de modelos de IA o muestras adversas para atacar sistemas impulsados por IA.

Si bien la IA funciona para proteger sistemas, redes e infraestructura, ésta debe ser protegida. La seguridad de la IA incluye tres perspectivas clave: la protección de sistemas impulsados por esta misma tecnología, que incluye la protección de datos y líneas de entrenamiento de IA, así como los modelos de ML; aprovechamiento de la IA para mejorar la defensa de seguridad, lo que

[21] Por ejemplo el proyecto Perdix de la *Strategic Capabilities Office* de los Estados Unidos en colaboración con el Naval Air Systems Command. Este proyecto se encuentra conformado por micro-drones autónomos capaces de ISR y otras misiones. Estos pueden ser lanzados en aire, mar y tierra, operando en enjambres pequeños o grandes para llevar a cabo sus misiones; *Cfr.* The Strategic Capabilities Office, "Perdix Fact Sheet", [Consultada el 06 de octubre de 2020], Disponible en: https://dod.defense.gov/Portals/1/Documents/pubs/Perdix%20Fact%20Sheet.pdf?ver=2017-01-09-101520-643

[22] Como el proyecto Roborder (R0bots for BORDER surveillance) de la Unión Europea, el cual tiene como objetivo: "desarrollar y demostrar un sistema de vigilancia de fronteras autónomo completamente funcional con robots móviles no tripulados que incluyen vehículos aéreos, de superficie de agua, submarinos y terrestres que incorporarán sensores multimodales como parte de una red interoperable"; *Cfr.* Comisión Europea, "autonomous swarm of heterogeneous R0bots for BORDER surveillance. Ficha informativa", [Consultada el 06 de octubre de 2020] Disponible en: https://cordis.europa.eu/project/id/740593/es; Proyecto R0BORDER, Consultada el 06 de octubre de 2020] Disponible en: https://roborder.eu

[23] Un ejemplo de este sistema es el proyecto Collaborative Operations in Denied Environment (CODE) de DARPA (Defense Advanced Research Projects Agency) de Estados Unidos, este proyecto tiene como objetivo mejorar la escalabilidad y rentabilidad de los sistemas de aeronaves no tripulados (unmanned aircraft systems, UAS) superando las limitaciones del control del piloto y operador de sensores, a través de algoritmos y software que, además de ser resistente a limitantes como el ancho de banda e interrupciones de comunicaciones, mejore la autonomía colaborativa de las UAS para trabajar juntos bajo la supervisión de un solo individuo; *Cfr.,* DARPA, Collaborative Operations in Denied Environment (CODE), [Consultada el 06 de octubre de 2020], Disponible en: https://www.darpa.mil/program/collaborative-operations-in-denied-environment

implica el uso de ML para comprender patrones, descubrir ataques y automatizar partes de procesos de ciberseguridad.

*c. Signature strikes*

"Simplemente eran civiles inocentes en el lugar equivocado en el momento equivocado", un error en el perfilamiento o la identificación.

> El 4 de febrero del año 2002, un dron Predator, operado por la CIA, "spotted" a tres hombres parados en Zawhar Kili, y abandonó el complejo mujahedeen localizado cerca de la ciudad de Khost en la provincia de Afganistan. Uno de ellos era alto; los otros estaban supuestamente actuando con reverencia en torno a él. Convencido de que los hombres eran objetivos legítimos -y esperando que el nombre alto fuera Osama bin Laden- la CIA disparó un misil Hellfire por el Predator, matando a los tres instantáneamente.[24]

Reportes posteriores determinarían que el hombre alto no era Osama bin Laden, además de describir con posterioridad que ninguno de los tres hombres eran miembros de Al-Qaeda o el talibán. Este ataque fue el primer ejemplo -conocido- de lo que se denomina "*signature strike",* es decir, un ataque de drones que arremete contra grupos de hombres que llevan cierta "firma" (característica) asociada con la actividad terrorista, cuya identidad es desconocida.[25] La diferencia con los "*personality strikes*", es que en estos existe un "alto nivel de confianza" sobre el conocimiento o identidad del objetivo. Estados Unidos ha defendido el uso de ese tipo de ataques, bajo el argumento de que son parte de "un conflicto armado con Al-Qaeda, el Talíban y fuerzas asociadas", lo que no los legitima en el ámbito del Derecho Internacional Humanitario. Estos ataques se desarrollan tanto en época de conflicto como o fuera de él, además pueden llegar a constituir una violación al Derecho Internacional Humanitario o al Derecho Internacional de los Derechos Humanos.

El uso de *signature strikes* basados en características o comportamiento determinado de una persona que lo identifique como un combatiente no permite establecer si la víctima del ataque era un objetivo legítimo, por lo tanto, violentarían el principio de distinción que establece el DIH. Un *signature strike*



[24] Sifton, John, "A Brief History of Drones", The Nation, 7 de febrero de 2012 [Consultado el 10 de diciembre de 2019] [Disponible en: http://www.thenation.com/article/166124/brief-history-drones#.]
[25] Klaidman, D. Kill or Capture: The War on Terror and the Soul of the Obama Presidency, Harcourt, 2012, p. 41.

solo será legal si se realiza respondiendo de forma afirmativa a dos cuestionamientos; primero ¿resulta la firma suficiente para establecer que la victima del ataque es un objetivo legítimo?; y, segundo, ¿resulta suficiente la firma para determinar que el individuo es realmente quien se comportó como la firma determinó? En caso de que exista duda sobre la persona y la firma, se debe presumir que se trata de un civil.

*d. La "guerra" de la desinformación*

*As with law, technology has a similar capacity to influence an individuals behaviour* (Filippi & Wright, 2018)

"AI podría poner en peligro totalmente la democracia".[26] El software artificialmente inteligente y los programas de aprendizaje automático continúan para crear mejores herramientas y perpetuar un fraude. Combinado con ataques cibernéticos y bots en redes sociales, los medios falsificados habilitados por la IA amenazan la estabilidad de una economía, democracia o régimen gubernamental.

La desinformación masiva, también se perfila como una amenaza para las sociedades. La información que transita por las redes no siempre es cierta y ha llegado a causar mucho daño, tanto que incluso las actividades de desinformación han llegado a ser consideradas como una amenaza a la seguridad nacional.[27] La combinación de automatización, perfilamiento y marketing puede tener un impacto significativo en la opinión pública durante importantes debates, elecciones y crisis políticas.

El funcionamiento de algoritmos y sistemas de recomendación automatizados que pueden crear "burbujas de filtro" -cámaras de eco totalmente automatizadas- en las que los individuos solo ven fragmentos de información que confirman sus propias opiniones o coinciden con su perfil, tienen efectos trascendentales para los procesos democráticos en la sociedad.



Este tipo de desinformación, conocida como "*fake news*" (noticias falsas), ha prosperado con la ayuda de las TIC, las redes sociales y como consecuencia

[26] Thomson, Nicholas, "Emmanuel Macron Talks to WIRED About France's AI Strategy", Wired, 31 de marzo 2018, [Consultada el 10 de febrero de 2020], Disponible en: https://www.wired.com/story/emmanuel-macron-talks-to-wired-about-frances-ai-strategy/

[27] La Estrategia Nacional de Seguridad española considera a las campañas de desinformación como "conflictos híbridos", caracterizados por incorporar operaciones de información, subversión, presión económica y financiera junto acciones militares", las cuales pueden llevarse a cabo tanto por actores estatales como no-estatales; *Cfr.* Presidencia del Gobierno, 2017, pp. 34 y ss.

de la situación fragmentada en que se encuentra el mundo, tanto en el ámbito político como social. El término hace alusión a una serie de fenómenos: desde intentos deliberados de socavar elecciones o la seguridad nacional, hasta cualquier punto de vista que desafíe el consenso de otro. Las prácticas de desinformación selectiva sobre la información de la opinión o noticas que sean de nuestro interés también traen repercusiones en nuestra conducta. Tenemos información ideológicamente alineada pero no sabremos si es cierta o falsa, por no tener un panorama completo con qué contrastarla.

El problema que se debe abordar es el uso de las ciber-tropas (*cybertroops*), es decir, el uso de herramientas como bots por gobiernos, partidos militares o políticos, los cuales gastan recursos significativos para generar contenido con el fin de manipular la opinión pública nacional y/o extranjera a instancias de ellos. Es probable que su papel en la generación de "opinión pública" aumente como un fenómeno global.

*e. Fallas "repentinas"*

Las interacciones inesperadas de sistemas autónomos ocasionan "fallas repentinas". Los sistemas autónomos pueden tomar decisiones increíblemente rápido, mucho más rápido de lo que los humanos pueden monitorearlos y restringirlos sin la ayuda de máquinas. Debido a la alta velocidad de los sistemas autónomos, las interacciones inesperadas y los errores pueden descontrolarse rápidamente. Debemos considerar a la ciberseguridad o el equivalente de vehículo autónomo de un choque repentino. O muchos carros autónomos, a horas "pico" en una gran ciudad.

El proceso de verificación y validación para sistemas autónomos que aprovechan el aprendizaje automático aún está en su etapa inicial y el bloqueo repentino sugiere que incluso los sistemas que funcionan mejor que los humanos en más del 99% de sus operaciones pueden ocasionalmente tener fallas catastróficas e inesperadas. Esto es especialmente preocupante dada la naturaleza adversaria de la guerra y el espionaje. Como usuarios deseamos emplear esta tecnología y tener la garantía de que los vehículos autónomos sean exitosos y seguros. Los adversarios militares o actores con fines maliciosos de los sistemas robóticos, serán menos amables.

*f. Replica o reutilización de sistemas de IA*

El robo y la réplica de sistemas de inteligencia artificial militares provocarán que estas armas cibernéticas caigan en las manos equivocadas.

Las herramientas y ciberarmas pueden proliferar de una manera única que es difícil de controlar. Primero, el ciberespacio es un dominio global: siempre que el atacante pueda superar las medidas de seguridad y defensa cibernéticas vigentes, se puede acceder a cualquier nodo de red e información que resida en ella desde cualquier parte del mundo. Al mismo tiempo, las herramientas cibernéticas se pueden reutilizar o rediseñar. La combinación de estas dos características significa que cuando las herramientas cibernéticas se han utilizado, robado, filtrado o están disponibles de otra manera, los actores que no sean los que las desarrollaron podrían encontrarlas, aplicarles ingeniería inversa y reutilizarlas para sus propios fines.

Finalmente, el hecho de que las herramientas y métodos cibernéticos puedan reutilizarse es uno de los factores que hacen que la atribución técnica rápida y confiable de los ciberataques resulte un proceso desafiante.

## 5. Regulación

El fundamento, en el desarrollo de instrumentos que regulen el empleo de la IA deben ser los instrumentos declarativos que traen consigo el enfoque de la vida y libertades, asociados con la dignidad humana. Centrados en el bien y fraternidad de la humanidad. En teoría los derechos funcionan independientemente de cualquier tecnología. Esta afirmación se encuentra respaldada por la Resolución A/HRC/20/L.13 aprobada por el Consejo de Derechos Humanos de la Organización de las Naciones Unidas, en donde se reconoce que el disfrute y protección de los DDHH debe estar garantizado tanto en el mundo online, como en el mundo offline.

Sin embargo, en la práctica el empleo de la tecnología impacta en sí y la forma en que los individuos disfrutan estos derechos. La dificultad de encontrar la unanimidad para aplicar dichos principios o una regulación consensuada entre los *stakeholders* sugiere que las posibilidades de la adopción y la aplicación confiable son bajas. A ello se le suman las preocupaciones en torno a cómo el empleo de las TIC afectan "los intereses de toda la comunidad internacional",[28] reconociendo que las tecnologías "también pueden ser empleadas



[28] "... the dissemination and use of information technologies and means affect the interests of the entire international community"; *Cfr.* Asamblea General de Naciones Unidas, Preámbulos de las Resoluciones A/RES/55/28 de 20

con finalidades distintas de los objetivos de mantener la estabilidad internacional y la seguridad".[29]

Facebook ha sido cuestionado por la Misión Investigadora nombrada por el Consejo de Derechos Humanos de la ONU para Birmania, al haber sido utilizada (su plataforma) por actores clave como: partidos políticos nacionalistas, miembros del gobierno, así como actores militares para diseminar una campaña de odio y desinformación incitando así a la violencia étnica contra los Rohingya[30] (población musulmana minoritaria) en Birmania en el año 2017.

Con 20 millones de usuarios de una población total de 53 millones, en el país asiático Facebook representa Internet. Alguna de la evidencia presentada en la Corte Internacional de Justicia,[31] fueron los posts publicados en la página de Facebook del General en Jefe Min Aung Hlaing (perfil ahora inexistente en Facebook), en donde se jactaba de la eficiencia con que "el problema Bengalí" se resolvería de una vez por todas.[32] Esto fue ejecutado a través de "*clearance operations*",[33] desplegadas el 25 de agosto del año 2017, las cuales fueron "diseñadas para infundir terror inmediato" en donde "apuntaban y aterrorizaban a la población rohingya", mismas que se extendieron por mas de dos meses. Al concluir más, reporta la Misión de la ONU, 40% de todas las aldeas en el norte de Rakhine habrían sido destruidas parcial o totalmente. Uno de los resultados

---

de Noviembre del año 2000; A/RES/56/19 de 29 de noviembre de 2001; A/RES/59/61 de 3 de diciembre de 2004; A/RES/60/45 de 8 de diciembre de 2005; A/RES/61/54 de 6 de Diciembre de 2006; A/RES/62/17 de 05 de diciembre de 2007; A/RES/63/37 de 2 de diciembre de 2008; A/RES/64/25 de 2 de diciembre de 2009.

[29] Preámbulos de las Resoluciones A/RES/58/32 de 08 de diciembre de 2003; A/RES/59/61 de 3 de diciembre de 2004; A/RES/60/45 de 8 de diciembre de 2005; A/RES/61/54 de 6 de diciembre de 2006; A/RES/62/17 de 5 de diciembre de 2007; A/RES/63/37 de 2 de diciembre de 2008; A/RES/64/25 de 2 de diciembre de 2009.

[30] Recientemente, en un fallo unánime, la Corte Internacional de Justicia dictó medidas cautelares para proteger a la comunidad rohinyá de Myanmar, de conformidad con lo establecido en la Convención para la Prevención y la Sanción del Genocidio (1948), determinando que "están desprotegidos en Myanmar y corren un peligro real e inminente de ser víctimas de un genocidio", además de ordenar conservar las pruebas que puedan demostrarlo; *Cfr.,* Corte Internacional de Justicia, APPLICATION OF THE CONVENTION ON THE PREVENTION AND PUNISHMENT OF THE CRIME OF GENOCIDE, (THE GAMBIA v. MYANMAR), Request for the indication of provisional measures, Procedimiento 23 de enero de 2020 [Consultado el 13 de febrero de 2020], Disponible en: https://www.icj-cij.org/files/case-related/178/178-20200123-ORD-01-00-EN.pdf

[31] Sesión pública celebrada el jueves 12 de diciembre de 2019, a las 10 a.m., en el Palacio de la Paz, Presidente Yusuf presidiendo, en el caso relativo a la aplicación de la Convención para prevenir y sancionar el delito de genocidio (Gambia c. Myanmar) [Consultado el 13 de febrero de 2020], Disponible en: https://www.icj-cij.org/files/case-related/178/178-20191212-ORA-01-00-BI.pdf

[32] The Bengali problem was a long-standing one which has become an unfinished job despite the efforts of the previous governments to solve it. The government in office is taking great care in solving the problem"; *Cfr.* UN HRC, Report of the detailed findings of the Independent International Fact-Finding Mission on Myanmar, UN doc. A/HRC/39/CRP.2 (17 Sep. 2018), para. 75 [Consultado el 10 de febrero de 2020], Disponible en: https://www.ohchr.org/Documents/HRBodies/HRCouncil/FFM-Myanmar/A_HRC_39_CRP.2.pdf

[33] Estas "clearance operations" constituyeron operaciones de limpieza étnica.

fue la huida de más de 725,000 rohingya a Bangladesh y aproximadamente 10,000 personas asesinadas.

Las medidas de control de armas (duales) requieren un entorno propicio para su implementación efectiva y una voluntad política. El entorno político actual, las tensas relaciones entre países como China, Estados Unidos, Rusia, el debilitamiento del multilateralismo, la erosión y polarización de los regímenes y las sociedades, seguido de una falta de *cybertrust* y el incremento en obtener la hegemonía y control sobre esta nueva tecnología influyen sobre la adopción de medidas o herramientas, vinculantes o no.

No es una novedad que existen diferentes opiniones sobre los objetivos del control de armas en la Comunidad Internacional, para Persi *et al.*[34], los siguientes cuatro objetivos del control de armas deben considerarse en cualquier medida para su implementación:

1. *Estabilidad. Eliminar los incentivos para un primer ataque, con la finalidad de prevenir una guerra accidental y reducir el riesgo de una escalada militar mediante una mayor previsibilidad y transparencia;*
2. *Seguridad: reducir los riesgos asociados a las operaciones militares;*
3. *Legalidad: garantizar la compatibilidad con las obligaciones internacionales, en específico con el Derecho Internacional Humanitario y el Derecho de los Derechos Humanos;*
4. *Eficacia: Proporcionar incentivos suficientes y por tanto buenas perspectivas, para la implementación de controles que produzcan las conductas deseadas por parte de los Estados interesados.*

En el ámbito internacional contamos con la experiencia de tratados e instrumentos internacionales que buscan el control de la proliferación de diversos tipos de armas. Si bien el aplicar los enfoques tradicionales de control de armas a las TIC ha tenido sus dificultades, con la situación actual podría resultar complicada la negociación próxima de un instrumento unilateral único que regule el uso militar de la IA. Lo que puede lograrse es el reconocimiento de instrumentos y compromisos existentes en el empleo de este tipo de tecnología.



Con base en la resolución del GGE del año 2015 en donde se reconoce que la Carta de las Naciones Unidas resulta aplicable al ciberespacio, podríamos

[34] Persi Paoli. G., Vignard. K., Danks. D, and Meyer. P., Modernizing Arms Control: Exploring responses to the use of AI in military decision-making, UNIDIR, Ginebra, p. 2.

aprovechar para el reconocimiento de los Convenios de Ginebra, en específico el Protocolo I[35] el cual refiere en su artículo 36:

> Artículo 36. Armas nuevas. Cuando una Alta Parte contratante estudie, desarrolle, adquiera o adopte una nueva arma, o nuevos medios o métodos de guerra, tendrá la obligación de determinar si su empleo, en ciertas condiciones o en todas las circunstancias, estaría prohibido por el presente Protocolo o por cualquier otra norma de derecho internacional aplicable a esa Alta Parte contratante.

Si bien el término "nuevas armas, medios y métodos de guerra" no se encuentra definido, lo que permite que estén sujetos a interpretaciones, el Comité Internacional de la Cruz Roja[36] señala que el término se refiere a "armas de todo tipo, ya sean antipersonal o anti-material", letales o no, incluidos los sistemas de armas.

De esta forma este artículo requiere a los estados realizar una revisión legal de todas las armas, medios y métodos de guerra nuevos, para determinar sí su empleo se encuentra prohibido por el Derecho Internacional. En esta tesitura los nuevos sistemas de armas que se desarrollen con la IA (sistemas de armas autónomos y otros) deben estar sujetos a dicha revisión. Este proceso se lleva a cabo en el ámbito nacional, la cooperación e intercambio de información que en materia de ciberseguridad se ha establecido al amparo de algunos tratados (incluidos los de libre comecio) pueden ayudar a los estados a comparir puntos de vista y en algún momento desarrollar instrumentos de revisión de estas armas. Esto último a través de un enfoque *multistakeholder*, respetando los Derechos Humanos.

Existen diversos esfuerzos internacionales por buscar un estándar en valores y principios para el empleo de la IA, a saber:

a. El 16 de mayo de 2018, Access Now y una coalición de grupos de derechos humanos y tecnología emitieron una Declaración abierta a la firma de organizaciones y gobiernos titulada *La Declaración de Toronto: Protección de los derechos a la igualdad y la no discriminación en los sistemas de aprendizaje automático*. El enfoque de la Declaración,



[35] Protocolo I adicional a los Convenios de Ginebra de 1949 relativo a la protección de las víctimas de los conflictos armados internacionales, 1977, [Consultada el 06 de octubre de 2020], Disponible en: https://www.icrc.org/es/document/protocolo-i-adicional-convenios-ginebra-1949-proteccion-victimas-conflictos-armados-internacionales-1977#GUERRA

[36] International Committee of the Red Cross (ICRC), "A Guide to the Legal Review of Weapons, Means and Methods of Warfare", ICRC, Ginebra, 2006, pp. 4-3.

de acuerdo con el Preámbulo, es examinar las implicaciones positivas y negativas de los sistemas de aprendizaje automático, garantizar que los derechos humanos estén protegidos, resguardar contra la discriminación, promover la inclusión, la diversidad y la equidad, y proporcionar soluciones a esas personas afectadas injusta o negativamente. Señalaron que los sistemas a menudo son opacos y pueden conducir casi sin esfuerzo a prácticas discriminatorias y represivas a menos que se establezcan salvaguardas para mitigar tales eventos. Al observar los muchos beneficios que surgen de los sistemas de aprendizaje automático, también pueden aparecer problemas que afectan la privacidad, la protección de datos, la libertad de expresión, la participación en la vida cultural y la igualdad ante la ley. Reconociendo que el Derecho Internacional establece claramente el deber de los estados de proteger los derechos humanos;

b. En abril del año 2018 la Comisión Europea emitió una Comunicación[37] relacionada con la IA, en la que proporcionó información sobre la adopción de la IA en la Unión;
c. El Comité de Asuntos Jurídicos de la Unión Europea emitió un Informe con recomendaciones a la Comisión de *Normas de Derecho Civil sobre Robótica*.[38] El Comité era consciente del hecho de que la humanidad está en el umbral de una nueva era donde los robots, los androides sofisticados y otras manifestaciones de IA provocarán una revolución industrial que afectará a toda la sociedad y tendrá implicaciones legales y éticas. En dichas Normas el Comité pidió a la Comisión que abordara los principios generales relativos al desarrollo de la robótica y la IA para uso civil;
d. El Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial de la Comisión Europea publicó un primer borrador de sus directrices éticas propuestas para el desarrollo y uso de IA en junio de 2018. Después de recibir comentarios durante un período de varios meses, emitió un borrador de *Directrices de ética para IA confiable,* el 18 de diciembre de 2018,[39] a saber, (1) aumentar las inversiones públicas y privadas en



[37] European Commission, *Communication*, [Consultado el 10 de diciembre de 2019], Disponible en: https:// ec.europa.eu/info/departments/communication_en
[38] Parlamento Europeo, "Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica" (2015/2103) [Consultado el 10 de diciembre de 2019], Disponible en: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_ES.html
[39] Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo Y al Comité de las Regiones, "Generar confianza en la inteligencia artificial centrada en el ser humano", Bruselas, 8.4.2019

IA para impulsar su aceptación, (2) prepararse para los cambios socioeconómicos y (3) garantizar un marco ético y legal adecuado para fortalecer los valores europeos;

e. La Estrategia de IA de Francia[40] se publicó en 2017. Derivado de dicho documento, el matemático y miembro del Parlamento, Cedric Villani, produjo el denominado *Informe Villani,*[41] en 2018, que describe el desarrollo de una política de datos agresiva; aumenta el potencial de la investigación francesa; apunta a cuatro áreas estratégicas; planes para el impacto de la IA en el trabajo; hace que IA sea más amigable con el medio ambiente; abre las cajas negras de IA; y asegura que la IA respalde la inclusión y la diversidad. Se dio prioridad a la atención de la salud, el medio ambiente, el transporte y la defensa.

En la implementación de estrategias o políticas de ciberseguridad, los Estados deben procurar que éstas garanticen la integridad de la infraestructura y de la información *en línea*, de forma tal que proteja a los usuarios de ataques cibernéticos que afecten los derechos a la intimidad o a la libertad de expresión y los derechos conexos. Y si bien el Estado tiene la obligación de proteger a las personas de que actores maliciosos violen sus derechos, esto no implica que pueda, bajo supuestas razones de "seguridad nacional" violentar a los ciudadanos. Debe existir un equilibrio.

Estas medidas no deben ser tomadas en secrecía, existe un interés público en conocer que se cuenta con una ruta o guía a seguir para la protección de la seguridad nacional, por ello el Estado debe informar y rendir cuentas sobre las medidas tomadas en materia de ciberseguridad para la seguridad nacional, tanto las que son directamente implementadas por éste como las que se encuentran desarrollándose por actores privados.

Lo que no implica que el Estado deba exponer la implementación y forma en que estas medidas funcionan, pero la comunidad necesita conocer que las medidas que se tomen son acordes a la protección de los derechos humanos, no arbitrarias ni violatorias de estos, más importante, que su implementación se realiza también bajo principios éticos, mismos que deben ser exigidos a las



---

COM(2019) 168 final [Consultado el 10 de diciembre de 2019], Disponible en: https://ec.europa.eu/transparency/regdoc/rep/1/2019/ES/COM-2019-168-F1-ES-MAIN-PART-1.PDF

[40] Macron, Emmanuel, "Discours sur líntelligence artificielle au Collège de France" [Consultado el 10 de diciembre de 2019], Disponible en: https://www.campusfrance.org/es/estrategia-sobre-la-inteligencia-artificial-francia-dobla-la-formacion-en-ia

[41] Villani, Cédric, "Rapport de Cédric Villani: donner un sens à l'intelligence artificielle (IA)", [Consultado el 10 de diciembre de 2019], Disponible en: https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf

empresas que desarrollan soluciones en materia de ciberseguridad para la defensa de países. Las actividades de las empresas y su relación con los Estados, en el rubro de las tecnologías y redes digitales presentan crecientes desafíos y pueden generar amenazas para el disfrute de los DDHH.

En el año 2018 se llevó a cabo una audiencia pública convocada por la Comisión Interamericana de Derechos Humanos (CIDH) y su Relatoría Especial para la Libertad de Expresión, en donde varias organizaciones regionales plantearon diversos problemas, uno de los temas fue la preocupación en torno a la vulneración del derecho a la privacidad y a la libertad de expresión a través de la vigilancia digital, software de espionaje, políticas de ciberseguridad, bloqueo de contenidos en línea.

Aunque no se enfocaba directamente en la tecnología basada en IA, dentro de las recomendaciones se señaló que las autoridades deben informar y rendir cuentas sobre las medidas tomadas en materia de ciberseguridad, tanto de aquellas directamente implementadas como de las que ejecutan intermediarios privados contratados por el Estado (párrafo 126).

## 6. Conclusiones

El uso ampliado de la IA y su aprendizaje automático, combinado con el crecimiento del mercado y la disminución de los precios, expandirá en gran medida el impacto a estos sistemas en la seguridad nacional. Por tanto, quien tenga una ventaja sobre la IA podrá ganar la iniciativa en la competencia para ganar el futuro.

Debido a su importancia en el desarrollo económico y militar, el desarrollo de la IA se ha elevado al estado de una estrategia nacional. Estados Unidos, China, Francia, la Unión Europea, todos han desarrollado estrategias para el uso de la IA. En nuestro país aún no existe un rumbo establecido, lo que no sólo nos pone en desventaja económica, sino que al existir una falta de principios y valores en el desarrollo, adquisición e implementación de esta tecnología, estamos a merced de que actores o empresas violenten principios éticos o derechos humanos.



El armamento de IA es inevitable y aunque no se ha logrado alcanzar un consenso para su empleo, tal vez resulte más sencillo acordar valores que políticas para su control. Si consideramos que en la actualidad no hay consenso suficiente en cuanto a la interpretación de si el Derecho Internacional Humanitario en el ciberespacio brinda protección legal para la población civil,[42] el

[42] Droege, Cordula, Chief Legal Officer and Head of the Legal Division, ICRC, Foreword, en ICRC "The Potential Human Cost of Cyber Operations. IRCR Expert Meeting" 14-16 November 2018, Ginebra, p. 3.

empleo de tecnología como la IA, con sus impactos aún desconocidos, agrava la situación y entorno de protección de los individuos.

Desafortunadamente, como la historia nos ha recordado en diversas ocasiones, un lenguaje poderoso en un documento oficial por sí solo no es suficiente para prevenir violaciones sistemáticas de los derechos humanos. De la misma forma que nos ha enseñado los peligros que trae consigo el desarrollo en secreto de armas, de las cuales desconocemos sus impactos reales (bomba atómica). La falta de un marco de gobernanza global para la tecnología corre el riesgo de fragmentar el ciberespacio, lo que podría disuadir el crecimiento económico, agravar las rivalidades geopolíticas y ampliar las divisiones dentro de las sociedades.[43]

*Artificial intelligence-driven profile* (perfiles impulsados por inteligencia artificial). "Info-wars", guerras de información y desinformación; el desarrollo de ataques en conflictos entre países con el uso de drones no tripulados. Estos sofisticados instrumentos de ataque, de acceso civil y militar son empleados con fines de combate, controlados por militares a través de monitores desde bases lejanas de su lugar de ataque. En la actualidad la decisión última para determinar el objetivo empleando esos drones no tripulados sigue siendo humana y en cualquier caso así debe seguir. Sin embargo, se está desarrollando tecnología que, con el empleo de IA, permitirá a los drones la toma de decisiones libre para determinar el objetivo, es decir, sin intervención humana.

Los algoritmos y el aprendizaje de las máquinas se realizan por humanos. Humanos que naturalmente emplean el contexto actual de sus vidas como marco de referencia de lo que debería ser ético, de valores morales, de prejuicios, conocimientos, experiencias y contextos. Todo esto es trasladado al aprendizaje de las máquinas.

En nuestras manos se encuentra que la IA nos posibilite el desarrollo de armas de destrucción masiva totalmente autónomas o nos permitirá transitar a una era de estabilidad y paz internacional.

La oportunidad determina el futuro. El advenimiento de la IA brinda una oportunidad sin precedentes para realizar la protección efectiva contra amenazas. Solo aprovechando firmemente esta oportunidad histórica con un fuerte conocimiento de la estrategia y la oportunidad, realmente haciendo un buen uso de la tecnología de inteligencia artificial, podremos aportar a la ciberseguridad, paz y estabilidad internacional.



[43] Foro Económico Mundial, "Global Risk Report 2020", WEF, Ginebra, p. 20.

## Referencias.

Asamblea General de Naciones Unidas, Preámbulos de las Resoluciones A/RES/55/28 de 20 de Noviembre del año 2000; A/RES/56/19 de 29 de noviembre de 2001; A/RES/59/61 de 3 de diciembre de 2004; A/RES/60/45 de 8de diciembre de 2005; A/RES/61/54 de 6 de Diciembre de 2006; A/RES/62/17 de 05 de diciembre de 2007; A/RES/63/37 de 2 de diciembre de 2008; A/RES/64/25 de 2 de diciembre de 2009.

Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones, "Generar confianza en la inteligencia artificial centrada en el ser humano", Bruselas, 8.4.2019 COM(2019) 168 final [Consultado el 10 de diciembre de 2019], Disponible en: https://ec.europa.eu/transparency/regdoc/rep/1/2019/ES/COM-2019-168-F1-ES-MAIN-PART-1.PDF

Corte Internacional de Justicia, APPLICATION OF THE CONVENTION ON THE PREVENTION AND PUNISHMENT OF THE CRIME OF GENOCIDE, (THE GAMBIA v. MYANMAR), Request for the indication of provisional measures, Procedimiento 23 de enero de 2020 [Consultada el 13 de febrero de 2020], Disponible en:
https://www.icj-cij.org/files/case-related/178/178-20200123-ORD-01-00-EN.pdf

Corte Internacional de Justicia, APPLICATION OF THE CONVENTION ON THE PREVENTION AND PUNISHMENT OF THE CRIME OF GENOCIDE, (THE GAMBIA v. MYANMAR), Request for the indication of provisional measures, Procedimiento 23 de enero de 2020[Consultada el 13 de febrero de 2020], Disponible en: https://www.icj-cij.org/files/case-related/178/178-20200123-ORD-01-00-EN.pdf

DARPA, Collaborative Operations in Denied Environment (CODE), [Consultada el 06 de octubre de 2020], Disponible en: https://www.darpa.mil/program/collaborative-operations-in-denied-environment

Diccionario Real Academia Española [Consultada el 13 de febrero de 2020], Disponible en: https://dle.rae.es/artificial?m=form

Droege, Cordula, "Foreword", en ICRC The Potential Human Cost of Cyber Operations. IRCR Expert Meeting, pp. 14-16 de noviembre 2018, Ginebra.

European Commission, Cybersecurity Our Digital Anchor. A European Perspective, European Comission, Joint Research Centre, Italia.

European Commission, Communication, [Consultada el 13 de febrero de 2020], Disponible en: https:// ec.europa.eu/info/departments/communication_en

Foro Económico Mundial, "Global Risk Report 2020", WEF, 2020, Ginebra [Consultada el 13 de febrero de 2020], Disponible en:http://www3.weforum.org/docs/WEF_Global_Risk_Report_2020.pdf

Gard, Tom, "Google´s new AI is better at Creating AI Than the Company´s Engineers", Futurism, 17 de mayo 2017 [Consultada el 13 de febrero de 2020], Disponible en: https://www.cnet.com/news/its-happening-googles-ai-is-building-more-ais/

High-Level Expert Group on Artificial Intelligence (HLEG), "A definition of AI: Main Capabilities and Disciplines", European Comission, 2019, p. 6 [Consultada 06 de octubre de 2020], Disponible en: https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

International Committee of the Red Cross (ICRC), "A Guide to the Legal Review of Weapons, Means and Methods of Warfare", ICRC, Ginebra, 2006

Klaidman, Daniel, Kill or Capture: The War on Terror and the Soul of the Obama Presidency, Harcourt, Nueva York, 2012.

Lee, Kai-Fu, AI SUPERPOWERS. CHINA, SILICON VALLEY, AND THE NEW WORLD ORDER, Houghton Mifflin Harcourt, 2018.

Macron, Emmanuel, "Discours sur líntelligence artificielle au Collège de France" [Consultado el 10 de diciembre de 2019], Disponible en: https://www.campusfrance.org/es/estrategia-sobre-la-inteligencia-artificial-francia-dobla-la-formacion-en-ia

National Institute of Standards and Technology (NIST), "Dictionary of Algorithms and Data Structures", 1998 [Consultada el 10 de febrero de 2020], Disponible en: https://xlinux.nist.gov/dads/HTML/algorithm.html

Thomson, Nicholas, "Emmanuel Macron Talks to WIRED About France's AI Strategy", Wired, 31 de marzo 2018, [Consultada el 10 de febrero de 2020], Disponible en: https://www.wired.com/story/emmanuel-macron-talks-to-wired-about-frances-ai-strategy/

OCDE, Algorithms and Collusion. Competition policy in the digital age, OCDE, 2017 [Consultada el 10 de febrero de 2020], Disponible en: https://www.oecd.org/daf/competition/Algorithms-and-colllusion-competition-policy-in-the-digital-age.pdf

Parlamento Europeo, "Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica" (2015/2103) [Consultado el 10 de diciembre de 2019], Disponible en: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_ES.html

Persi Paoli. G., Vignard. K., Danks. D, and Meyer. P., Modernizing Arms Control: Exploring responses to the use of AI in military decision-making, UNIDIR, Ginebra, p. 2.

Preámbulos de las Resoluciones A/RES/58/32 de 08 de diciembre de 2003; A/RES/59/61 de 3 de diciembre de 2004; A/RES/60/45 de 8 de diciembre de 2005; A/RES/61/54 de 6 de diciembre de 2006; A/RES/62/17 de 5 de diciembre de 2007; A/RES/63/37 de 2 de diciembre de 2008; A/RES/64/25 de 2 de diciembre de 2009.

Protocolo I adicional a los Convenios de Ginebra de 1949 relativo a la protección de las víctimas de los conflictos armados internacionales, 1977, [Consultada el 06 de octubre de 2020], Disponible en: https://www.icrc.org/es/document/protocolo-i-adicional-convenios-ginebra-1949-proteccion-victimas-conflictos-armados-internacionales-1977#GUERRA

Sesión pública celebrada el jueves 12 de diciembre de 2019, a las 10 a.m., en el Palacio de la Paz, Presidente Yusuf presidiendo, en el caso relativo a la aplicación de la Convención para prevenir y sancionar el delito de genocidio (Gambia c. Myanmar) [Consultado el 10 de diciembre de 2019], Disponible en: https://www.icj-cij.org/files/case-related/178/178-20191212-ORA-01-00-BI.pdf

Sifton, John, *A Brief History of Drones*, The Nation, 7 de febrero de 2012, [Consultado el 10 de diciembre de 2019], Disponible en: http://www.thenation.com/article/166124/brief-history-drones#

UN HRC, "Report of the detailed findings of the Independent International Fact-Finding Mission on Myanmar", UN doc. A/HRC/39/CRP.2, 17 de septiembre de 2018, [Consultado el 10 de diciembre de 2019], Disponible en: https://www.ohchr.org/Documents/HRBodies/HRCouncil/FFM-Myanmar/A_HRC_39_CRP.2.pdf

United Nations Institute for Disarmament Research (UNIDIR), Algorithmic Bias and the Weaponization of Increasingly Autonomous Technologies. A Primer, UNIDIR, 2018.

United Nations Institute for Disarmament Research (UNIDIR), Swarm Robotics: Technical and Operational Overview of the Next Generation of Autonomous Systems, UNIDIR, *2020,* [Consultada el 06 de octubre de 2020], Disponible en: https://unidir.org/publication/swarm-robotics-technical-and-operational-overview-next-generation-autonomous-systems

Villani, Cédric, "Rapport de Cédric Villani: donner un sens à l'intelligence artificielle (IA)", [Consultado el 10 de diciembre de 2019], Disponible en: https://www.ai-forhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf

Wang, Yilun, Kosinski, Michal, "Deep neural networks are more accurate than humans at detecting sexual orientation from facial image", Journal of Personality and Social Psychology, febrero 2018, Vol. 114, Issue 2, pp. 246-257. [Consultado el 10 de diciembre de 2019], Disponible en: https://www.gsb.stanford.edu/faculty-research/publications/deep-neural-networks-are-more-accurate-humans-detecting-sexual

Will Roper en McCullough, Amy, "The Looming Swarm", 22 de marzo 2019, [Consultada 06 de octubre de 2020], Disponible en: https://www.airforcemag.com/article/the-looming-swarm/