Documento de Trabajo 04/2019

Usos militares de la inteligencia artificial, la automatización y la robótica (IAA&R)

Artificial intelligence, automation and robotics (AIA&R) in the military.

Trabajo incluido en el Plan Anual de Investigación del Centro Superior de Estudios de la Defensa Nacional (CESEDEN) para el año 2019, como Grupo de Trabajo de Cooperación Nacional nº 1, asignado al Centro Conjunto de Desarrollo de Conceptos (CCDC)

*

Organismo solicitante del estudio:
Centro Superior de Estudios de la Defensa Nacional (CESEDEN)

Usos militares de la inteligencia artificial, la automatización y la robótica (IAA&R)

Artificial intelligence, automation and robotics (AIA&R) in the military

Maquetado en octubre de 2019 por el Instituto Español de Estudios Estratégicos (IEEE)

Centro Superior de Estudios de la Defensa Nacional (CESEDEN)

Nota: Las ideas y opiniones contenidas en este documento son de responsabilidad de los autores, sin que reflejen, necesariamente, el pensamiento del Ministerio de Defensa, del CESEDEN o del IEEE.

Índice

Introducción

Ángel Gómez de Ágreda

Usos militares de la inteligencia artificial, la automatización y la robótica (IAA&R)

Hasta muy recientemente, el arte de la guerra era la manifestación más sofisticada y compleja de la gestión humana. Nada movía tantos recursos, tantas voluntades ni tantos intereses como cuando la política dejaba de poder ejercerse por medios pacíficos y entraban en juego los generales y sus estados mayores.

Más allá del valor individual, de las tácticas de combate y de las estrategias militares, la guerra en su concepción más científica ha sido siempre un enorme desempeño logístico. La capacidad para generar, sostener y emplear los medios necesarios para vencer ha sido determinante a lo largo de la historia bélica.

Se atribuye a Napoleón la reflexión de que «los amateurs hablan sobre táctica mientras los profesionales estudian logística», y a Sun Tzu la que dice que «la línea entre el orden y el desorden reside en la logística». El estratega chino también dejaba para la posterioridad su clasificación de los factores de la guerra: primero, los cálculos; segundo, las cantidades; tercero, la logística; cuarto, el equilibrio de poder; y quinto, las posibilidades de victoria.

Datos. Saber de qué dispongo; qué puedo generar; dónde, cómo y cuándo puedo emplearlo, y con qué efecto. Conocer lo mismo sobre el enemigo. Y, finalmente, tener la capacidad para llevar a la práctica aquello que es posible manejando los invisibles hilos de las comunicaciones. Las guerras se han ganado casi siempre antes de empezar las batallas o, al menos, se han perdido irremediablemente antes de causar o sufrir la primera víctima.

Cuando la guerra tenía lugar en los campos de batalla, el Comandante necesitaba tener una visión integral de sus fuerzas y las adversarias, de la situación en el combate y de la evolución de la misma. El genio militar permitía a Alejandro, a Aníbal o a César revertir situaciones adversas, pero una parte fundamental de la victoria se había obtenido desplegando en las llanuras de Gaugamela, atravesando los Alpes o construyendo puentes sobre el Rin o murallas alrededor de Alesia.

Ahora, la guerra ya no ocurre tanto en el campo de batalla como entre la gente. Insurgencias y grupos terroristas aprovechan tácticas asimétricas para generar efectos que requieren de una logística mucho menos sofisticada. Se alimentan de la propia población enemiga y de sus recursos, y basan su fortaleza en la explotación de los

principios y valores de la sociedad a la que atacan.

Más allá incluso de esas tácticas, las nuevas tecnologías traen la guerra al interior de cada uno de nosotros. La «guerra en la gente» explota sentimientos más que razones, afectos más que efectos. Es una guerra basada en el conocimiento íntimo de cada adversario y de sus reacciones. Una guerra en la que los datos ya no se limitan a los combatientes uniformados. La opinión pública y la opinión publicada, la influencia y la reputación, las operaciones psicológicas en fin, adquieren una importancia fundamental en el manejo de la voluntad de propios y ajenos.

En un mundo en el que la biosfera física cohabita con la biosfera lógica basada en el ciberespacio, la complejidad y la niebla de la guerra de Clausewitz se tornan más espesas que nunca. No se dejan de añadir dimensiones a controlar, entornos que vigilar, amenazas, riesgos y, no deben olvidarse, oportunidades que explotar. El ciudadano moderno se ha habituado a contar con multitud de datos antes de la toma de cualquier decisión y esa costumbre ha generado una alta dependencia respecto de ellos.

Datos. Igual que el petróleo en su momento, que la electricidad en nuestras ciudades, que el dinero, los datos se han convertido en la materia prima sobre la que construir nuestra capacidad militar. Datos con mayúsculas, Big Data, miles de millones de datos individuales, irrelevantes por sí mismos pero que dibujan, como en un cuadro impresionista, la imagen de conjunto del campo de batalla global del siglo XXI. Datos alineados y ordenados, estructurados para que cobren sentido.

La profesora Inmaculada Mohíno introduce en su capítulo los modos en que se pueden recopilar esos datos. Lo hace centrándose en el campo de batalla, pero su aproximación es extrapolable a la gestión del día a día de los ejércitos. La guerra contemporánea es ubicua y permanente, y la gestión de los ejércitos también tiene que serlo. Lo que ocurra en el frente dependerá de cómo se haya organizado la retaguardia y los datos deberán poder transferirse entre unas funciones y otras.

La profesora de la Universidad de Alcalá nos adentra en el mundo de los sensores capaces de recolectar parámetros del combatiente individual, de las unidades en su conjunto; de las plataformas terrestres, marítimas y aéreas -tripuladas o no-; del medio ambiente y del enemigo. Datos de parámetros físicos, cognitivos y emocionales que permiten conocer las capacidades concretas de cada soldado en función de todas las circunstancias que le rodean.

Datos que se combinan mediante una serie de reglas y técnicas para trascender su valor individual y presentar al decisor una imagen comprensible y comprehensiva. El campo de batalla se transforma en un juego de mesa en el que todas las piezas están a la vista mostrando el valor que tiene cada una y su capacidad remanente. Para el combatiente individual, los datos que incorpora él mismo al sistema se combinan también con los de multitud de otros sensores para permitirle ver mucho más allá del horizonte físico y del horizonte lógico al que podría tener acceso por sí mismo.

Los vientos de la tecnología permiten ya levantar las nieblas guerreras. Sin embargo, incorporan nuevas vulnerabilidades en forma de la dependencia que generan, del

insaciable apetito de más y mayor granularidad en la información, de la búsqueda de la certeza absoluta –la «parálisis por el análisis» –, y de las posibilidades de injerencia del adversario en nuestro sistema de gestión de datos.

Si el enemigo tiene acceso a nuestros datos, si somos incapaces de proteger la confidencialidad de los mismos, no estaremos perdiendo solo la ventaja de ver sin ser vistos, sino que estaremos propiciando esa misma prerrogativa a nuestro rival. Si sus capacidades llegan más allá, podría estar en condiciones de visualizar nuestras bases de datos y de negarnos a nosotros el acceso a nuestro propio sistema. Incluso, en un ejercicio de desinformación, podría alterar los datos en los cuáles basamos nuestros análisis para que la información que recibamos no sea veraz y, por lo tanto, genere decisiones incorrectas. La niebla se espesaría hasta convertirse en ceguera.

A la capacidad para recopilar, procesar y distribuir la información tenemos, pues, que sumar la de defenderla de ataques adversarios y la de operar cuando todo ese flujo de conocimiento desaparece o se ve limitado. El diseño de un sistema resiliente debe incluir la forma de minimizar la pérdida de operatividad en caso de intrusión o de falta de acceso a los datos.

El combatiente y cada plataforma se convierten así en agentes de generación de información que se transmite a un procesador central para su tratamiento. Y, al mismo tiempo, también se transforman en consumidores del conocimiento generado por ese proceso de integración de información. El flujo de ida y de vuelta es constante y tiene en cuenta miles de millones de inputs individuales y docenas de técnicas para su combinación.

Todo ello se produce de forma automatizada, casi sin participación activa por parte del soldado. Cámaras, micrófonos, tejidos inteligentes, electrocardiógrafos, termómetros, pulsímetros y cientos de pequeños aparatos recopilan y transmiten mucha más información de la que su portador llega a apreciar y, por supuesto, de la que sería capaz de comunicar. La inteligencia artificial que analiza sus datos sabe mejor que él el estado en que se encuentra.

Un estado que no es solamente físico, que no incluye solo su ubicación geográfica o la munición remanente, sino que interpreta también su forma de moverse, sus constantes vitales y su entorno para dibujar también un mapa con su estado de ánimo y su capacidad real para ejecutar cualquier acción del mismo modo que ya ocurre con los ciclistas profesionales.

Esa misma inteligencia artificial permite ya el manejo de determinadas funciones de las máquinas a través de órdenes generadas desde los impulsos cerebrales. No es ciencia-ficción. Los primeros desarrollos están ya documentados y su aplicación en el campo de batalla no tardará en llegar. Igual que, con solo pensarlo, nuestro cerebro da instrucciones a nuestros brazos o piernas, en un futuro próximo podrá hacer lo mismo con un exoesqueleto que proteja al soldado de agresiones y accidentes al tiempo que potencia la fuerza y resistencia de sus articulaciones. Proyectos como el «combatiente del futuro» o versiones noveladas de humanos con capacidades físicas incrementadas como los excelentes relatos de Dale Brown hace ya tiempo que imaginan o diseñan

estos equipos.

La profesora Mohíno nos habla también de las conexiones que se pueden establecer directamente con el cerebro humano, en las que está avanzando de forma espectacular la neurociencia. La realidad virtual y la realidad aumentada permitirán entornos de combate y de entrenamiento mucho más eficientes que los actuales.

La tendencia actual apunta también a una mayor autonomía de las plataformas militares. Drones y todo tipo de vehículos adoptan tecnologías que, en muchas ocasiones, son de uso dual cívico-militar para moverse o para operar, o para aliviar en buena medida la carga de trabajo del operador humano.

La miniaturización y la actuación en forma de enjambres también se abren paso y encuentran su nicho de actividad. Los nano-drones que describiera Michael Crichton en Presa o micro-drones como el que aparece en la película El ojo en el cielo ofrecen posibilidades increíbles y una resiliencia difícilmente replicable por parte de sistemas de armas más grandes.

La tarea fundamental del combatiente en todo tiempo pasa a ser la de proveer de datos al sistema para que éste pueda correlacionarlos y ofrecerle al Comandante una imagen global del campo de batalla. Luego, ese mismo combatiente, con sus datos y muchos más, se convertirá en un vector óptimo para ejecutar las instrucciones del Mando. Unas órdenes que se acomodarán, además, a las posibilidades reales medidas por la integración de toda la información disponible.

Esa integración de los datos en una imagen que sirva al Comandante para la gestión de sus capacidades –tanto en tiempo de guerra como en la operativa diaria de los ejércitos– es el tema que trata la profesora Rocío Barragán en su capítulo sobre la Common Operational Picture (COP), la imagen operacional común que visualiza millones de inputs de manera que la decisión sea más sencilla e intuitiva de adoptar.

La profesora Barragán, actualmente en Eurocontrol, aprovecha su experiencia en la Escuela Técnica Superior de Ingeniería Aeronáutica y del Espacio (ETSIAE) de la Universidad Politécnica de Madrid (UPM) para dar una visión de las distintas técnicas de machine learning, el aprendizaje y auto-aprendizaje de las máquinas, y la minería de datos a la hora de gestionar la información proporcionada por los sensores.

El proceso descrito más arriba de utilización de los datos obtenidos por los sensores tiene lugar en esta fase. No necesariamente sobre la integración simple de los datos, sino posibilitando el aprendizaje profundo de las máquinas para la generación de escenarios previsibles y opciones de actuación para el Mando. Las posibilidades abarcan desde lo doctrinal, estratégico y geopolítico hasta lo táctico y lo técnico.

Aunque la decisión permanezca en manos del Comandante, el camino que lleva hasta ella está cada vez más pavimentado por algoritmos y procesos en los que la inteligencia artificial proporciona los argumentos en que basarla. Se libera al decisor del proceso de integración de los datos y se le presenta una versión simplificada de la realidad que integra muchos más elementos de los que jamás habría podido llegar a

tener en cuenta manualmente.

Esta integración de los datos no es sencilla. Requiere de la capacidad para entender señales muy distintas entre sí procedentes de plataformas o sensores de múltiples orígenes. Implica la ponderación de la importancia de cada uno de esos datos y el modo en que se relaciona con los demás, la exclusión de los irrelevantes y la evitación de sesgos en su inclusión o peso específico.

Eso sí, una vez puesto en marcha el proceso, todas las fases de la toma de decisión se ven apoyadas con juicios mucho más fundados que las intuiciones y las genialidades del Comandante. Apoyadas, que no condicionadas. O ese debería ser, al menos, el límite establecido, como veremos más abajo.

El planeamiento de la operación, la preparación de las fuerzas y la ejecución de la misión contarán con métricas precisas para maximizar la eficacia de cada acción. Todo ello en un entorno que se reevalúa de forma constante para mantenerse actualizado en todo momento en función de la evolución de la situación y de las acciones del enemigo.

Ningún plan del futuro resistirá tampoco el primer disparo, como apuntaba el Mariscal Moltke o, de forma más prosaica pero muy gráfica, el campeón de los pesos pesados de boxeo, Mike Tyson: «todo el mundo tiene un plan hasta que te dan un puñetazo en la boca». Sin embargo, la capacidad para reevaluar la situación y actualizar los planes permitirá una flexibilidad mucho mayor.

También supondrá un importante riesgo de escalada de las hostilidades si las decisiones se automatizan suficientemente. La inteligencia artificial no es siempre la mejor solución a un problema y la delegación de las decisiones conlleva riesgos propios. El grado de autonomía aceptable en cada situación deberá ser evaluado del mismo modo que lo es en el caso humano a través de las reglas de enfrentamiento (RoE, en inglés). La responsabilidad es algo que, por otro lado, no puede delegarse y que siempre se exigirá a los agentes de voluntad. Esos agentes deberían ser siempre humanos.

El capítulo de Rocío Barragán expone someramente las distintas técnicas que están ya disponibles para obtener esta visión integral del escenario. No solo en el campo de batalla, sino en la gestión diaria de las actividades de cualquier unidad militar.

La clave es la disponibilidad de los datos sobre los que trabajar y el dominio de las distintas técnicas para la elaboración de información relevante basada en ellos. Si el resultado obtenido es la mejor línea de acción para acometer al enemigo, o la forma más eficiente de transportar suministros a una operación, o la gestión de los turnos de guardia de una tripulación dependerá solo de la forma en que empleemos esa base de datos inicial.

Al final, se trata nada más ni nada menos que de explotar las «7 uves»: el Volumen de datos Válidos, Variados y Veraces, obtenidos y procesados Velozmente para extraer un Valor añadido de su Visualización. El valor del Big Data reside en la simplificación

de esos miles de millones de datos dispersos en una visualización intuitiva.

El capitán de navío Enrique Cubeiro, Jefe de Estado Mayor del Mando Conjunto de Ciberdefensa (MCCD) y su anterior Jefe de Operaciones, nos lleva, no obstante, más a la mitigación de los riesgos que, sobre esos datos, pueden darse desde el ciberespacio que a la explotación de las ventajas que pueden obtenerse de los mismos. Cubeiro diseccciona las distintas amenazas que podrían poner en jaque nuestra capacidad para utilizar los datos obtenidos, o manipularlos para hacernos llegar a conclusiones erróneas.

En un entorno inteligente, ya sea el humano o el de las máquinas, lo que se pretende es afectar el ciclo de decisión adversario o acelerar el propio. Para eso nace Internet y esa ha sido la obsesión constante de las instituciones en su utilización. La mejor herramienta no puede, no debe, ser susceptible de quedar inutilizada por una acción hostil porque supondría la pérdida de la iniciativa que está en la base de la libertad de acción del Mando.

De este modo, la ciberseguridad tiene que dar protección a las soluciones de IA al mismo tiempo que estas mismas mejoran las capacidades de la ciberseguridad para hacerlo. Una relación simbiótica. Si el mundo de la IA se apoya en las redes cibernéticas, su protección debe ser una prioridad de propia supervivencia. Por mucho que el objetivo final no sea sobrevivir como herramienta sino retener la habilidad de prestar un servicio.

La capacidad para identificar patrones, para deducir tendencias, para encontrar vulnerabilidades, para reaccionar rápidamente ante un suceso imprevisto y para tantas otras funciones resulta crucial para mantener un elevado nivel de ciberseguridad sobre las mismas redes a través de las cuáles discurren los datos que alimentan a esa IA.

En ese sentido, la inteligencia artificial no deja de ser otro eslabón más en la cadena competitiva entre ataques y defensas. Sus aplicaciones serán utilizadas tanto en identificar vulnerabilidades para su explotación como para identificarlas con el fin de defenderlas. En ambos casos, conseguir unos algoritmos altamente eficaces será fundamental.

La ciberseguridad, la protección de las propias redes que permiten generar soluciones de inteligencia artificial, será una de las funciones principales y más urgentes que se deben acometer. Sin embargo, el entorno operativo es mucho más amplio. Llega incluso más allá del mismo ámbito lógico cuando se mezcla con el de la información y el de los afectos. Entramos ahí en una «zona gris»[1] en la que abundan los matices, los tonos

1 «Zona del espectro de los conflictos donde predominan las actuaciones situadas al margen del principio de buena fe entre estados (bona fide) que pese a alterar notablemente la paz no cruzan los umbrales que permitirían o exigirían una respuesta armada.» Estado Mayor de la Defensa, CCDC, «PDC-01 Doctrina para el Empleo de las FAS», disponible en: http://www.defensa.gob.es/ceseden/Galerias/ccdc/documentos/02_PDC-01_xAx_Doctrina_empleo_FAS.pdf, fecha de la consulta 03.09.2019.

apenas distinguibles. Un mundo monocromo en los márgenes de lo reglamentado que nos describe el teniente coronel Marín en su capítulo.

Esa guerra en el filo de la legalidad, de la aceptación internacional, es un conflicto híbrido en el que las armas no tienen por qué tener ese nombre, en el que pueden ser meras herramientas, procedimientos o declaraciones.

En el siguiente capítulo se abordan las múltiples manifestaciones de esta forma de confrontación. En ella, las herramientas asimétricas dejan de ser una alternativa utilizada por el bando que se considera menos fuerte para ser utilizadas indistintamente por ambas partes. La clave es la gestión de la complejidad de todos los factores implicados y el aprovechamiento de cualquier resquicio para debilitar la voluntad del adversario. Una «muerte de los mil cortes» en la que ninguna de las agresiones sobrepasa el umbral de la agresión merecedora de una respuesta armada, pero en la que cada acción socava la resiliencia del otro.

Al igual que sucede en la metáfora de la rana en agua hirviendo, las amenazas híbridas evolucionan de forma sigilosa y continua hasta que consiguen su objetivo. Incluyen las sanciones económicas y los actos terroristas, pero también la utilización de la tecnología para alterar las percepciones y los relatos. La guerra cibernética, con su discreción y dificultad para atribuir autorías a los ataques es una de las señas de identidad.

Pero la inteligencia artificial está permitiendo novedosas formas de alcanzar los mismos efectos que antes requerían de una intervención armada. Los tratamientos automatizados de audio y video empiezan a estar presentes en los enfrentamientos entre naciones. Al menos, la sospecha de falsificaciones profundas –los deep-fakes- aparecen en las alegaciones sobre el origen de conflictos como el que mantiene dividido al Consejo de Cooperación del Golfo.

Estas herramientas no suelen generar brechas, sino explotar las ya existentes magnificando sus efectos, polarizando las opiniones y exacerbando los ánimos de los contendientes.

Los sistemas de armas están ganando en autonomía, en el frente y en la retaguardia, pero están creciendo, principalmente, en lo que respecta a la conectividad. El capítulo presenta algunos ejemplos de armamento más o menos autónomo y conectado en uso en la actualidad, pero sobre todo se centra en adivinar tendencias y aplicaciones futuras de la inteligencia artificial.

Igual que los drones están mejorando su capacidad para ejecutar acciones en un modo «enjambre» coordinando sus movimientos en un baile coreografiado previamente o improvisado a modo de bandada de estorninos, el campo de batalla del futuro dependerá más de la capacidad de las armas para actuar como un solo elemento que del poder individual de cualquiera de ellas. Es la integración de los efectos lo que producirá realmente un resultado deseado.

Los sensores de que hablaba la profesora Mohino, los datos que proporcionan y su integración según describe la profesora Barragán, protegidos como describe el capitán de navío Cubeiro permitirán utilizar los sistemas de armas autónomos en un entorno como el que nos describe el teniente coronel Marín.

No es una guerra nueva, ni es más o menos violenta por el hecho de que las máquinas la ejecuten o ayuden a hacerlo. La guerra es un fenómeno humano y son los hombres los que la llevan a cabo. Tampoco es un mero cambio de herramientas, de armas con las que pelear. La guerra es un acto político y social que muta con la política y la sociedad que la llevan a término. La tecnología permite adaptar los instrumentos a las necesidades de la guerra, pero también condiciona el entorno social que da lugar a la misma. Entender la guerra implica entender a la sociedad que la hace. La inteligencia artificial cambia ambas y lo hace a un ritmo exponencial.

Capitulo 1

De las células a los bits

Inmaculada Mohíno Herranz

Resumen

La monitorización a través de sensores en el soldado y en el campo de batalla es algo que se hace necesario en la actualidad. La Inteligencia Artificial (IA) toma fuerza a cada momento, y gracias a ello se puede extraer valor de los datos. Los datos existen hace muchos años y están almacenados en grandes servidores esperando que alguien les dé el valor que potencialmente poseen.

En el campo de batalla pueden ser utilizados diversos sensores registrando información de la posición, el estado físico, emocional y mental, así como las capacidades de un soldado en tiempo real para hacer frente a diferentes situaciones. Con respecto al enemigo pueden reconocer sus armas, estrategias y estudiar su ejecución para contraatacar o defender de un modo más eficiente. La información proporcionada por los diferentes sensores puede ser utilizada de manera individual o combinada con los datos proporcionados por otros sensores, de modo que se pueden crear sistemas con una inteligencia y capacidad muy superior a la humana.

En el presente capítulo pretenden exponer algunos elementos que harán realidad aspectos que hasta ahora creemos ciencia ficción.

Palabras clave:

Inteligencia Artificial, sensores, aprendizaje automático, algoritmos, toma de decisiones.

From cells to bits

Abstract

The monitorization using sensors in the soldier and in the battlefield is necessary nowadays. Artificial Intelligence gains strength at every moment, so, value can be extracted from the data. Data exist many years ago and they are stored in large servers waiting for someone to give them the value they potentially have.

In the battlefield, several sensors can be used, capturing information about the position, physical, emotional and mental state, as well as the capabilities of a soldier in real time to deal with different situations. With respect to the enemy, they can recognize their weapons, strategies and study their execution, to counterattack or defend in a more efficient way. The information provided by the different sensors can be used by itself or be combined with data provided by other sensors. This way, systems with higher capabilities than humans could be created.

The present chapter presents some elements that will make reality aspects that until now we believe science fiction.

Keywords:

Artificial Intelligence, sensors, machine learning, algorithms, decision-making.

Introducción

En la literatura se encuentran numerosas maneras de definir Inteligencia Artificial (IA). Una de ellas podría ser: «*el estudio de las computaciones que hacen posible percibir, razonar y actuar*». Entre las definiciones, una conclusión interesante es que la IA pretende imitar al ser humano. El ser humano aprende con datos como imágenes, sonidos, comportamientos de otros, olores, etc. Además de datos existen reglas, así como técnicas aprendidas en casa, en el colegio y en la vida. Ausubel describe cómo aprenden las personas con su teoría del aprendizaje significativo. La construcción de nuevos conocimientos tiene lugar gracias a la observación y al registro de acontecimientos que se relacionan con los conocimientos previos[2].

«Machine learning» (ML) en castellano «aprendizaje automático», es el término que engloba las distintas técnicas existentes para que las máquinas adquieran la capacidad de aprender de los datos. Las máquinas, al igual que los humanos precisan de datos, técnicas y métodos para aprender. Una vez que las máquinas aprenden son puestas a prueba con datos nunca vistos y clasifican, predicen o reconocen patrones con unos ciertos valores de precisión, exactitud, sensibilidad y especificidad.

Las máquinas pueden aprender con cierta probabilidad casi cualquier cosa, siempre y cuando existan características significativamente relevantes y grandes cantidades de datos de ingesta al sistema. El mismo se compone de ciertas fases cuando hablamos de machine learning. A continuación se expone un típico diagrama de bloques que representan dichas fases.

Figura 1: Diagrama de bloques del sistema de aprendizaje

El primer bloque denominado ***señales de entrada***, representa la entrada que pueden ser las señales registradas por los sensores. A continuación, la etapa de ***extracción de características***, en la cual se extraen unas u otras características en función de la señal o los datos a estudiar. Por ejemplo, si se va a utilizar una señal fisiológica como el electrocardiograma, algunas características podrían ser las pulsaciones por minuto o la amplitud de la onda Q del complejo QRS, entre muchas otras.

2 Ausubel, D., Novak, J., & Hanesian, H. «Educational Psychology: A Cognitive View (2nd Ed.)», Holt, Rinehart & Winston, New York, 1978.

Una vez que se tienen las características, sería conveniente, estudiar cuales son «buenas» o no, en función del problema a tratar. Es decir, en función de qué se pretenda predecir, unas características serán más significativas que otras en el aprendizaje. Por esta razón se pueden utilizar técnicas de ***selección de características***. Existen numerosos tipos de algoritmos en la selección de características, algunos como los conocidos «algoritmos evolutivos o algoritmos genéticos»[3]. Estos algoritmos están basados en la técnica de la evolución biológica en la naturaleza, que combinan los principios de supervivencia del individuo más apropiado mediante un intercambio de características entre individuos de una población de posibles soluciones, de manera que se constituye un procedimiento de búsqueda capaz de aplicarse a cuestiones de optimización en diferentes áreas. Dichos algoritmos genéticos trabajan con una población de varios individuos, es decir, varias soluciones. Cada individuo representa una solución y se denomina «cromosoma»[4]. En el proceso de evolución existen varias fases que son: generación de la población inicial, evaluación de la función de adaptación, selección de los individuos a reproducir, reproducción, mutación e inserción de los hijos a la población. Estas fases se repiten hasta que se cumple la condición de finalización.

Lo que se pretende con este algoritmo es obtener el individuo más apropiado, es decir, las características más beneficiosas en la tarea a resolver. Existen numerosas técnicas de selección, por lo que lo conveniente es utilizar varias para contrastar los resultados.

Una vez que las mejores o más apropiadas características han sido seleccionadas, ya es posible entrenar el sistema con el/los clasificador/es seleccionado/s. Por tanto la última fase ***clasificación/regresión*** es la de clasificar si queremos obtener clases o se hará regresión si lo deseado es obtener un número.

En la literatura se puede encontrar un número de clasificadores muy elevado, por lo que lo apropiado será testear varios clasificadores de diversa índole. Éstos delimitan las clases con fronteras, las cuales serán diferentes en función del clasificador utilizado. La elección del clasificador es una parte fundamental en el aprendizaje automático, ya que afecta directamente a la eficiencia del sistema. Para encontrar el clasificador más apropiado es conveniente buscar un compromiso entre precisión, exactitud, sensibilidad y especifidad.

3 Holland, J. H. «Adaptation in natural and artificial systems», University of Michigan Press, 1975.

4 GOLDBERG, D. E. «Genetic algorithms – in search, optimization and machine learning.» Addison-Wesley Publising Company, Inc., 1989. ISBN: 9780201157673.

Utilización de sensores en el campo de batalla

La Figura 2 representa de forma esquemática los bloques en los que se puede utilizar la sensorización y la inteligencia artificial en el campo de batalla.

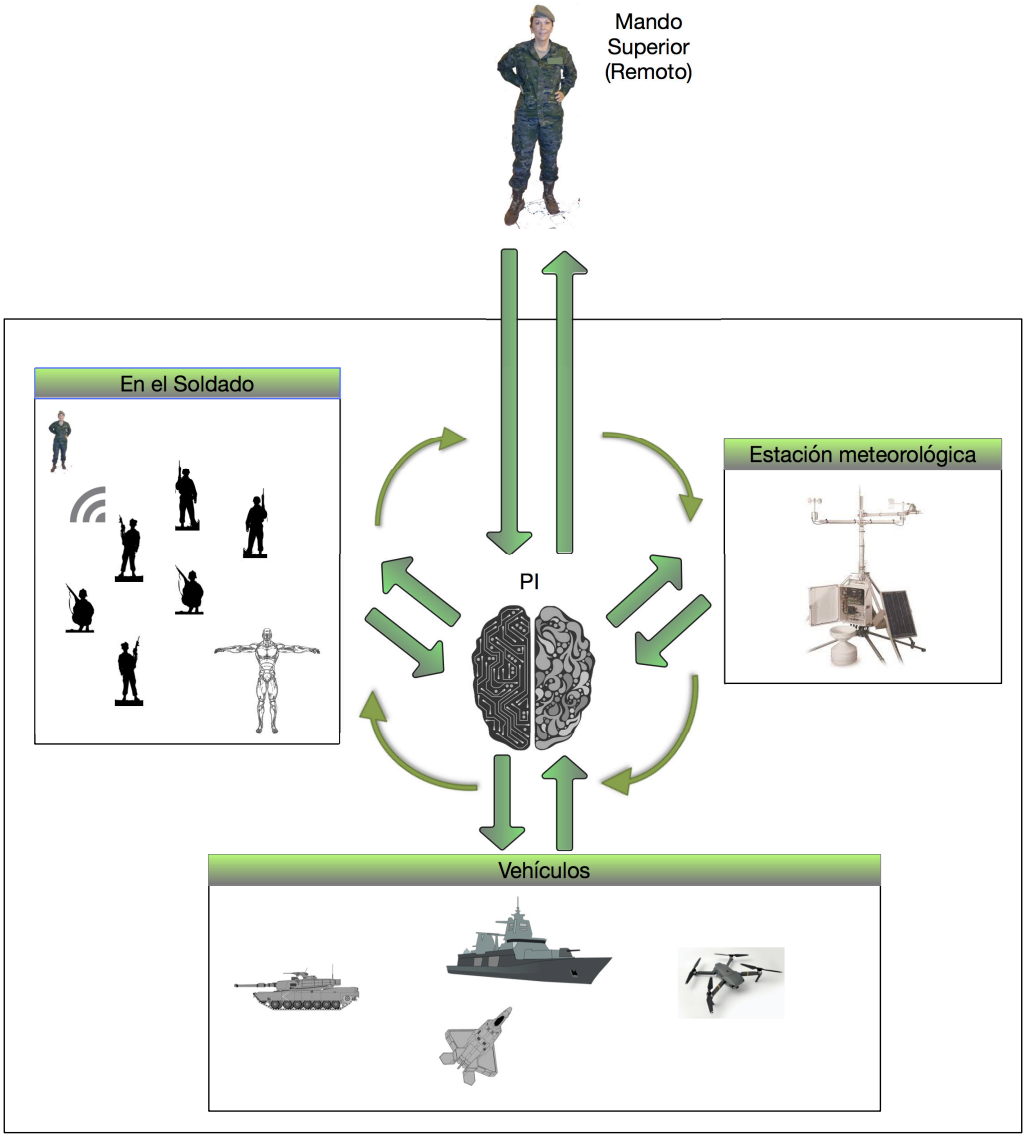


Figura 2: Utilización de sensores en el campo de batalla

Una parte básica para extraer el máximo potencial de la inteligencia artificial de los sistemas expuestos es la **interconexión**. Es fundamental que exista un intercambio de información continuo entre los sistemas para que cada uno de ellos pueda reaccionar lo antes posible ante amenazas inminentes o futuras. Para ello, será necesario que los protocolos de intercambio de información sean suficientemente robustos y seguros, para que no haya pérdida de información así como eviten intrusiones enemigas.

Otra parte importante es el sistema que gestione la información. Para ello se puede utilizar un ***«Procesador de Información»*** (PI) cuya función es la de procesar y gestionar la información pertinente en cada momento. Dicho procesador puede ser el encargado de predecir numerosos valores en función de las entradas individuales o combinadas. Como se observa en la imagen se va a realizar una visión general de los principales sistemas utilizados en el campo de batalla y la inteligencia que poseen y se prevé poseerán. La inteligencia artificial añadida a cada uno de los sistemas expuestos permitirá facilitar la toma de decisiones, así como la asistencia humanitaria, la supervivencia y seguridad de los soldados entre diversas posibilidades que ofrece la nueva era tecnológica.

Además de la interconexión y el PI, será necesario sensorizar al soldado, vehículos y sistemas de captación de información del entorno, como una estación meteorológica. La disposición de numerosos sensores en el soldado para captar información relevante del estado físico y mental de cada uno de ellos, del ambiente, como es el audio y la imagen, y la utilización del exoesqueleto entre otros aspectos que serán tratados más adelante. Por otro lado, los sensores conectados en los distintos vehículos permitirán obtener información relevante con respecto al estado de cada uno, situación, y estudio del enemigo. Finalmente, se utilizará una estación meteorológica portátil, que proporciona una importante cantidad de información sobre el clima y permite formar una escena futura con respecto al mismo.

Una vez que la cantidad ingente de información recogida de todos y cada uno de los sensores es procesada, se hace fusión de información que dotará de valor añadido a los datos recopilados y proporcionará unas claras pautas que la autoridad pertinente utilizará en la toma de decisiones.

Por lo tanto, la información ofrecida por la inteligencia artificial conforma «una ayuda en la toma de decisiones». Aunque, con una alta probabilidad, los sistemas serán autónomos y en muchas ocasiones tomarán sus propias decisiones, estarán siempre basados en las líneas de código escritas por un humano, en función de lo que la ética dicte[5].

En el soldado

En esta sección se van a exponer las diferentes partes necesarias para comprender los datos obtenidos a partir de la monitorización a través de sensores en el soldado. Esta

5 GROTH, Olaf, NITZBERG, Mark y ESPOSITO, Mark, WRules for Robots. Why We Need a Digital Magna Carta for the Age of Intelligent Machines.» International Reports, 2018. Disponible en: https://www.kas.de/documents/252038/253252/7_dokument_dok_pdf_52115_2.pdf/1a5564f5-77ea-a5bc-228b-279d885c313b?version=1.0&t=1539647628555, fecha de la consulta 02.02.2019.

sección se divide en varias partes, las señales fisiológicas en el soldado, los elementos necesarios para registrar dichas señales, como son los electrodos y los dispositivos de registro y la importancia del registro de dichas señales. También puede ser mejorada la eficiencia del soldado utilizando la neurotecnología, que será descrita en adelante. Finalmente, se comenta el uso de exoesqueletos y de sistemas de localización.

El cuerpo humano emite información de manera constante, consciente e inconsciente todo el tiempo. La recolección de dicha información se lleva a cabo a través de numerosos sensores dispuestos sobre del cuerpo.

A lo largo de esta sección, se va a tratar de exponer que tipo de información puede ser registrada mediante diversos sensores.

Monitorización del cuerpo humano

En el soldado, pueden ser obtenidas numerosas señales fisiológicas que contienen información muy relevante en cuanto el estado físico, mental y emocional del soldado.

Las señales que son tratadas en el presente documento se obtienen mediante medidas no invasivas. Algunas de las más destacables pueden ser:

- Electrocardiograma, representando la señal eléctrica del corazón.
- Impedancia Torácica, conductividad eléctrica en el tórax.
- Actividad eléctrica de la piel.
- Señales procedentes del cerebro o registro cerebral:
    - Electroencefalograma, señal eléctrica del cerebro.
    - Potenciales evocados, respuestas del cerebro provocadas por estímulos sensoriales.
- Electromiograma, señal eléctrica muscular.
- Temperatura del cuerpo y de la piel.
- Voz.

La **señal de voz** es una herramienta con potencial suficiente como para ayudar a la identificación y verificación del usuario, así como la capacidad de comandar el control de vehículos, armas, e incluso exoesqueletos, entre otras posibilidades. En la actualidad, existen algoritmos entrenados en reconocer, identificar y verificar al hablante. Es necesario mejorar y ampliar dichas funcionalidades y crear nuevas el futuro próximo.

Por otro lado, en cuanto al tratamiento de **otras señales fisiológicas**, es bien sabido que es posible conocer el estado físico de un humano. Mediante la señal de electrocardiograma se puede obtener información del estado del corazón del humano, sí está o no sano, además de numerosas patologías. El efecto de la carga emocional, mental y física en las señales fisiológicas es un tema que suscita mucha curiosidad, y numerosos científicos han estudiado sus relaciones. El sistema límbico está compuesto por un conjunto de complejas estructuras dispuestas encima y alrededor del tálamo y otras estructuras subcorticales[6]. Ello incluye el hipotálamo, el hipocampo, la amígdala y otras áreas cercanas que parecen ser las responsables de nuestra respuesta emocional.

El hipotálamo regula el hambre, el dolor, niveles de placer, y mucho más. También regula el funcionamiento del sistema nervioso autónomo el cual regula el pulso, la presión del sangre y la excitación en ciertas circunstancias. Numerosos artículos demuestran la relación de señales fisiológicas como el electrocardiograma, impedancia torácica y la actividad electrodermal con la actividad mental, emocional y física[7],[8].

Si añadimos señales como la respiración, la sudoración o la presión sanguínea, se puede conocer con más precisión el estado de estrés, emocional e incluso mental del sujeto bajo estudio. También sería posible saber si el soldado ha sido herido y en qué medida. Estudios demuestran que la fusión de diversas señales proporciona resultados más acertados que con solo una señal, como se podría prever.

Una parte muy importante es el **registro cerebral** (electroencefalograma y potenciales evocados) ya que a partir de las señales registradas será posible manejar diversas herramientas «con el pensamiento». Ya existen estudios en los que se utilizan las señales cerebrales para estudiar la carga mental y la detección de la mentira, entre otros.

Para detectar el electroencefalograma y utilizar los potenciales evocados, sería necesaria la utilización de un casco. En la Figura 3, se puede observar un ejemplo de un casco con diversos sensores dispuestos alrededor del mismo.

6 BOEREE, George C. «The emotional nervous system». 2009.Disponible en; http://www.mc3cb.com/pdf_ap_articles_2/2015_12_27_Limbic_The%20Emotional%20Nervous%20System.pdf, fecha de la consulta 02.09.2019.

7 PICARD, Rosalind W. VYZAS, Elias y HEALEY, Jennifer, «Toward machine emotional intelligence: Analysis of affective physiological state». 2001. Disponible en: https://ieeexplore.ieee.org/document/954607, fecha de la consulta 02.09.2019.

8 MOHINO-HERRANZ, I., GIL-PITA, R., FERREIRA, J., ROSA-ZURERA, M., SEOANE, F. «Assessment of mental, emotional and physical stress through analysis of physiological singlas using smartphones. Sensors». 2015. Disponible en: https://www.mdpi.com/1424-8220/15/10/25607/pdf, fecha de la consulta 02.09.2019.

Figura 3: Casco con diferentes sensores

Los sensores utilizados en el ejemplo cubren varios aspectos. En cuanto a **visualización**: cámaras 360º para visualizar la escena completa, una cámara frontal para visualizar en tiempo real lo que visualiza el soldado. Con respecto al **audio**, el casco podría disponer de un array de micrófonos 360º, además del micrófono cercano a la boca del soldado para registrar su voz. Además, se registra el **electroencefalograma** mediante varios electrodos dispuestos en el interior del casco. Por último, **gafas inteligentes**, las cuales proporcionan información al soldado con respecto a las señales vitales, de la posición y estado fisiológico propio y de sus compañeros, además de las condiciones meteorológicas que lo rodean, tanto presentes como futuras haciendo uso de la predicción realizada.

Textiles inteligentes

Otras señales fisiológicas pueden ser registradas mediante electrodos, más concretamente, textrodos, es decir, electrodos textiles (textiles inteligentes). Dichos electrodos están dispuestos en prendas, como chalecos, camisetas y guantes. La situación de los electrodos debe ser localizada en áreas del cuerpo donde el registro de la señal sea correcto.

Los textiles inteligentes, del inglés *Smart Textiles*, son textiles avanzados tecnológicamente hablando, desde textiles que protegen de los rayos UV, hasta textiles capaces de cambiar su color, visualizar mensajes, capturar señales fisiológicas (como las citadas), evitar lesiones o modificar su propia estructura para absorber la energía

de una bala.

El avance en textiles es muy importante y lo que se espera será mucho más.

A continuación, se explican algunos textiles muy útiles para el soldado[9]:

- Textiles con microcápsulas para termorregulación. Encuentran el equilibrio entre el calor generado por el cuerpo y el calor liberado al ambiente.
- Materiales de memoria de forma. Estos materiales ajustarán la protección al viento en función de la intensidad y la temperatura.
- Nanotecnología y membranas. Equilibrio entre la impermeabilidad y óptima transpiración
- Textiles antimicrobianos.
- Textiles que cambian de color en función de distintas variables.
- Textiles electrónicos también conocidos como e-textiles del inglés Electronic

Textiles[10].

- Prenda compuesta por una red de fibras ópticas y conductoras cuyo objetivo es detectar las heridas, localizarlas de forma exacta y detectar su naturaleza. También prendas que cambian su composición para realizar un torniquete si fuese necesario.
- Prenda que recopila la información de diversas señales fisiológicas. De manera que registra y envía información acerca de la presión arterial, saturación de oxigeno en la sangre, temperatura del cuerpo, temperatura de la piel.
- Tejidos captadores de energía. Desarrollado con hilos fotovoltaicos para capturar la energía solar y convertirla en energía eléctrica para cargar baterías de los dispositivos.

Otra prenda útil es el **chaleco antifragmentos/antibalas**, en los que se buscan principalmente tres propiedades relevantes, como son **dureza, resistencia y fuerza.** La dureza se refiere a cuánto resiste un material la penetración de una fuerza externa. La fuerza se refiere a la cantidad de fuerza requerida para cambiar la forma de un material. En cuanto a la resistencia, un material resistente podría absorber una gran cantidad de energía antes de la fractura, pero podría deformarse al hacerlo, lo cual no sería aceptable

9 AYORA, Alberto, «Tejidos inteligentes: La tecnología detrás de las prendas», 2016, disponible en: https://www.desnivel.com/material/material-noticias/tejidos-inteligentes-la-tecnologia-detras-de-las-prendas/, fecha de la consulta 02.09.2019.

10 STOPPA, Matteo y CHIOLERIO, Alessandro, «Wearable Electronics and Smart Textiles: A Critical Review», Sensors, 2014, disponible en: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4168435/, fecha de la consulta 02.09.2019.

para la armadura corporal. El material ideal tendrá altos grados de las tres cualidades. El material subyacente debe ser lo suficientemente duro para desviar los asaltos sin romperse; debe resistir completamente la deformación, mientras que también tiene que tener la capacidad de absorber la energía y deformarse modestamente, para que la energía de un impacto balístico no afecte a la salud del soldado. Encontrar un material con estas cualidades es tarea complicada. El diamante tiene algunas cualidades que lo convierten en un material atractivo para la futura armadura. El grafeno también posee cualidades interesantes. Pero ambos tienen sus virtudes y sus defectos, de modo que se sigue investigando con diamantes sintéticos y polímeros de dos dimensiones entre otros, con lo que se llegará a una protección casi inexpugnable[11].

Importancia del registro de señales fisiológicas

El sentido de capturar las señales citadas es el valor que tiene utilizar la información extraída para predecir el futuro y lo más importante, para la ayuda en la toma de decisiones ya que proporciona información que el ser humano no posee.

A partir de los datos obtenidos, el mando y el propio soldado serán capaces de conocer la posición, el estado físico, el estado mental y emocional de cada uno de los soldados. Con dicha información, el mando será capaz de tomar decisiones con respecto a varios puntos de vista como son la estrategia a llevar a cabo, si hay que asistir a algún soldado porque ha sido herido y además con los drones es posible conocer cómo son los enemigos, qué tipo de armas utilizan para estudiar y predecir sus estrategias.

Neurotecnología

La neurotecnología se define como un conjunto de herramientas que pueden influir en el sistema nervioso del ser humano, especialmente en el cerebro. Algunas áreas de esta tecnología se utilizan para mejorar y reparar la función cerebral y permitir a los investigadores visualizar el cerebro.

Esta tecnología puede ser invasiva y no invasiva. En el caso de invasiva, se requiere cirugía para intervenir ciertas áreas del cerebro. En cuanto a la no invasiva, que será siempre más sencilla de implantar, utiliza electrodos para capturar o alterar las señales

11 SCHARRE, Paul, FISH, Lauren, KIDDER, Katherine y SCHAFER, Amy, «Emerging Technologies. Super Soldiers», Center for a New American Security, 2018. Disponible en: https://www.cnas.org/super-soldiers, fecha de la consulta 02.09.2019.

cerebrales, así como los diferentes estados sensoriales del sistema nervioso y el cerebro. Además del conocido electroencefalograma que utiliza electrodos comunes se pueden realizar magnetoencefalografías que es una técnica para medir los campos magnéticos que genera la actividad eléctrica del cerebro. Para capturar dichos campos magnéticos es preciso el uso de ciertos sensores muy sensibles.

Para mejorar el rendimiento del soldado, existe lo que se conoce como estimulación magnética transcraneal repetitiva, que causa que las neuronas en el cerebro se activen. Esto puede provocar que se alteren las conexiones en el cerebro y, modificando la sinapsis, mejorar el rendimiento motor y la cognición. En algunos experimentos se ha demostrado que aumenta la vigilancia y la cognición bajo el estado de fatiga. Estudios de la Fuerza Aérea de Estados Unidos demostraron que se reducen los tiempos de reacción así como las falsas alarmas al apuntar, lo que puede ser una mejora en una tarea del campo de batalla[12],[13].

Exosuits y exoesqueletos

Los exoesqueletos son creados con el objetivo inicial de aumentar la capacidad de un soldado de soportar la carga y desplazamiento de peso, así como facilitar el movimiento y aumentar el rendimiento del soldado en términos de fuerza, resistencia y protección. Éstos incluyen, por el momento, el dispositivo como tal, un cable, un controlador y el motor.

En la literatura, algunos autores diferencian entre Exoesqueleto y Exosuits[14]. La principal diferencia viene dada por que éstos últimos no soportan carga, sino que ayudan a las articulaciones a facilitar el movimiento y reducir la fatiga, aumentando la resistencia y ahorrando energía.

12 SCHARRE, Paul, FISH, Lauren, «Human Performance Enhancement», Centre for a New American Security. 2018. Disponible en: https://www.cnas.org/publications/reports/human-performance-enhancement-1, fecha de consulta 02.09.2019.

13 NELSON, J. T., McKINLEY, R. A., GOLOB, E.J., WARM, J.S. y PARASURAMAN, R., «Enhancing vigilance in operators with prefrontal cortex transcranial direct current stimulation (tDCS)», NeuroImage. Volumen 85, Part 3, 15 enero 2014, pag. 909-917.

14 SCHARRE, Paul, FISH, Lauren, KIDDER, Katherine y SCHAFER, Amy, «Emerging Technologies. Super Soldiers», Center for a New American Security, 2018, disponible en https://www.cnas.org/publications/reports/emerging-technologies-1, fecha de consulta 02.09.2019.

Exoesqueletos rígidos

El exoesqueleto rígido es una cáscara externa, que se coloca fuera del cuerpo proporcionando asistencia al movimiento y al soporte de carga. Un prototipo militar más conocido es el Tactical Assault Light Operator Suit (TALOS). Se está investigando en exoesqueletos que sean también capaces de soportar el peso de un casco provisto con más funcionalidades. Como se propone en secciones anteriores, se podría añadir un sistema array de micrófonos y cámaras dispuestas alrededor del mismo, proporcionando audio y visión 360º para cada uno de los soldados.

Además, los exoesqueletos podrán mejorar la puntería del soldado, proporcionando una estabilidad y reduciendo considerablemente los movimientos involuntarios propios de un ser humano.

Los exoesqueletos pueden ser principalmente clasificados en función de la actuación: por motor eléctrico, actuador neumático, hidráulico, actuador lineal eléctrico y motor de combustión interna. Cada uno tiene sus ventajas e inconvenientes, como son el ruido, la seguridad y el peso, entre otros.

Son varias las dificultades para crear el exoesqueleto ideal. Una de las principales limitaciones es la alimentación de estos dispositivos. La energía que necesita el dispositivo es elevada y el objetivo es que sea duradera, lo que añadiría sobrecarga a la equipación del soldado.

Otra importante dificultad que se encuentran los ingenieros en el diseño de estos dispositivos son los materiales a utilizar. En las pruebas iniciales utilizan el acero y el aluminio. Sin embargo, no son materiales muy apropiados para esta tarea, ya que el acero es pesado y necesitaría emplear demasiada energía para superar su propio peso, y en cuanto a las aleaciones de aluminio son muy ligeras, pero fallan por fatiga muy rápidamente, lo que podría provocar daños en el usuario. Otros materiales que se utilizan son el titanio y la fibra de carbono, ya que son materiales más resistentes y más livianos, lo que minimiza el coste de energía necesaria para mover el propio exoesqueleto, aunque se incrementa el coste de manera considerable.

Otro problema o reto a superar es el diseño del actuador (transforma la energía en la activación de las partes articuladas del exoesqueleto). Podemos encontrar los actuadores hidráulicos, potentes y precisos, pero pesados. También pueden ser neumáticos, aunque son más impredecibles para movimientos precisos. Además, es complicado lograr que el sistema sea ergonómico y tenga todos los posibles movimientos humanos, ya que puede impedir o dificultar movimientos de cadera y de hombros, al ser articulaciones esféricas, estando situado el centro de rotación en el interior del cuerpo. Aunque por otro lado el sistema no debe poder moverse de una manera que exceda el rango del movimiento del cuerpo humano, ya que podría dañar al mismo.

Los exoesqueletos pueden ser diseñados como completos o por módulos, es decir, hay exoesqueletos que se colocan en la cintura y se disponen a lo largo de las piernas

para ayudar a las articulaciones como las rodillas para tareas como el transporte de cargas pesadas en terreno irregular. También en la cadera, o en el tobillo, de manera que se transfiere la carga al suelo. En módulos, como es obvio, el consumo de energía de las baterías se reduce de manera considerable pero también la funcionalidad del mismo. El compromiso se establece en función de las necesidades.

Otras funcionalidades que serán añadidas a los exoesqueletos en un futuro no muy lejano será su uso mediante control por voz y mental. Aunque pueda parecer de ciencia ficción, será realidad porque ya hay mucho trabajo hecho y como siempre, el que primero lo haga ganará la primera batalla[15].

Exosuits soft

Éstos son definidos por el IEEE del inglés Institute of Electrical and Electronics Engineers (Instituto de Ingenieros Eléctricos y Electrónicos) como: «dispositivos que utilizan textiles para interactuar con el cuerpo y aplicar pares de torsión a través de fuerzas de tracción sobre el exterior del cuerpo en paralelo con los músculos utilizando la estructura ósea para soportar cargas compresivas»[16]. Por lo que se entiende, que dichos trajes mejoran el rendimiento de los músculos y huesos, incidiendo en una reducción del coste metabólico para realizar determinadas tareas.

Localización del soldado

Una parte importante de la monitorización del soldado es también conocer en qué posición se encuentra cada soldado, con cierta tolerancia, a partir de la señal de GPS. Existen sistemas que lo realizan con una precisión más que aceptable. Con la información de la posición de cada soldado puede resultar realmente útil estudiar estrategias haciendo uso de métodos de machine learning. En el caso de interiores, la localización del soldado puede realizarse mediante otro tipo de señales como sonoras, ultrasonidos o similares.

15 GROSS, Michael Joseph, «The Pentagon's Push to Program Soldiers' Brains», 2018, disponible en: https://www.theatlantic.com/magazine/archive/2018/11/the-pentagon-wants-to-weaponize-the-brain-what-could-go-wrong/570841/, fecha de consulta 02.09.2019.

16 ASBECK, A.T., De ROSSI, S.M.M, GALIANA, I., DING, Ye, WALSH, C. J., »Stronger, Smarter, Softer: Next-Generation Wearable Robots,« 2014, Robotics & Automation Magazine, IEEE, 21 no. 4

En los vehículos

Drones

La aviación no tripulada abarca un amplio abanico de aeronaves. El inicio de las aeronaves no tripuladas está vinculado con el desarrollo de los llamados «torpedos aéreos», antecesores de los misiles crucero, que después se desarrollaron a través de ramas de las bombas no guidas (no propulsadas), los blancos aéreos (drones), los modelos recreacionales y/o deportivos de radiocontrol, las aeronaves de reconocimiento, de investigación, de combate, e incluso algunos modelos de vuelo extra-atmosférico[17].

El término vehículo aéreo no tripulado (Unmanned Aerial Vehicle, UAV) se hizo común en los años 90 para describir aeronaves robóticas y reemplazó el término vehículo aéreo pilotado remotamente (Remotely Piloted Vehicle, RPV), surgen como aviones pilotados de manera remota. Sus capacidades de control autónomo están aumentando y se prevé que la inteligencia de la que están dotados se vea mejorada en poco tiempo. Las aeronaves con fines militares se conocen como Vehículos Aéreos de Combate No Tripulados (UCAV) destinadas a realizar tanto misiones de reconocimiento, de ataque u otras. Estas aeronaves tienen múltiples ventajas con respecto a otros vehículos. Además de ser autónomos o controlados de manera remota, cuentan con la posibilidad de un acceso rápido y sencillo a zonas a las que los humanos o equipos de rescate les puede resultar imposible de llegar. Los drones son capaces de acceder a una zona, evaluar la situación, obtener imágenes, audio y muestras.

Además, estos sistemas podrían rastrear una zona, encontrar a un soldado herido e incluso tratarlo ya que pueden estar provistos de brazos robóticos e instrumental médico para evaluar e incluso tratar heridas, realizar una reanimación cardio-pulmonar, e incluso llevar a cabo operaciones quirúrgicas de emergencia.

Estos sistemas, al estar provistos de cámara y micrófonos, pueden contener la inteligencia suficiente para cotejar la biometría de distintos seres humanos. Con esto, verificar e identificar a un soldado en concreto. Así, la persona que es identificada como mando, puede controlar el dron y determinadas acciones ejercidas por el mismo.

17 Cuerno Rejado, Cristina, «Origen y desarrollo de los Sistemas de Aeronaves Pilotadas por Control Remoto. Aplicaciones y Operación con Drones /RPAS», 2015, disponible en: http://drones.uv.es/origen-y-desarrollo-de-los-drones/, fecha de consulta 02.09.2019.

Mini-drones

Otra idea posiblemente útil para el ejército puede ser el uso de mini-drones. Con mini-drones se hace referencia en el presente capítulo a drones de pequeño tamaño, como puede ser el de una abeja. Debido a su pequeño tamaño se presupone la elevada dificultad en añadir inteligencia al mismo y problemas de eficiencia energética[18], por lo que una posibilidad sería que el conjunto fuera compuesto de varios dispositivos tipo abeja cada uno de los cuales dispusiera de un tipo de sensor, como son una cámara, micrófono, sensor de Infrarrojos, sensor de radiación, entre otros muchos.

Con esto, de forma corporativa se podría hacer fusión de información procedente de cada uno de los sensores y así tener una imagen visual, acústica e incluso ambiental de una zona concreta.

Además, dichos dispositivos podrán ser provistos de sistemas de autodestrucción provocando la misma a partir de cierta señal del soldado o mando.

También se les puede añadir inteligencia con interconexión entre cada uno de los dispositivos y volar en banda, como los estorninos, con objetivos en grupo. Son enjambres de mini drones con un objetivo común, en los que podría haber intercambio o relevo de funciones según las circunstancias o la consecución del objetivo común.

Vehículo Aéreo

Para entrenar a pilotos se utiliza inteligencia artificial con el fin de simular un vuelo real en combate habiendo alcanzado capacidades similares e incluso superiores a las de pilotos reales. El siguiente paso es que dichos dispositivos sean capaces de tripular una aeronave como un caza y además tomar decisiones. En este punto debemos hablar sobre la ética del uso de la inteligencia artificial en la toma de decisiones. Este tema es muy controvertido hasta el momento, y se llegará a un acuerdo más pronto que tarde, ya que como se puede observar es necesario.

A todos los vehículos tripulados hay que proveerlos de inteligencia artificial. Así, los vehículos podrían tener la capacidad de conocer el estado emocional del piloto, soldado o la persona o personas que se encuentran en el mismo. Hay diversas formas de conseguirlo con la utilización de sensores, ya sea integrados en el propio asiento (sensores de señales fisiológicas), cámaras o micrófonos, o en el propio equipo del soldado, entre otros. Es posible, y será obligado en el futuro, conocer algunas emociones,

18 TUCKER, Patrick, «DARPA Plans Bugbot 'Olympics' to Foster Breakthrough in Tiny Machines», Defense One, 2018, disponible en: https://www.defenseone.com/technology/2018/07/darpa-plans-bugbot-olympics-foster-breakthrough-tiny-machines/149847/, fecha de consulta 02.09.2019.

como son el pánico, estrés y relajación, de aquellas personas cuyas decisiones pueden afectar a las vidas de otros seres humanos como a las de sí mismos.

Tanto los vehículos de las fuerzas armadas como los del mundo civil deberán estar provistos de esta inteligencia de predicción emocional, cuyo objetivo será evitar accidentes provocados por el estado físico o psíquico del piloto. La detección precoz de emociones facilita la toma de nuevas decisiones por parte del sistema y del personal a cargo, pues puede enviar alertas a quien concierne e, incluso en ciertos casos de emergencia, llegar a tomar el control.

Vehículo Naval

Algunos de los vehículos utilizados por la Fuerza Naval son los vehículos submarinos no tripulados de búsqueda y caza de minas[19]. A estos vehículos cuya función es fundamental para la seguridad general, se les está añadiendo inteligencia en cuanto a la detección, clasificación e identificación de minas enterradas en entornos de reducida visibilidad. Se trata de vehículos controlados por control remoto, teniendo capacidades de dominio, control y comunicación plena. También los hay programados para el reconocimiento de las profundidades del fondo marino y posterior rescate.

Mediante sistemas de reconocimiento de imágenes de RADAR es posible determinar donde se encuentra un blanco determinado, esto es posible desde hace mucho tiempo. Pero se añade la inteligencia de detectar qué tipo de barco/submarino es el que se está detectando y así conocer si se trata o no de un enemigo.

Por otro lado, una parte interesante podría ser que una vez el vehículo detecte al enemigo pueda ser capaz de evitar o predecir las contramedidas del mismo. De esta manera los sistemas de inteligencia artificial deberían estar entrenados en el estudio del enemigo y de las posibilidades que hay en un momento determinado y en un lugar concreto. Esto se basa en el estudio de las posibilidades de maniobra del enemigo y la probabilidad de que cada posibilidad suceda.

Los vehículos terrestres del ejército están provistos de diversos mecanismos de defensa. En la actualidad, la misión principal es la observación e inteligencia.

Parece que la tendencia por las Fuerzas Armadas de diversos países es utilizar vehículos terrestres no tripulados del inglés Unmanned Ground Vehicle (UGV). En principio estos equipos tendrán como principales funciones las de búsqueda, detección, análisis y neutralización de explosivos[20]. Otras funcionalidades que se pueden atribuir

19 BOLOIX TORTOSA, Jaime, «Impacto de la Robótica y la Inteligencia Artificial en el Empleo y Efectividad de la Fuerza Naval», Trabajo Fin de Master, Universidad Complutense de Madrid, 2018.

20 ROS PAU, Antonio, «Los futuros vehículos terrestres no tripulados de las Fuerzas Armadas francesas», 2018, disponible en: https://www.defensa.com/otan-y-europa/futuros-vehiculos-terrestres-

serían: inspección de coches, revisión de cavidades y falsos techos, apertura de paquetes sospechosos, recolección de municiones y explosivos. También para lugares de difícil acceso, estos vehículos son de gran ayuda y pueden ser remotamente utilizados para la visualizar la escena, tomar muestras, e incluso ayudar a personas.

El vehículo terrestre puede ser provisto de inteligencia artificial de diversas maneras. Una de ellas puede ser la de localización del enemigo por el sonido. Añadiendo un array de micrófonos en el vehículo y con algoritmos de machine learning el sistema sería capaz de decir por donde vienen los disparos y así reorientar su dirección, además de detectar el tipo de arma que se está utilizando. En los vehículos terrestres pueden situarse numerosos sensores, como son cámaras 360º, cámaras electro-ópticas con visión térmica y nocturna mejorando considerablemente los sistemas de vigilancia y observación. Con esta información es posible reconocer al enemigo, haciendo uso de la disciplina científica conocida como «visión artificial».

En el ambiente

Estación meteorológica

En el campo de batalla, un barco o portaaviones, una **estación meteorológica** puede ser realmente útil para que el mando estratégico tome ciertas decisiones a partir de las medidas registradas y las predicciones realizadas.

Una estación meteorológica es una instalación destinada a medir y registrar con una frecuencia determinada diversas variables meteorológicas. Están compuestas por numerosos sensores para detectar cada uno de los parámetros atmosféricos relevantes y un microprocesador, el cual determina una condición meteorológica.

Algunos de los sensores son **termómetros** que registran la temperatura ambiente, del suelo y del subsuelo y un **termógrafo** para registrar fluctuaciones de temperatura. Para medir la cantidad de agua se utiliza un **pluviómetro**. El **barómetro** se usa para medir la presión atmosférica y el **piranómetro** para registrar la radiación solar, el **heliógrafo** para medir la duración e intensidad de los rayos solares. Otro dato relevante es la velocidad y dirección del viento, para lo cual se utiliza el **anemómetro** y la **veleta**, respectivamente.

Las estaciones meteorológicas móviles deben estar provistas de baterías con muy bajo consumo, y ser recargables a través de los elementos de la naturaleza, como por

no-tripulados-fuerzas-armadas, fecha de consulta 02.09.2019.

ejemplo estar provistas de placas solares.

Los datos aportados por las estaciones meteorológicas serán enviados bien a un maestro, es decir, un procesador central como el que se ha denominado PI o directamente de manera remota al lugar en el que se encuentra el mando superior.

Con los datos registrados y procesados no solamente es posible ver en tiempo real qué está ocurriendo en términos de clima en un momento determinado sino, además, ver el histórico de datos y por supuesto hacer predicciones de qué va a ocurrir en las próximas horas e incluso días. Esto puede ayudar sustancialmente en la toma de decisiones.

Conclusiones

Se abre un nuevo mundo de posibilidades con el uso de la inteligencia artificial en el ámbito militar. La información de la que se puede disponer es inmensa y puede ayudar a la toma de decisiones. Y no en mucho tiempo, los nuevos sistemas serán autónomos, es decir, serán capaces de tomar sus propias decisiones, siempre y cuando la ética lo permita.

Estos nuevos sistemas ofrecen una amplio abanico de posibilidades como: incrementar la seguridad de los soldados, mejorar la ayuda humanitaria en cuanto al rastreo de zonas catastróficas, la búsqueda y rescate de personas en peligro, el estudio del enemigo para mitigar los ataques, entre otras muchas posibilidades. Aunque por otro lado, es un arma de doble filo, con estos sistemas inteligentes es posible desarrollar armas con capacidad de destrucción sin límites.

La conclusión más relevante es que la utilización de sensores y de inteligencia artificial creará una nueva concepción en el mundo militar. La investigación en este área es y será fundamental ya que en estos momentos hay una gran competición en muchos países. Además, no debemos olvidar que estamos hablando de tecnologías duales, y que por tanto, toda investigación en el área militar es luego utilizada en el mundo civil, mejorando la vida de todos. Por tanto, invertir en inteligencia artificial en defensa es invertir en la mejora de la calidad en la vida de todos nosotros.

Bibliografía

ASBECK, A.T., De ROSSI, S.M.M., GALIANA, I., DING, Ye, WALSH, C. J. «Stronger, Smarter, Softer: Next-Generation Wearable Robots,» 2014, Robotics & Automation Magazine, IEEE, 21 no. 4

AYORA, Alberto, «Tejidos inteligentes: La tecnología detrás de las prendas», 2016. Disponible en: https://www.desnivel.com/material/material-noticias/tejidos-inteligentes-la-tecnologia-detras-de-las-prendas/, fecha de la consulta 02.09.2019.

AUSUBEL, D., NOVAK, J., & HANESIAN, H. «Educational Psychology: A Cognitive View (2nd Ed.)», Holt, Rinehart & Winston, New York, 1978.

BOEREE, George C. «The emotional nervous system», 2009, disponible en; http://www.mc3cb.com/pdf_ap_articles_2/2015_12_27_Limbic_The%20Emotional%20Nervous%20System.pdf, fecha de la consulta 02.09.2019.

BOLOIX TORTOSA, Jaime, «Impacto de la Robótica y la Inteligencia Artificial en el Empleo y Efectividad de la Fuerza Naval», Trabajo Fin de Master, Universidad Complutense de Madrid, 2018.

CUERNO REJADO, Cristina, «Origen y desarrollo de los Sistemas de Aeronaves Pilotadas por Control Remoto. Aplicaciones y Operación con Drones /RPAS», 2015, disponible en: http://drones.uv.es/origen-y-desarrollo-de-los-drones/, fecha de consulta 02.09.2019.

GOLDBERG, D. E. «Genetic algorithms – in search, optimization and machine learning.» Addison-Wesley Publising Company, Inc., 1989. ISBN:9780201157673.

GROSS, Michael Joseph «The Pentagon's Push to Program Soldiers' Brains», 2018, disponible en: https://www.theatlantic.com/magazine/archive/2018/11/the-pentagon-wants-to-weaponize-the-brain-what-could-go-wrong/570841/, fecha de consulta 02.09.2019.

GROTH, Olaf, NITZBERG, Mark y ESPOSITO, Mark, «Rules for Robots. Why We Need a Digital Magna Carta for the Age of Intelligent Machines» International Reports, 2018. Disponible en: https://www.kas.de/documents/252038/253252/7_dokument_dok_pdf_52115_2.pdf/1a5564f5-77ea-a5bc-228b-279d885c313b?version=1.0&t=1539647628555, fecha de la consulta 02.02.2019.

HOLLAND, J. H. «Adaptation in natural and artificial systems», University of Michigan Press, 1975.

MOHINO-HERRANZ, I., GIL-PITA, R., FERREIRA, J., ROSA-ZURERA, M., SEOANE, F. «Assessment of mental, emotional and physical stress through analysis of physiological singlas using smartphones», Sensors, 2015, disponible en: https://www.mdpi.com/1424-8220/15/10/25607/pdf, fecha de la consulta 02.09.2019.

NELSON, J. T., McKINLEY, R. A., GOLOB, E. J., WARM, J. S. y PARASURAMAN, R. «Enhancing vigilance in operators with prefrontal cortex transcranial direct current stimulation (tDCS)», Neurolmage, Volumen 85, Part 3, 15 enero 2014, pag. 909-917.

PICARD, Rosalind W., VYZAS, Elias y HEALEY, Jennifer, «Toward machine emotional intelligence: Analysis of affective physiological state», 2001, disponible en: https://ieeexplore.ieee.org/document/954607, fecha de la consulta 02.09.2019.

ROS PAU, Antonio, «Los futuros vehículos terrestres no tripulados de las Fuerzas Armadas francesas», 2018, disponible en: https://www.defensa.com/otan-y-europa/futuros-vehiculos-terrestres-no-tripulados-fuerzas-armadas, fecha de consulta 02.09.2019.

SCHARRE, Paul, FISH, Lauren, KIDDER, Katherine y SCHAFER, Amy, «Emerging Technologies. Super Soldiers», Center for a New American Security, 2018, disponible en: https://www.cnas.org/super-soldiers, fecha de la consulta 02.09.2019.

SCHARRE, Paul, FISH, Lauren, «Human Performance Enhancement» Centre for a New American Security. 2018, disponible en: https://www.cnas.org/publications/reports/human-performance-enhancement-1, fecha de consulta 02.09.2019.

SCHARRE, Paul, FISH, Lauren, KIDDER, Katherine, and SCHAFER, Amy, «Emerging Technologies. Super Soldiers», Center for a New American Security, 2018, disponible en https://www.cnas.org/publications/reports/emerging-technologies-1, fecha de consulta 02.09.2019.

STOPPA, Matteo y CHIOLERIO, Alessandro, «Wearable Electronics and Smart Textiles: A Critical Review», Sensors, 2014, disponible en: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4168435/, fecha de la consulta 02.09.2019.

TUCKER, Patrick, «DARPA Plans Bugbot 'Olympics' to Foster Breakthrough in Tiny Machines», Defense One, 2018, disponible en: https://www.defenseone.com/technology/2018/07/darpa-plans-bugbot-olympics-foster-breakthrough-tiny-machines/149847/, fecha de consulta 02.09.2019.

Capitulo II

Integración de datos para obtener la Common Operational Picture a nivel operacional y estratégico

Rocío Barragán Montes

Resumen

Una de las ventajas de la incorporación del big data a la sociedad es la posibilidad de tomar decisiones mejor informadas. La cantidad de datos que se generan diariamente en todos los ámbitos y las técnicas que se están desarrollando y empezando a implementar en la industria, permiten imaginar cómo sería el futuro considerando algoritmos basados en los datos, al contrario de los tradicionales algoritmos basados en programación. Además, el análisis de esos datos a través de minería de datos y métodos estadísticos de análisis predictivo puede facilitar la comprensión del dominio al que se aplica e incluso ofrecer un rango de posibles soluciones.

En el campo de la Defensa, la toma de decisiones por parte del Comandante debe estar apoyada en un conocimiento de la situación detallado, conocimiento que se puede ver incrementado por la agregación de datos de las diversas fuentes, integrándolos y analizándolos para ser visualizados de la manera más adecuada para el descubrimiento de información por parte de las personas. Este capítulo ofrece una visión de las distintas técnicas de minería de datos y *machine learning* que pueden mejorar la adquisición del conocimiento de la situación operacional a través de una visión operacional común por parte del Comandante.

Palabras clave:

Inteligencia Artificial, Big-Data, aprendizaje automático, Minería de datos, algoritmos, imagen operacional, toma de decisions.

Data integration into a Common Operational Picture at the operational and estrategic level

Abstract

One of the main advantages of the integration of big data into society is the chance of achieving a better-informed decision-making process. The amount of daily-generated data in all fields and the newly developed and gradually adopted techniques in the industry allow a first vision of a future where data drive algorithms in opposition to traditional programming. Moreover, data analysis through data mining and statistical predictive analytics may ease the domain understanding and provide a range of possible solutions.

In the Defence area, the Commander decision making must be supported on a detailed situation awareness, which can be enhanced capturing, integrating, analysing data and visualising them in the most appropriate display for human to discover information.

This chapter provides a vision of different data mining and machine learning techniques which can improve the knowledge acquisition of the common operational picture by the Commander.

Keywords:

Artificial Intelligence, Big-Data, machine learning, data-mining, algorithms, common operational picture, decision-making.

Introducción

En la sociedad actual, diariamente se generan enormes cantidades de datos, aunque solamente una pequeña parte es transformada en información. Las mejores tomas de decisión deben estar basadas en información extraída de datos y que la intuición sólo se aplique cuando no sea posible obtener conocimiento de otras fuentes reales y medibles. Más aún, la evaluación cuantitativa del rendimiento o de cualquier área susceptible de ser medida necesita de una estructura que soporte el almacenamiento, transformación, compartición y análisis de datos y difunda la información a todos los involucrados. Y finalmente, para hacer la información accesible al ser humano, ésta debe ser mostrada de una forma adecuada, compatible con las capacidades y habilidades de las personas.

Los datos también son la base para la inteligencia artificial, que se está incluyendo progresivamente en todos los ámbitos de la sociedad. Su consideración en las Fuerzas Armadas (FAS), no solo por ser tendencia, es crucial para ocupar una posición prevalente en el futuro. No obstante, utilizada maliciosamente también puede ser una amenaza a la seguridad y ser capaz de hacer tambalear un país. Su uso controlado y la detección de injerencias indeseadas son críticos para la seguridad nacional.

El mando es una de las áreas centrales de capacidad militar del futuro definidas en un informe de la European Defence Agency (EDA)[21]. Adicionalmente, en dicho informe se analizan siete tendencias en los requisitos en las capacidades militares futuras: compartición de información, toma de decisiones, cooperación civil y militar, movilidad, ciberespacio, capacidades no-letales y soldado mejorado. Distintas aplicaciones en estos campos se analizan en este documento, pero en este capítulo, nos centramos en las implicaciones futuras en el ejercicio del mando y del control que es la adquisición y mantenimiento de la conciencia situacional mediante una imagen operacional común (Common Operational Picture, COP) basada en datos, que permita: la detección inmediata de los factores que influyen en los resultados, la rapidez en la toma de decisiones, la distribución ágil de órdenes y la compartición de información con todos los actores involucrados, civiles, militares y comerciales, nacionales e internacionales.

El objetivo de este capítulo es proponer métodos de minería de datos y de *machine learning* que ayuden a la obtención de la COP de datos por parte del Comandante para que pueda desempeñar sus funciones con una imagen más completa, actualizada y real de los factores que influyen en su toma de decisiones, incluyendo componentes predictivos.

21 KEPE, Marta y otros, «Exploring Europe's Capability Requirements for 2035 and Beyond», European Defence Agency, 2018. Disponible en: https://www.eda.europa.eu/docs/default-source/brochures/cdp-brochure---exploring-europe-s-capability-requirements-for-2035-and-beyond.pdf, fecha de la consulta 03.09.2019.

Minería de datos, Big Data, *Machine learning* e Inteligencia Artificial

La integración de datos para su explotación por parte del Comandante debe estar basada en dos conceptos muy ampliamente utilizados: la minería de datos y el big data. Big data se refiere a la extracción, manipulación, almacenaje de los datos y búsqueda de patrones entre ellos, mientras que la minería de datos o *data mining* se refiere a la extracción de conocimiento a partir de grandes cantidades de datos. Dicho con otras palabras, la minería de datos requiere de un conocimiento del dominio al que se quieren aplicar las técnicas de big data.

Machine learning, o aprendizaje automático, se diferencia de un algoritmo tradicional programado por un humano en que el aprendizaje automático modeliza un comportamiento esperado inferido del comportamiento habitual de un sistema. Para realizar esta modelización automática, la máquina necesita conocer cuál es el comportamiento habitual en base a un gran número de datos, con suficiente granularidad, no sesgados ni manipulados. Por otro lado, el uso de la inteligencia artificial trata de simplificar la necesidad de personal en tareas automatizables, posibilitando la dedicación de las personas a otras tareas de más alta capacitación. Los expertos consideran que los métodos de *machine learning* proporcionan predicciones, mientras que la inteligencia artificial produce acciones, por lo que *machine learning* podría considerarse un subconjunto de la inteligencia artificial o la base de ella.

En cualquier escenario futuro en el ámbito de la Defensa, el Alto Mando siempre estará presente. El Comandante no se verá afectado por la relocalización de las tareas inherente a la inteligencia artificial. Al contrario, será el recipiente último de la información y el responsable de las decisiones, por lo que cuesta creer que las delegue en una máquina. La inteligencia artificial deberá ser diseñada para tratar de condensar la información procesada a nivel estratégico de una manera óptima, sin impedir el descenso al máximo nivel de detalle y el acceso a los datos que dan lugar a la información facilitada. De esa forma se posibilitará la toma rápida de decisiones y la distribución de órdenes a los implicados para la ejecución de las acciones concretas necesarias.

El conocimiento del dominio de Defensa es clave a la hora de diseñar las herramientas de minería de datos, *machine learning* e inteligencia artificial que serán estratégicamente útiles. Es interesante reseñar que, durante las etapas de diseño y validación de la implantación de técnicas de minería de datos en los sistemas de mando y control, la participación del Comandante es fundamental para asegurar que el sistema se adapta a las necesidades del usuario final. Asimismo, la validación de que las funcionalidades desarrolladas cumplen con los requisitos también deberán involucrar al personal que será usuario de las herramientas y, de manera particular, al Comandante.

La interoperabilidad es un parámetro a tener en cuenta durante el desarrollo de las herramientas de big data. Entendida como un marco general y común con el objetivo de flexibilizar la compartición de datos entre distintas entidades, nacionales

o internacionales, la interoperabilidad es un requisito que cualquier proyecto de big data en el ámbito de la Defensa deberá tener en cuenta. Facilitará la inclusión gradual de datos y funcionalidades provenientes de distintas fuentes y que cubran distintas necesidades y al mismo tiempo, minimizará el tiempo de implantación de las mismas.

En cualquier proyecto de minería de datos es necesario diseñar una seguridad a nivel de usuario para que los datos y la información sean mostrados solo a los perfiles adecuados. Además, la ciberseguridad, tal y como explica el actual Jefe de Estado Mayor del Mando Conjunto de Ciberdefensa, Enrique Cubeiro en el capítulo del presente documento: «*Inteligencia Artificial para la seguridad y defensa del Ciberespacio*», deberá estar presente en el diseño de cualquier sistema para evitar de raíz el acceso no autorizado de agentes externos. Será necesario la definición de perfiles de usuarios, el diseño de la arquitectura y de los elementos de ciberseguridad necesarios, el flujo de información entre los distintos actores involucrados, creadores o consumidores de datos, civiles o militares, etc.

Rol del Comandante: Mando y Control

La combinación de valentía, liderazgo ético, buen juicio, intuición, conciencia situacional y la capacidad de considerar visiones contrarias ayudan al Comandante a tomar decisiones informadas en situaciones complejas[22]. En este capítulo exploraremos de qué forma una presentación adecuada y en el momento justo de la información necesaria derivada de datos objetivos y en tiempo real puede proveer de un mejor juicio y conciencia situacional al Comandante.

Se considera al Comandante como el centro de la operación. Sin embargo, debe estar interconectado con toda la red tanto para nutrirse de información como para guiar estratégicamente a las distintas fuerzas bajo su mando. Para ello, la creación de un nuevo sistema basado en datos enriquecería la COP, ya que mejoraría las distintas etapas clave del proceso de mando: la recopilación e integración de información, la planificación, preparación y ejecución y la evaluación continua. Asimismo, el flujo de datos desde sus originadores a los consumidores de la información, sería facilitado por este sistema. De esta forma, el propio sistema haría llegar a los subordinados una visión global e informaría a los distintos miembros del escalafón de cómo sus acciones contribuyen al progreso de la operación.

Se debe posibilitar la contribución de cada uno de los actores en las distintas etapas del proceso de mando, desde los sistemas automáticos que acompañan a cada soldado

22 US Army, Joint Operations. Joint Publication 3-0, Incorporating change 1, Joint Chiefs of staff Washington DC, USA, 2018. Disponible en: https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_0ch1.pdf?ver=2018-11-27-160457-910, fecha de la consulta 03.09.2019.

(ver el capítulo «De las Células a los Bits» de la profesora Inmaculada Mohíno) a los analistas. El nivel más bajo de contribución y originador de una gran cantidad de datos serían los soldados y su equipamiento automático, la vigilancia realizada por los distintos sistemas como drones o satélites, la inteligencia recopilada sobre el terreno, etc. Por el otro extremo, el penúltimo nivel en el flujo de la información y los datos serían los mandos de las diferentes secciones de un Estado Mayor. Sus responsables deben acceder a la COP de datos para visualizar sus propios datos, de los que son responsables y que son necesarios para sus tareas y para facilitar al Comandante una visión ejecutiva del estado de sus secciones. Un esquema de los distintos niveles de mando y control y del flujo de la información se encuentra en la Figura 1. Sucesivamente, en cada uno de los escalafones del mando la información fluirá hacia arriba, posibilitando el ejercicio del control. Los datos que la originan pueden ser consultados al nivel de detalle más adecuado. El mando se ejercerá en sentido inverso a través de la COP de datos.

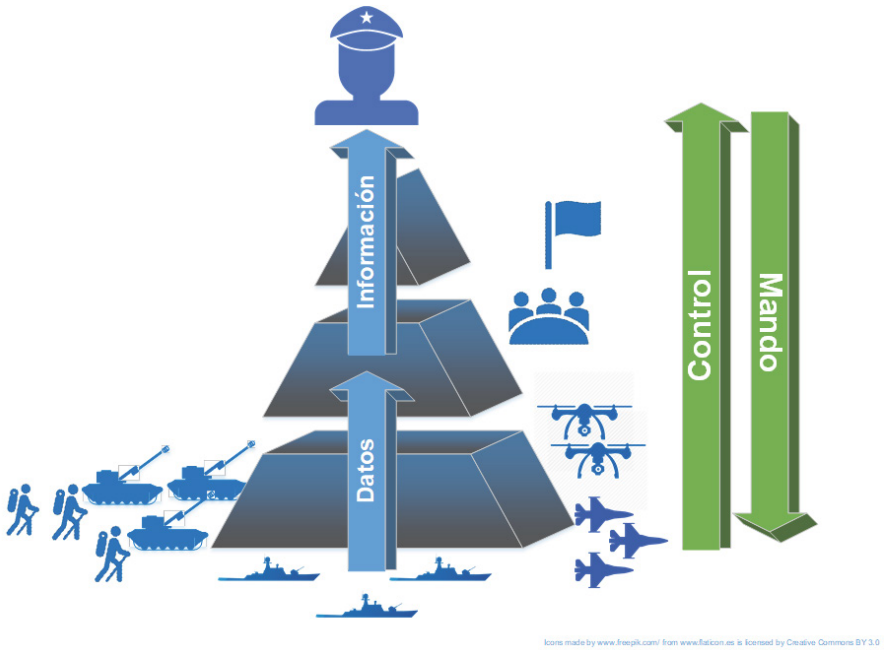


Figura 1. Flujo de conversión de datos en información para el ejercicio del mando y control

A continuación, se definen las principales funciones del Comandante y la relación que existe entre ellas para obtener la COP que se verá fortalecida por el uso de herramientas de *machine learning* y minería de datos.

Recopilación e integración de inteligencia

Las distintas fuentes de datos con información relevante para el Comandante deben estar a su alcance en un formato adecuado para la extracción del conocimiento. El estado de los combatientes, las tropas, los sistemas, el enemigo, las medidas de evaluación del rendimiento en tiempo real, etc. formarán parte de la COP basada en datos, por lo que la integración de los mismos debe estar disponible y accesible. Este concepto se conoce como«datos para la decisión» (*Data to Decision*, D2D). Se debe

poder asegurar un acceso rápido y efectivo a los datos a aquellos agentes implicados para garantizar unos tiempos de toma de decisión lo suficientemente cortos, incluso en escenarios altamente complejos y congestionados.

Una parte importante del futuro de la incorporación de los datos en la visión estratégica del Comandante es el despliegue de la nube de combate (Combat Cloud). Entendida como una red interconectada para la distribución de datos y el intercambio de información dentro de un espacio de batalla, cada usuario, plataforma o nodo autorizado contribuye y recibe información esencial de forma transparente y puede utilizarla en toda la gama de operaciones militares. La capacidad de recopilar datos e integrarlos en un sistema de información abierto y adaptable mejorará significativamente la capacidad de mando y control y la agilidad operativa de las fuerzas en combate[23]. Un esquema de la transmisión y centralización de la información en el entorno de combate presente en dicha fuente puede observarse en IHS Markit[24]. La información de todos los actores se debe incorporar al sistema para que fluya entre todos los actores del C4ISR (Command, Control, Communication and Computers, Intelligence, Surveillance and Reconnaissance).

En cualquier proyecto de minería de datos o de machine learning una parte muy importante del esfuerzo recae en el diseño de las ETL (extracción, transformación y carga, por sus siglas en inglés: Extract, Transform and Load). Estos procesos de manipulación de datos son parte de la nube de combate y son fundamentales para que los datos fluyan desde los originadores hasta los sistemas de explotación de la información en el formato y el tiempo adecuados.

En conclusión, será necesario invertir en la creación de una nube de combate eficiente, completa y segura, incluyendo el flujo de datos e información entre todos los actores afectados, pero solo entre ellos, y que sea escalable e interoperable, para posibilitar la incorporación progresiva de todos los datos que se vayan generando.

Planificación

Es el proceso mediante el cual el Comandante y resto de personal traducen su visión en un plan de acción centrado en los resultados esperados. Determina la manera de usar las capacidades militares en tiempo y lugar para conseguir objetivos considerando

23 Estado Mayor de la Defensa, Centro Conjunto de Desarrollo de Conceptos.«Entorno operativo 2035». Madrid, 2019, disponible en: https://publicaciones.defensa.gob.es/entorno-operativo-2035-libros-papel.html, fecha de la consulta 11/07/2019.

24 EDWARDS, James y otros, «Jane's by IHS Markit: C4ISR and Network Centric Warfare: Current Trends and Projected Developments», 2019, disponible en: https://www.janes.com/article/87673/intel-briefing-c4isr-network-centric-warfare-current-trends-and-projected-developments, fecha de la consulta 03/09/2019.

los riesgos asociados[25]. El conocimiento del entorno operativo es fundamental para la definición de planes y, por tanto, la COP juega un papel importante en esta fase del proceso operacional. La planificación de las tropas traduce la visión del Comandante en un plan de acción para la preparación y la ejecución, centrándose en los resultados esperados. Como afirmó Peter Drucker: «todo aquello que no se puede medir no se puede mejorar», por lo tanto, en esta fase es necesario definir métricas o indicadores de todos aquellos factores que se quieren evaluar o mejorar con unos valores ideales a alcanzar. Éstos medirán el grado de éxito de la operación.

La presencia de datos de todos los actores involucrados en la operación proveerá a la COP de una visión global efectiva, tanto en combate como en paz, para facilitar la planificación. La visualización de datos y los métodos de minería de datos predictivos y prescriptivos ayudarán a la toma de decisión acerca del plan de acción más adecuado para optimizar los indicadores de la operación.

Preparación

Incluye las actividades llevadas a cabo por las unidades para mejorar sus capacidades para ejecutar una operación. Algunas de estas actividades pueden ser refinar el plan, llevar a cabo maniobras, inteligencia, vigilancia, coordinación, inspecciones, etc. Estas actividades pueden dar una idea de los resultados esperados en una futura operación, proporcionar los objetivos de rendimiento de determinadas métricas y llevar a cabo una simulación de las herramientas que serán de utilidad en las operaciones, incluyendo el sistema de mando y control y la actualización de la COP de datos.

La preparación será un pilar fundamental de la COP, ya que proveerá de datos que se pueden convertir en información para la toma de decisiones, por ejemplo, mediante maniobras de vigilancia, información de inteligencia o labores logísticas. Asimismo, utilizando la COP de datos se puede definir el valor alcanzable de las métricas que luego servirán para evaluar el éxito de la misión.

Ejecución

La ejecución consiste en llevar a cabo un plan aplicando fuerzas de combate para cumplir la misión considerando el conocimiento de la situación para evaluar el progreso y tomar y refinar decisiones. Está considerablemente relacionado con el

25 US Army, Joint Planning. Joint Publication 5-0, Joint Chiefs of staff Washington DC, USA, 2017, disponible en: https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp5_0_20171606.pdf, fecha de la consulta 03.09.2019.

proceso de evaluación continua ya que la evolución de la fase de ejecución debe ser monitorizada mediante indicadores o métricas. Las herramientas que pueden evaluar anticipadamente cuál es el resultado de una determinada acción facilitan en gran medida el proceso de toma de decisiones en el momento de la ejecución.

Una de las funciones del mando es delegar en los subordinados la mayoría de las decisiones de acuerdo al concepto de operaciones previamente definido con el objetivo de minimizar el tiempo necesario durante la ejecución. Por lo tanto, la producción y seguimiento en tiempo real de todos los datos que afectan a la operación será posible en la COP de datos. Una COP ágil, que permita la sincronización de los datos y la transmisión de intenciones del Comandante de manera coordinada, será de gran ayuda para optimizar los tiempos de reacción.

Evaluación continua

Se refiere a la monitorización constante y la evaluación de la situación actual, del enemigo y del progreso de la operación. Es uno de los objetivos del Comandante y se verá enriquecida con herramientas que faciliten la conciencia situacional. La evolución de las tropas, localización del enemigo, efectos letales y no letales, y en general el entorno operacional pueden ser mejor descritos a través de indicadores cuantitativos que faciliten información sobre el estado y el grado de cumplimiento de objetivos.

Además, la evaluación del impacto de una actividad en un determinado indicador debe ser el objetivo de cualquier acción y la consecución de un determinado valor el fin de la misma. Algunas de las características a medir serían:

- Reacciones del enemigo y sus vulnerabilidades.
- Monitorización del avance de las actividades en relación con el estado final previsto por el Comandante.
- Evaluación de la operación en términos de efectividad y medida del rendimiento.

A través de la COP de datos el control y seguimiento de los factores que intervienen en el cálculo de los indicadores de éxito de la misión pueden ser accedidos y la causa de la degradación de la métrica puede ser detectada con exactitud. De esa manera, el mando se puede distribuir de manera rápida entre los actores involucrados y a todos los niveles.

El uso de métricas e indicadores se ve altamente beneficiado por las herramientas de minería de datos. Los componentes que intervienen en cada métrica pueden ser desgranados y las relaciones causa-efecto descubiertas para poder evitar o solucionar cuando un determinado rendimiento medido por un indicador se está viendo degradado. Las acciones a llevar a cabo pueden ser distribuidas rápidamente considerando una nube de combate consistente, lo cual asegurará además que cada miembro de las fuerzas recibe en tiempo y manera adecuadas las valoraciones objetivas

sobre su contribución al éxito de la operación. Una métrica de efectividad, como el ejército Estadounidense sugiere[26], proporciona un criterio para evaluar los cambios en el sistema, ya sea en su comportamiento, sus capacidades o el entorno operacional, de tal forma que se miden los resultados de las acciones tomadas y su proximidad o lejanía a los resultados esperados y se puede actuar en consecuencia.

En la Figura 2 se encuentra un gráfico de la relación entre las diferentes funciones que contribuyen a una COP eficiente basada en datos.

Figura 2. Distintas funciones del comandante y su contribución a una COP basada en datos

Aplicaciones de técnicas de *Machine learning*

Desde el comienzo del uso de máquinas en la revolución industrial se han estudiado las disciplinas y tareas en las que las máquinas deberían integrarse para producir mayor beneficio al trabajo realizado exclusivamente por personas. Más recientemente, a mediados del siglo pasado y en el ámbito del control del tráfico aéreo, Fitts [27] hacía referencia a habilidades en las cuales las máquinas eran mejores y otras en las que las personas eran mejores. Su lista MABA-MABA (iniciales de Machines Are Better At

26 US Army,«Operations FM 3-0», Department of the Army, Washington DC, USA, 2017, disponible en: https://fas.org/irp/doddir/army/fm3-0.pdf, fecha de la consulta 03.09.2018.

27 FITTS, Paul M.«Human Engineering for an Effective Air-Navigation and Traffic-Control System», Ohio State University Research Foundation, Washington DC, USA, 1951, disponible en: https://apps.dtic.mil/dtic/tr/fulltext/u2/b815893.pdf, fecha de consulta 11.02.2019.

- Men Are Better At), aunque aplicado en su día al ámbito de la Navegación Aérea y el Control del Tráfico Aéreo, se ha utilizado con éxito como base para estudiar la implantación de las máquinas en tareas en las cuales, o son mejores, o más rápidas, o más baratas que si son llevadas a cabo por parte de los humanos. Actividades que requieran velocidad, memoria a corto plazo, computación o consistencia en la salida de la información procesada, pueden ser llevadas a cabo más eficientemente por ordenadores. No obstante, Fitts también reconoce que, considerando las actividades en las que los humanos son mejores, éstos siempre deberían estar presentes en tareas de: razonamiento, percepción e interpretación de la información. Como se ha comprobado en distintos experimentos, las máquinas sin supervisión humana tienden a degradar sus respuestas, ya que carecen de nociones de ética o empatía, en contraposición a las actividades humanas cuyos resultados se adaptan, puesto que la flexibilidad es una característica en la cual, los humanos son mejores.

La garantía de que las personas siguen bajo control al utilizar sistemas radica en la elección del grado de automatización más adecuado para cada tarea, según los definidos por Parasuraman[28]. Los autores definen los distintos niveles en la delegación de funciones a las máquinas, como se muestra en la Tabla 1, desde el más básico, el 1, en el que el ordenador no ayuda a la toma de decisión y es el humano el único que participa en esa tarea, pasando por los niveles en los que la máquina ofrece posibles decisiones valorando cada una de las opciones, hasta aquellos niveles en los que la intervención humana es opcional. El máximo grado de automatización, el 10, sería aquél en el que la máquina toma sus decisiones sin informar al hombre. En los niveles a partir del 6 el ordenador puede realizar acciones de manera autónoma siendo el humano prescindible.

1	El ordenador no ayuda, el humano toma todas las decisiones y acciones
2	El ordenador sugiere un conjunto completo de posibles decisiones o acciones
3	El ordenador limita la selección a unas pocas alternativas
4	El ordenador sugiere una alternativa
5	El ordenador sugiere una acción que ejecuta si el humano la aprueba
6	El ordenador permite al humano un tiempo determinado para evitar una acción automática
7	El ordenador ejecuta automáticamente y necesariamente informa al humano
8	El ordenador informa al humano solo si éste le pregunta
9	El ordenador informa al humano solo si el ordenador lo decide
10	El ordenador decide todo, actúa autónomamente, ignorando al humano

Tabla 1 Distintos niveles de automatización según Parasuraman, Sheridan y Wickens

28 PARASURAMAN, R. SHERIDAN, T.B. y WICKENS, C.D. «A Model for Types and Levels of Human Interaction with Automation», Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 30.3, pag. 286–297, IEEE, 2000, disponible en: https://doi.org/10.1109/3468.844354, fecha de consulta 11.02.2019.

Tareas como la alerta cuando las reservas de munición bajan por debajo de un determinado porcentaje podrían ser automatizadas con el máximo nivel sin perder el control de la situación, considerando siempre la posibilidad de que el cumplimiento de la tarea, es decir, que efectivamente la alerta se efectúe, sea monitorizado por un humano.

Sin embargo, en dominios sensibles, con alta responsabilidad y sin posibilidad de errores, altos grados de automatización no son siempre la mejor opción. En concreto, las tareas que debe llevar a cabo el Comandante deberían moverse entre los niveles 2 y 8 de la escala de Parasuraman, puesto que la toma de decisiones debe estar supervisada siempre por el humano.

Machine learning es una disciplina que saca partido de las mejores capacidades de las computadoras en el procesamiento masivo de datos, mientras que el papel de las personas radica en entrenar a las máquinas y en analizar el resultado obtenido por ellas, ya sea información o modelos de predicción, para aplicarlo de la mejor manera al dominio en cuestión. Los procesos de machine learning constan generalmente de dos pasos:

1. Un entrenamiento en el que la máquina aprende y valida cómo se comportan los datos y como resultado genera un modelo predictivo.
2. La utilización del modelo creado para cualquier nuevo conjunto de datos. Éste ofrece un resultado esperado de los datos con un alto grado de exactitud.

Los datos aparecen presentados en instancias. Cada una de esas instancias tiene una serie de atributos. Con los datos de entrenamiento se genera un modelo que podrá predecir el comportamiento de futuras instancias en base al valor de sus atributos.

Una vez que se ha creado un modelo de predicción lo suficientemente preciso, se puede estudiar la importancia de cada uno de los atributos con los que se ha construido el modelo. Esto puede ayudar al Comandante a establecer prioridades de entre todos los factores que influyen en el éxito de una operación.

En el dominio del mando y control, el centro del sistema es el Comandante, el cual recibirá toda la información extraída por la máquina en base a modelos creados a partir de los datos para que sea él el que tome las decisiones informadas más adecuadas. Por lo tanto, como ya se ha dicho, en este entorno no tendría sentido utilizar el nivel 9 o 10 de automatización en la escala de Parasuraman. Sin embargo, las técnicas de *machine learning* sí pueden proveer al Comandante de modelos de predicción utilizando los distintos tipos de aprendizaje automático que se desglosan a continuación.

A continuación, se explican brevemente distintos tipos de aprendizaje automático o *machine learning*. En cada uno de ellos y a modo de ejemplo, se esboza su posible aplicación en algunos procesos necesarios para la toma de decisiones del Comandante.

Supervisado

El *machine learning* supervisado tiene como fin generar un modelo que prediga el valor de un atributo de salida considerando el valor de unos atributos de entrada. El atributo de salida puede ser numérico, un valor booleano (1-acierto, 0-error), o la clasificación a la que pertenece la instancia entre un conjunto de posibilidades. En este tipo de aprendizaje, en el conjunto de datos de entrenamiento, el atributo «respuesta» ha sido previamente etiquetado, con su valor numérico, booleano o con la clasificación a la que pertenece cada instancia. En otras palabras, las técnicas de *machine learning* supervisadas son aquellas que han necesitado la ayuda de un humano para enseñar a la máquina lo que es correcto e incorrecto. Son principalmente métodos de clasificación y de regresión.

Clasificación

En los métodos de clasificación la máquina dispone de un conjunto de datos con una serie de atributos, uno de ellos una clasificación de la instancia. El sistema aprenderá cuál suele ser la clasificación de las instancias en base al resto de atributos para que, en el futuro, cuando se presenten nuevas instancias, la máquina decida cuál es su clasificación con un alto grado de acierto.

Una posible aplicación de técnicas de clasificación podría ser un modelo de predicción que, en base a la identificación de carros de combate del enemigo, a su número, tipo y posición, pueda evaluar cuál es el potencial ofensivo de sus fuerzas. En tiempos de paz, varias de las actividades logísticas llevadas a cabo en los acuartelamientos pueden ser manejadas de manera más automática utilizando la clasificación. Por ejemplo, el nivel de operatividad de las fuerzas propias puede ser predicho en base a varios factores como los periodos de descanso de las tropas, la existencia de reservas de combustible, el mantenimiento de vehículos, etc.

Regresión

El método de regresión tiene como objetivo prever el valor del atributo respuesta de cada instancia nueva en base al valor del resto de sus atributos. Se entrena de manera supervisada, es decir, existe un conjunto de datos de aprendizaje correctamente etiquetados a partir de los cuales se crea el modelo predictivo. Análogamente a la clasificación, en la regresión el valor de los atributos de entrada determina con un alto grado de certidumbre el valor del atributo respuesta. El caso más sencillo sería una regresión lineal con un atributo de entrada y otro de salida (bivariante), en la cual, el valor del atributo respuesta crece o decrece al mismo ritmo que el atributo de entrada. Un caso interesante de regresión serían las series temporales, donde la temporalidad es uno de los atributos de entrada y que en mayor o menor medida determina el valor

del atributo de salida.

La aplicación más inmediata de los métodos de regresión en el ejército sería la predicción de los indicadores del rendimiento de la operación, la efectividad, por ejemplo. Otras aplicaciones de este tipo de aprendizaje son: determinar cuánto influyen las condiciones climatológicas en un área determinada en la operatividad de un enjambre de drones o cómo una fecha u hora en concreto puede afectar a la presencia de civiles en una zona. Se pueden usar series temporales para predecir resultados, por ejemplo, logísticos, de consumo de combustible de un acuartelamiento, o cómo la recepción de correo influye en la moral de las tropas.

No supervisado

A diferencia del supervisado, en el *machine learning* no supervisado se emplean en el entrenamiento conjuntos de datos no etiquetados ni clasificados de ningún modo. Los algoritmos no supervisados proporcionan el conocimiento inferido del análisis del conjunto de datos de entrada sin contexto sobre su significado. El aprendizaje automático no supervisado no necesita una intervención previa de ninguna persona ya que se basa en la búsqueda de relaciones entre atributos por parte de la máquina para determinar o bien agrupaciones de instancias, o bien atributos que son irrelevantes en las relaciones entre instancias. A la primera técnica se le denomina clusterización (o *clustering*) y a la segunda reducción de dimensionalidad.

Clustering

El *clustering* tiene como objetivo clasificar un conjunto de datos en diferentes agrupaciones o clústeres, en los cuales los miembros de un grupo común son similares entre ellos y diferentes de los miembros de otros grupos. Este método aprovecha las altas capacidades de las máquinas en el manejo masivo de datos para encontrar patrones que no son evidentes o perceptibles por el ojo humano.

Esta técnica utilizada en el ámbito de la Defensa puede ayudar a tomar las decisiones que optimizan el daño final, ya sea recibido o causado cuando no es posible identificarlo por las personas. Un ejemplo sería una relación compleja entre distintos factores que intervienen en el cálculo de un indicador de éxito de la operación. Agrupar las instancias representativas de fuerzas enemigas de tal forma que se identifiquen aquellas en las que una determinada acción les causa el mayor daño, o bien, aquellas que pueden causarnos mayor desgaste, permitiría planificar una maniobra que proteja nuestras fuerzas con la mayor efectividad. Otra posible utilidad sería la búsqueda de ataques en nuestro sistema en red, tal y como presenta el Capitán de Navío Enrique Cubeiro en su capítulo «*Inteligencia Artificial para la seguridad y defensa del Ciberespacio*». Debido a la gran cantidad de factores que intervienen en una red compleja que muchas veces los humanos no somos capaces de rastrear y a las distintas formas que pueden tomar

los ataques no autorizados, los métodos de *machine learning* no supervisados, como es el *clustering*, son los que mejor pueden revelar un mal funcionamiento del sistema y la presencia de una posible amenaza.

Reducción de dimensionalidad

Se refiere al proceso de reducir el número de variables aleatorias o atributos bajo consideración por medio de la obtención de un conjunto de variables principales con el fin de facilitar el estudio y el cómputo de un conjunto de datos. Algunas soluciones necesitan muchos datos, por lo que un filtrado previo, tanto de cantidad como de tipo de datos, resulta de gran utilidad para optimizar el rendimiento de la construcción y la predicción de los modelos. Para realizar la reducción de dimensión existen dos tipos de estrategias a seguir: se puede escoger un subconjunto menor a partir del conjunto principal de datos (selección de características) o reducir la dimensión global del conjunto principal de datos a una menor (extracción de características) mediante una combinación, lineal o no lineal, de atributos.

El método de reducción de dimensionalidad puede aplicarse como paso previo a todas las actividades y normalmente no necesita el conocimiento de un dominio para proponer qué atributos deben desecharse. Es por ello que para la COP de datos se utilizaría como paso previo para optimizar la extracción de información de los modelos predictivos que pudieran resultar más pesados de computar.

Aprendizaje profundo o Deep learning

El *deep learning* funciona en base a unidades interconectadas y agrupadas en diferentes capas que simulan el comportamiento de las neuronas en el cerebro humano. Cada neurona determina la existencia de una determinada característica en el objeto a clasificar y tiene un peso asociado. Con cada objeto que clasifique la máquina, el valor del peso se irá redefiniendo, de tal forma que vaya ponderando la contribución de esa característica en la clasificación del objeto. Los algoritmos de *deep learning* permiten obtener y extraer conocimiento de enormes volúmenes de datos multimedia, ya sean grandes repositorios de imágenes, vídeos, textos o audios. Estos algoritmos construyen de manera automática jerarquías de características que van de lo más sencillo a lo más complejo hasta crear una red neuronal multicapa. Pueden ser supervisados o no, puesto que puede haber intervención humana que le ayude a determinar la clasificación de instancias de entrenamiento o no.

Sin embargo, debe haber monitorización de los resultados por parte de personas conocedoras del dominio. El aprendizaje profundo es una de las técnicas más extendidas en la actualidad y ya está en funcionamiento en muchos sistemas, entre ellos militares. El reconocimiento de lugares, personas u otras situaciones en base a sensores sin actuación de las personas se utiliza actualmente en drones de combate,

en diversas armas autónomas y en vigilancia. El sistema que usa *deep learning* puede evaluar cualquier situación incluyendo el seguimiento del objetivo o la presencia de civiles y puede proveer al Comandante de una información muy valiosa para ejercitar el mando.

Asimismo, la digitalización de imágenes es otra área con aplicación al uso militar. Un uso para el Ejército del Aire podría ser detectar el valor analógico de los distintos instrumentos a bordo, la actitud de alabeo o cabeceo de una aeronave o, en general, información que puede ser adquirida de manera visual por la máquina para ofrecer un valor numérico a continuación, con el cual se pueden establecer métricas o evaluar lo que se percibe visualmente ofreciendo un indicador cuantitativo. Esta técnica facilita el conocimiento de la situación en cada momento de la batalla y el entendimiento de las decisiones que el soldado ha ido tomando.

Otra aplicación real y actual del uso de *deep learning* es la extracción de semántica del lenguaje natural a partir de la presencia u orden de determinadas palabras o frases. El reconocimiento y transcripción del lenguaje natural en conversaciones entre soldados durante una misión puede ser interesante para el análisis y la evaluación posterior del éxito de la misión, así como para encontrar puntos de mejora.

Muchos de los conceptos relativos a la desinformación que explica el Teniente Coronel Francisco A. Marín Gutiérrez, del Mando Conjunto de Ciberdefensa en este libro, usan el aprendizaje profundo, por ejemplo, la generación de audio, imágenes y vídeo falsificados o la creación de noticias falsas o *fake news*. Por ejemplo, tal y como comentaba el Teniente Coronel en su capítulo «*La inteligencia artificial en los ámbitos cibernético y cognitivo: su utilización en apoyo a la desinformación»,* Cambridge Analytica está acusado de recopilar información de usuarios de redes sociales para manipular a la opinión pública en diversos procesos electorales en todo el mundo mediante robots que generaron noticias falsas construidas en base a sus opiniones y que después fueron distribuidas por la red por millones de usuarios. En el lado opuesto, también se pueden utilizar técnicas de aprendizaje profundo para detectar cuándo se está produciendo una suplantación de individuos por *bots*.

Otros peligros del aprendizaje profundo también se conocen. Se ha demostrado que es posible falsear datos del conjunto de entrenamiento, haciéndole creer a la máquina que un determinado objeto, una imagen o un audio, no es lo que parecería ser. Engaños que no serían posibles a humanos, sí han sido posibles en máquinas[29]. Es por ello que se considera fundamental la presencia de las personas en tareas de monitorización en procesos críticos.

Aprendizaje de refuerzo

29 EDWARDS, Chris, «Hidden Messages Fool AI», Communications of the ACM, Vol. 62.1, pág. 13–14, ACM, New York, 2019, disponible en: https://doi.org/10.1145/3290412, fecha de consulta 03.09.2019.

En el aprendizaje de refuerzo el modelo es generado y realimentado automáticamente en cada nueva ejecución del algoritmo. A la máquina se le programan una serie de reglas básicas que regirán cada ejecución y una función de recompensa como resultado de la misma. La recompensa puede ser una función calculada por la propia máquina o bien recibir el visto bueno de una persona externa. La máquina aprenderá con cada ejecución si las decisiones que ha tomado le han reportado una mejor recompensa y en base a eso tomará mejores decisiones en el futuro. Esta técnica se puede utilizar en herramientas *What-if*, en las que la máquina ofrece distintas posibles acciones y el resultado de cada una de ellas para que el Comandante decida cuál es la más adecuada en cada momento. En base a las acciones tomadas por el Comandante, la máquina seguirá aprendiendo para hacer sus futuras propuestas de acción más aproximadas al criterio del Comandante.

Aplicaciones de Minería de datos

La minería de datos o data mining se refiere a la extracción de conocimiento a partir de grandes cantidades de datos. En prácticamente todos los ámbitos empresariales se está llevando a cabo una modernización de las herramientas de monitorización del negocio por medio del análisis masivo de datos. Iniciativas como Big Data Value Association[30] fomentan su incorporación en el ámbito empresarial sugiriendo además mejores prácticas y muchas de las definiciones que se usan en este capítulo están extraídas de dicha fuente. La minería de datos depende en gran medida de la manera de mostrar la información, por lo que el toque final a cualquier proyecto de minería de datos debe ponerlo una herramienta de visualización que tenga en cuenta las capacidades del humano para la adquisición del conocimiento de una manera clara y concisa.

El Comandante también se puede beneficiar de la extracción de la información a partir de una gran cantidad de datos que se generan diariamente. El estudio estadístico de los datos a través de novedosas herramientas puede complementar las utilizadas tradicionalmente para obtener la COP. Para ello es fundamental la integración de los datos en tiempo real a través de la nube de combate. Algunas actividades generan (o se prevé que generen) una gran cantidad de datos como, por ejemplo, los sistemas de protección y monitorización del soldado presentados en el capítulo «*De las células a bits*» de la profesora Inmaculada Mohíno o los robots de reconocimiento facial y de movimiento. Otro ejemplo serían los enjambres de drones, que podrían aportar no sólo información sobre el campo de batalla sino también del estado del propio

30 Big Data Value Association, European Big Data Value. Strategic Research and Innovation Agenda, October 2017, disponible en: http://www.bdva.eu/sites/default/files/BDVA_SRIA_v4_Ed1.1.pdf, fecha de la consulta 23.03.2019.

enjambre, el tiempo atmosférico, las fuerzas enemigas, su posición y movimientos, etc. El tratamiento adecuado de esos datos y la presentación eficiente al Comandante a través de la COP de datos es fundamental para sus objetivos. Además, el momento más adecuado para llevar a cabo determinadas acciones es una variable fundamental para el Comandante y puede ser mejor determinado con la colaboración de herramientas de análisis de datos.

El mundo se encamina a un futuro basado en el comportamiento de los datos y es por ello que se hace necesario que cualquier herramienta de análisis de datos utilice una seguridad y ciberseguridad para asegurar que los datos son accedidos sólo por aquellos agentes autorizados.

Análisis de Datos o Data Analytics

El término *data analytics* hace referencia al conjunto de técnicas y herramientas que permiten analizar de forma conjunta todos los datos, con el fin de encontrar relaciones entre ellos, detectar tendencias y encontrar casos atípicos o *outliers*. Es una parte fundamental de la inteligencia del negocio (por sus siglas en inglés, *Business Intelligence*, BI) relacionada con el manejo de datos. Se trata de la aplicación de técnicas estadísticas para obtener una visión global del estado de la organización basándose en los datos. Su análisis conduce a una visión operacional más completa y más rápida y a la toma de decisiones más informadas.

Análisis predictivo

El análisis predictivo es una aplicación de la minería de datos que consiste en la extracción de información existente en los datos, y su utilización para predecir tendencias y patrones de comportamiento, pudiendo aplicarse sobre cualquier evento futuro desconocido. Se trata de predecir con cierta exactitud el valor de una métrica o el grado de alcance de un determinado estado.

El análisis predictivo se puede usar para predecir actividades ilegales. La compra de material capaz de fabricar una bomba, la existencia previa de antecedentes delictivos, la visita a determinadas páginas web, etc. pueden ser precursores que indiquen una alta posibilidad de que un sujeto cometa un acto delictivo que comprometa la seguridad nacional. Asimismo, el movimiento de las fuerzas navales o terrestres y la ocupación de determinadas posiciones pueden ser precursores de una acción ofensiva. La relación entre los distintos factores y la conversión en una amenaza es revelada por una serie de conceptos estadísticos y debe ser explotado mediante una COP que presente la información de una manera adecuada para el Comandante.

Análisis prescriptivo

El objetivo del análisis prescriptivo de datos es ayudar en la toma de decisiones al recomendar las mejores acciones, preventivas o reactivas, basándose en los resultados del análisis descriptivo y predictivo, y, además, presentar el potencial impacto de dichas acciones.

El análisis prescriptivo mostrado al Comandante a través de herramientas *what-if* en la COP facilita que el control se pueda ejercer de manera más efectiva. Cuáles serían las diferencias en el resultado de la misión si el Comandante ejecuta una acción u otra. La información mostrada de la manera apropiada y la posibilidad de utilizar las mismas herramientas para distribuir las órdenes a los implicados contribuirá muy positivamente a la eficiencia del mando y control. El cálculo del tiempo restante para la misión considerando múltiples factores, o el momento más adecuado para tomar determinadas acciones también se puede enriquecer utilizando análisis prescriptivo.

Visualización de datos

La visualización de datos es una disciplina que ayuda a entender los datos de una manera visual, para facilitar posteriormente la comprensión y el análisis de la información, y la toma de decisiones. También se suele definir como «analítica descriptiva», ya que nos ayuda a entender qué ha pasado o qué está pasando en un momento dado. Este punto es clave para obtener una perspectiva de alto nivel de la situación operacional, la integración de fuerzas y del cumplimiento de objetivos es por ello que la visualización de los resultados de los procesos de minería de datos debe ser una parte importante de la COP de datos. No es extraño que la denominación tradicional la herramienta de visualización de datos sea cuadro de mando (en inglés *dashboard*).

La futura COP integrando *data analytics* incluirá información sobre el mando de la misión y sus métricas, posibilitando su seguimiento y englobando las variables que las determinan:

- La misión: objetivos medibles, nivel de cumplimiento proveniente de las fases de planificación y preparación.
- Enemigo: estado, posición, movimientos.
- Terreno y clima: situación geográfica, características que afectan al estado de las tropas, amigas y enemigas.
- Tropas y apoyo disponible: número, tipo, capacidades y condiciones. Se

actualizará en tiempo real con los datos de la nube de combate.

- Tiempo disponible para el cumplimiento de hitos y para la toma de decisiones.
- Consideraciones civiles.

La COP de datos incluirá el resultado de la aplicación todas las técnicas de minería de datos y *machine learning* anteriormente citadas que se estén utilizando sobre los datos provenientes de la nube de combate. La información se mostrará de la manera adecuada para la óptima comprensión por parte del Comandante del estado global de la misión, la aportación de cada una de las partes y para la distribución de órdenes y de la visión del Comandante hacia cada uno de los agentes implicados.

Un ejemplo de las posibilidades que pueden tener las COP de datos se pueden ver en juegos de guerra por ordenador. Típicamente se muestran las posiciones geográficas y el estado de cada una de las tropas propias a través de métricas, desde su capacidad de ataque hasta su estado físico y anímico. Es posible profundizar en los factores que influyen en cada uno de los datos mostrados, y así determinar con más exactitud las causas de los valores de cada uno de los indicadores. Se facilita también la posibilidad de filtrar por cada tipo de armamento, desplegando el detalle de unidades que lo forman y su capacidad operativa con las condiciones actuales. Se debería tener la posibilidad de incorporar las previsiones realizadas por el análisis predictivo y los distintos módulos de *machine learning* que se apliquen. De esa forma también se podrían mostrar gráficamente qué elementos enemigos son más críticos para ayudar al Comandante en la toma de decisiones. A posteriori, debería ser posible un análisis dinámico de los distintos factores que han intervenido en la batalla para poder sacar conclusiones.

Esta sería una COP en tiempos de guerra. En paz, la información a recibir incluirá más indicadores del grado de preparación de la misión, fijar los objetivos de la planificación, pero incorporando todos los datos reales que forman la nube de combate.

Conclusiones

En este capítulo nos hemos centrado en la incorporación de datos a la COP que necesita el Comandante para su toma de decisiones. Hemos visto distintas técnicas de minería de datos y de *machine learning* que aplicadas a distintos aspectos en el ámbito de la Defensa pueden generar información partiendo de la gran cantidad de datos que los distintos sistemas, humanos y técnicos, generan día a día.

El uso de la COP facilita los procesos que son llevados a cabo por el Comandante. Pero una COP basada en datos aporta al proceso de **control** la información objetiva y medible sobre el estado de cada uno de los factores que influyen en el éxito de la

operación. La distribución de las órdenes al último nivel de detalle involucrado es también posible, aportando valor añadido al proceso de **mando**. Inspirándonos en la frase del famoso estadístico W. Edwards Deming: «Creemos en Dios. Todos los demás, que traigan datos», el aporte que la adición de datos a la COP será, entre otros: fiabilidad, objetividad, causalidad, agilidad, rapidez, facilidad y proporcionará una clara ventaja a la tradicional imagen estática del estado de las tropas y la operación.

En el futuro se esperan otras formas de conflictos, por ejemplo, en el ámbito virtual. La monitorización del correcto estado de los sistemas, los accesos no deseados y la integridad de los datos en el ciberespacio será más fácil de llevar a cabo mediante herramientas de tratamiento masivo de datos.

Bibliografía

Big Data Value Association, *European Big Data Value. Strategic Research and Innovation Agenda*, October 2017, disponible en: http://www.bdva.eu/sites/default/files/BDVA_SRIA_v4_Ed1.1.pdf, fecha de la consulta 23.03.2019.

EDWARDS, Chris,«Hidden Messages Fool AI», *Communications of the ACM*, Vol. 62.1, pág. 13–14, ACM, New York, 2019, disponible en: https://doi.org/10.1145/3290412, fecha de la consulta 03.09.2019.

EDWARDS, James y otros,«Jane´s» by IHS Markit: C4ISR and Network Centric Warfare: Current Trends and Projected Developments», 2019. Disponible en: https://www.janes.com/article/87673/intel-briefing-c4isr-network-centric-warfare-current-trends-and-projected-developments, fecha de la consulta 03/09/2019.

ESTADO MAYOR DE LA DEFENSA (EMAD). Centro Conjunto de Desarrollo de Conceptos.«Entorno operativo 2035». Madrid, 2019. Disponible en: https://publicaciones.defensa.gob.es/entorno-operativo-2035-libros-papel.html, fecha de la consulta 11/07/2019.

FITTS, Paul M., Human Engineering for an Effective Air-Navigation and Traffic-Control System, Ohio State University Research Foundation, Washington DC, USA, 1951, disponible en: https://apps.dtic.mil/dtic/tr/fulltext/u2/b815893.pdf, fecha de consulta 11.02.2019

KEPE, Marta y otros,«Exploring Europe´s» Capability Requirements for 2035 and Beyond», European Defence Agency, 2018. Disponible en: https://www.eda.europa.eu/docs/default-source/brochures/cdp-brochure---exploring-europe-s-capability-requirements-for-2035-and-beyond.pdf, fecha de la consulta 03.09.2019.

PARASURAMAN, R., SHERIDAN, T.B., y WICKENS, C.D.,«A Model for Types and Levels of Human Interaction with Automation», Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 30.3, pag. 286–297, IEEE, 2000,

disponible en: https://doi.org/10.1109/3468.844354, fecha de la consulta 11.02.2019.

US Army,«Joint Operations. Joint Publication 3-0», Incorporating change 1, Joint Chiefs of staff Washington DC, USA, Joint Chiefs of staff, 2018. Disponible en: https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_0ch1.pdf?ver=2018-11-27-160457-910, fecha de la consulta 03.09.2019.

US Army,«Joint Planning. Joint Publication 5-0», Joint Chiefs of staff Washington DC, USA, 2017. Disponible en: https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp5_0_20171606.pdf, fecha de la consulta 03.09.2019.

US Army,«Operations FM 3-0», Department of the Army, Washington DC, USA, 2017, disponible en: https://fas.org/irp/doddir/army/fm3-0.pdf, fecha de la consulta 03.09.2018.

Capitulo III

La inteligencia artificial en el campo de la información: su utilización en apoyo a la desinformación

Francisco Antonio Marín Gutiérrez

Resumen

La finalidad de este capítulo es presentar las posibilidades de empleo de la Inteligencia Artificial en las operaciones que actualmente se desarrollan en el campo de la información. La combinación en un mismo conflicto de éste tipo de operaciones junto con otras de carácter convencional forma parte de lo que actualmente se denominan como acciones híbridas, constituyendo su descripción la primera parte del texto.

Tras analizar cómo queda recogido el concepto de lo híbrido en la normativa española se realiza una breve descripción de lo que supone la guerra híbrida en los ámbitos no tradicionales, como el cognitivo, dentro del cual se desarrollan las denominadas operaciones de desinformación.

Ya en la segunda parte de este trabajo se realiza una enumeración de las posibilidades de empleo de la Inteligencia Artificial tanto en apoyo a la desinformación como en contra de la misma, siendo una de las conclusiones obtenidas que, aunque de momento no existe una solución técnica que contrarreste totalmente el impacto de la desinformación, el uso de la Inteligencia Artificial puede contribuir a mitigar sus efectos.

Palabras clave

Ámbito cognitivo, amenaza híbrida, bots, desinformación, Inteligencia Artificial, noticias falsas.

Artificial Intelligence in the information field: the use of misinformation

Abstract

The purpose of this chapter is to present the possibilities of use of Artificial Intelligence in the operations that are currently being carried out in the information area. The combination in the same conflict of such operations together with conventional ones is a portion of what is currently referred to as hybrid actions, with the first part of the text being its description.

After analyzing the concept of «hybrid» in Spanish legislation, a brief description of what hybrid warfare entails in non-traditional domains, such as cognitive, where the so-called disinformation operations are performed.

In the second part of this work an enumeration of the possibilities of employment of Artificial Intelligence is carried out both in support of the misinformation and against it, being one of the conclusions obtained that, although there is currently no technical solution that completely counteracts the impact of misinformation, the use of Artificial Intelligence can help mitigate its effects.

Keywords

Artificial Intelligence, bots, cognitive domain, disinformation, hybrid threat, fake news.

Milicia es la vida del hombre contra la malicia del hombre, pelea la sagacidad con estratagemas de intención. Nunca obra lo que indica, apunta, sí, para deslumbrar amaga al aire con destreza y ejecuta en la impensada realidad, atenta siempre a desmentir. Echa una intención para asegurarse de la émula atención, y revuelve luego contra ella venciendo por lo impensado…

(Baltasar Gracián. Oráculo manual y Arte de Prudencia)[31]

La inteligencia artificial en el campo de la información: su utilización en apoyo a la desinformación

Con esta cita, extraída de una sentencia denominada *Obrar de intención, ya segunda, y ya primera,* se pretende evidenciar que los procesos de manipulación de la verdad no resultan nada nuevo – la obra fue publicada en 1647 -, y que las operaciones que actualmente se desarrollan en el campo de la información tienen una base teórica que es constante en la Historia. No obstante, el estado actual de la tecnología ofrece unas posibilidades hasta no hace mucho insospechadas para este tipo de operaciones, desempeñando en ellas un papel imprescindible la denominada Inteligencia Artificial (IA). Estas posibilidades son aprovechadas por actores estatales y no-estatales con la finalidad de crear un clima de confusión para provocar desestabilización. La combinación en un mismo conflicto de este tipo de operaciones con otras de carácter convencional es una parte importante de lo que se conoce como acciones híbridas, y el papel que en ellas puede desempeñar la IA se desarrolla a continuación.

Lo híbrido: amenazas y guerras híbridas

El término híbrido ha adquirido una relevancia cada vez mayor en los últimos tiempos y buena muestra de ello es que al introducir la palabra en los buscadores de internet más utilizados, aparezcan 1.330.000.000 resultados en inglés, reduciéndose a la «modesta» cantidad de 50.700.000 si se realiza la búsqueda en español (en el caso de amenaza híbrida aparecen 81.400.000 y 607.000 resultados, respectivamente). Pero comencemos definiendo el término, y para ello acudiremos al diccionario de la RAE, donde encontramos la siguiente acepción de híbrido: Dicho de una cosa: que es

31 GRACIÁN, Baltasar,«Oráculo manual y Arte de Prudencia», Biblioteca Virtual Miguel de Cervantes, Alicante, 1999, disponible en: http://www.cervantesvirtual.com/obra/oraculo-manual-y-arte-de-prudencia--0/, fecha de la consulta 26.06.2019.

producto de elementos de distinta naturaleza[32]. Esta definición básica, como veremos a continuación, resulta perfectamente aplicable a la hora de analizar los actuales conflictos internacionales.

La búsqueda de los antecedentes de lo que entendemos por amenazas o conflictos híbridos nos llevan hasta la National Defense Strategy (Estrategia de Defensa Nacional, NDS) de los Estados Unidos de 2002. Éste documento, el primero de su categoría redactado tras los ataques terroristas sufridos el 11 de septiembre de 2001, tenía como objetivo definir la forma en que los Estados Unidos de América debían de hacer frente a los futuros desafíos y en ella, tras declarar que América es una nación en guerra, se recogían los cuatro grandes desafíos – reiterados en la versión de la NDS de 2005[33] - que amenazaban los intereses de los Estados Unidos:

- **Tradicionales**: aquellos planteados por Estados que, en forma de conflictos militares, emplean capacidades y fuerzas militares reconocidas como tales.
- **Irregulares**: proceden de aquellos grupos y/o estados que utilizan métodos«no convencionales» para contrarrestar las ventajas tradicionales de los adversarios más fuertes.
- **Catastróficos**: incluyen la obtención, posesión y empleo de armas de destrucción masiva o de métodos que produzcan efectos similares a dichas armas.
- **Disruptivos**: pueden proceder de adversarios que desarrollan y utilizan tecnologías novedosas para negar a los Estados Unidos sus actuales ventajas en dominios operacionales clave.

Merece la pena resaltar como, de manera bastante clarividente, éste documento afirmaba que en el futuro los oponentes más capaces pueden tratar de combinar una capacidad verdaderamente disruptiva con formas de guerra tradicional, irregular o catastrófica, una afirmación que se ha convertido en realidad en nuestros días.

A partir de la idea de los cuatro desafíos de la NDS, varios autores acuñaron el concepto de híbrido para designar aquella categoría de conflicto en la que un adversario combina de forma simultánea varias de las formas de enfrentamiento por ellos representadas. Las fuentes más conocidas consideran que el precursor del concepto es el artículo Future Warfare: The Rise of Hybrid Wars[34], publicado en noviembre de 2005, no obstante, ya existe una referencia anterior en una comunicación de Erin M. Simpson, del Department of Government de la Universidad de Harvard, fechada en

32 RAE (Real Academia Española de la Lengua),«Diccionario de la lengua española», disponible en: http://dle.rae.es/?id=KIg05mN, fecha de la consulta 26.06.2019.

33 US DoD (Department of Defense),«The National Defense Strategy of the United States of America», Washington, 2005, disponible en: http://www.au.af.mil/au/awc/awcgate/nds/nds2005.pdf, fecha de la consulta 26.06.2019.

34 HOFFMAN, Frank y MATTIS, James«Future Warfare: The Rise of Hybrid Wars», Revista Proceedings, nov 2005, US Naval Institute, Maryland, 2005.

abril de 2005 titulada Thinking about Modern Conflict: Hybrid Wars, Strategy, and War Aims[35].

Independientemente del interés en determinar quién fue el primero en utilizar el término, podemos afirmar que el uso de estrategias híbridas en un conflicto no es algo nuevo, aunque sí resulta novedosa la forma en la que una amplia gama de instrumentos políticos, civiles y militares se combinan y aplican de forma coherente y coordinada contra una vulnerabilidad específica de las naciones u organizaciones contra los que se dirigen, todo ello con el fin de alcanzar objetivos estratégicos.

La globalización, respaldada por los avances tecnológicos, en particular en el ámbito de las comunicaciones y del ciberespacio, ha motivado que tanto las naciones como las organizaciones internacionales hayan sufrido un incremento en el número de vulnerabilidades que pueden ser explotadas en una variedad de escenarios más allá del enfrentamiento militar.

Entre los indicadores de los modernos escenarios de conflictos híbridos destacan ciberataques de una sofisticación cada vez mayor – en los que ya se ha comenzado a utilizar la Inteligencia Artificial como herramienta -, complejas campañas de propaganda y desinformación, así como la presión política y económica dirigida y coordinada. Todo ello representa un desafío para la defensa pues esta combinación de desafíos excede de la amenaza militar tradicional.

En definitiva, las estrategias que promueven el uso de amenazas híbridas pretenden, en última instancia, dificultar, retrasar e impedir la oportuna toma de decisiones y socavar la capacidad de una nación o de una alianza – que bien pudieran ser la OTAN o la Unión Europea - para responder a dicha amenaza de forma rápida, firme y eficaz.

Lo *híbrido* en la normativa española

El término *híbrido* también ha quedado reflejado en la doctrina de seguridad y defensa española y, en línea con la actual opinión predominante en la OTAN, se considera adecuado hablar de amenazas o conflictos híbridos y no de guerras híbridas. El principal motivo es que hablar de *guerra* predispone a pensar únicamente en actividades bélicas, excluyendo la amplia gama de opciones restantes, y la realidad es que las agresiones actuales exceden de lo meramente militar.

Si revisamos los principales documentos de seguridad y defensa, encontraremos que el término híbrido aparece ya en la Estrategia de Seguridad Nacional (ESN) de

35 SIMPSON, Erin.«Thinking about Modern Conflict: Hybrid Wars, Strategy, and War Aims», comunicación del 9 de abril de 2005 en el Annual Meeting of the Midwest Political Science Association (Panel 13-10 Strategies for Modern War).

2017[36]. El documento recoge ampliamente las amenazas híbridas a lo largo de todo el texto, y ya en la carta de inicio del Presidente del Gobierno éste se refiere a ellas como una combinación de amenazas convencionales y no convencionales orientadas a la desestabilización de nuestra forma de vida, y cuya identificación y atribución resultan especialmente complicadas.

Posteriormente, la introducción proporciona la que se puede considerar como una acertada definición de amenaza híbrida:

> *Se trata de acciones combinadas que pueden incluir, junto al uso de métodos militares tradicionales, ciberataques, operaciones de manipulación de la información, o elementos de presión económica, que se han manifestado especialmente en procesos electorales. La finalidad última que se persigue es la desestabilización, el fomento de movimientos subversivos y la polarización de la opinión pública.*

Continuando con la ESN de 2017, en ella de nuevo se hace referencia a este tipo de amenazas en el capítulo 2 (Dinámicas de transformación de seguridad global) y, sobre todo, en el capítulo 4 cuando, al hablar de las amenazas y desafíos para la seguridad nacional, se afirma que a los tradicionales conflictos armados se unen formas adicionales de agresión e influencia, haciendo especial referencia a la ambigüedad y la dificultad de atribución son factores constantes de los denominados conflictos híbridos, definiéndolos como:

> *Aquellos que incorporan operaciones de información, subversión, presión económica y financiera junto a acciones militares. Estas acciones, perpetradas tanto por actores estatales como no-estatales, tienen por objeto la movilización de la opinión y la desestabilización política.*

El marco doctrinal de las Fuerzas Armadas españolas ha incorporado también el concepto y en este sentido, el siguiente documento al que haremos referencia es el Concepto de empleo de las FAS (CEFAS)[37], documento de referencia que establece el marco de actuación, la forma en que llevarán a cabo sus misiones y las características generales que deben reunir las fuerzas militares españolas en las operaciones. Dentro del mismo, en el capítulo dedicado a las Dinámicas de transformación de seguridad global, se destaca el crecimiento de los denominados conflictos y acciones híbridas, destacando que éste tipo de acciones son aquellas perpetradas tanto por Estados como

36 DSN (Departamento de Seguridad Nacional),«Estrategia de Seguridad Nacional», Madrid, 2017, disponible en https://www.dsn.gob.es/es/estrategias-publicaciones/estrategias/estrategia-seguridad-nacional-2017, fecha de la consulta 26.06.2019.

37 EMAD (ESTADO MAYOR DE LA DEFENSA),«Concepto de empleo de las Fuerzas Armadas, cambio 2», Madrid, 30 de mayo de 2018, disponible en: http://www.emad.mde.es/Galerias/EMAD/files/CEFAS-cambio-2.pdf, fecha de la consulta 26.06.2019.

por actores no estatales que combinan el empleo de medios militares con ataques cibernéticos, elementos de presión económica o campañas de influencia por las redes sociales.

Al referirse al marco estratégico específico militar, se considera que en este entorno serán cada vez más probables las operaciones que se desarrollen en la zona gris[38] del espectro, en la que concurran acciones diversas, con mayor o menor grado de ambigüedad y visibilidad, que persiguen crear un clima de desinformación y confusión para provocar desestabilización, tales como ciberataques, acciones de manipulación de la información, sabotajes, revueltas, etc, normalmente en un entorno de baja intensidad. La combinación de ésta clase de acciones con el uso de medios convencionales dará lugar a lo que conocemos como acciones híbridas. Finalmente añade que las actuales amenazas basan principalmente su fuerza en el fácil acceso y empleo de las tecnologías más avanzadas en operaciones híbridas que pueden combinar acciones en las dimensiones física, virtual o de opinión, con tácticas y procedimientos asimétricos y/o terroristas, llevados a cabo por actores, estatales y no-estatales, incluyendo organizaciones extremistas violentas, con implantación global muchas de ellas, y apoyadas en postulados ideológicos o religiosos radicales.

Y no podemos olvidar otro documento de relevancia, la Doctrina para el empleo de las Fuerzas Armadas[39], publicación doctrinal militar del más alto nivel que describe la forma de empleo de las Fuerzas Armadas y establece las normas fundamentales con las que estas operan. Dentro del mismo se caracteriza a la amenaza híbrida por emplear, de forma simultánea, procedimientos convencionales junto a tácticas irregulares y actividades terroristas, actos de crimen organizado, ataques en el ciberespacio, presión política, así como múltiples herramientas de información y desinformación, incluyendo las noticias falsas y la mentira en sí misma. Destacar también que señala como principal característica de ésta amenaza el que trata de alcanzar sus objetivos evitando cruzar el umbral que define un conflicto abierto, de manera que no se provoque una escalada militar.

Por todo lo anterior queda claro que el concepto está sobradamente definido e integrado en nuestra legislación, quedando integrado en el amplio abanico de amenazas que afectan a la defensa nacional y a las que tienen que hacer frente las Fuerzas Armadas españolas.

38 La denominada«zona gris» es aquella zona del«espectro de los conflictos donde predominan las actuaciones situadas al margen del principio de buena fe entre estados (bona fide) que pese a alterar notablemente la paz no cruzan los umbrales que permitirían o exigirían una respuesta armada». EMAD, Centro Conjunto de Desarrollo de Conceptos,«PDC-01 (A). Doctrina para el empleo de las Fuerzas Armadas», CESEDEN, Madrid, 2018, disponible en: https://publicaciones.defensa.gob.es/pdc-01-a-doctrina-para-el-empleo-de-las-fas-libros-papel.html, fecha de la consulta 26.06.2019.

39 EMAD,«PDC-01 (A), ibidem.

La guerra híbrida y los ámbitos no tradicionales

Las operaciones de carácter militar se llevan a cabo en espacios, físicos y no físicos, denominados ámbitos, dotados de características propias y diferenciadas que condicionan las aptitudes y procedimientos de las fuerzas que deben operar en ellos. Frente a los ámbitos considerados tradicionales – terrestre, marítimo y aeroespacial -, tenemos otros dos, el cognitivo y el ciberespacial, que además de tener una estrecha relación son aquellos en los que la Inteligencia Artificial puede ser aplicada de forma más efectiva. En lo que respecta al ámbito cognitivo, su definición más adecuada la encontramos en la«Doctrina para el empleo de las Fuerzas Armadas»[40], donde se recoge que es«un ámbito intangible, inherente al ser humano y consustancial a su capacidad de juicio y de toma de decisiones. Este ámbito alcanza a las voluntades de todas las personas afectadas por el conflicto y los sistemas de inteligencia artificial, por lo que impregna al resto de ámbitos.» En el mismo apartado, se especifica que«su principal limitación es que para operar en él se manejan aspectos intangibles y de difícil evaluación, como los valores, las percepciones, la conciencia, las actitudes y los prejuicios. Se entiende por percepción la interpretación subjetiva, elaboración personal o representación mental, fruto de la interiorización de la información y los estímulos recibidos del entorno.» Por tanto, la información recibida por el ser humano y su elaboración son una parte significativa del ámbito cognitivo.

El Centro Conjunto de Desarrollo de Conceptos (CCDC) se encuentra inmerso en dos estudios sobre el ámbito cognitivo que pretenden profundizar en el mismo, sin que este capítulo pretenda prejuzgar sus resultados.

En lo referente al ámbito ciberespacial, tratado en detalle en otro capítulo de este trabajo, la ya mencionada Doctrina lo define como el«ámbito artificial compuesto por infraestructuras, redes, sistemas de información y otros sistemas electrónicos, por su interacción a través de las líneas de comunicación sobre las que se propaga y el espectro electromagnético, así como por la información que es almacenada o transmitida a través de ellos. Es trasversal a los demás ámbitos y no está sujeto a un espacio geográfico determinado, y le caracterizan su extensión, el anonimato, la inmediatez y su fácil acceso».

Ampliando lo expuesto en el marco doctrinal, el Jefe del Estado Mayor de la Defensa (JEMAD) manifestó recientemente que el conflicto actual redefine el campo de batalla tradicional, amplia el espacio de enfrentamiento y busca nuevos escenarios de confrontación.

Y esta tendencia es algo que se hace patente tanto en el plano físico como en el plano virtual,«donde estamos asistiendo a un protagonismo del enfrentamiento en el ciberespacio y en el ámbito cognitivo del ser humano, otorgando a las personas y,

40 EMAD,PDC-01 (A), ibidem.

sobre todo, a sus percepciones un papel central»[41].

Resulta evidente que en los últimos años se evidencia un continuo incremento en el número de ataques realizados contra un país en el campo de la información, es decir, contra su opinión pública, utilizando todos los medios de comunicación disponibles para la difusión de narrativas falsas. Este tipo de acciones afecta a todas las naciones y por ello la Comisión Europea encargó en 2018 a un grupo independiente de alto nivel, un estudio multidisciplinar acerca del fenómeno de la desinformación[42]. Una de las principales conclusiones del estudio es que el problema real es la desinformación – definida como información falsa, inexacta o engañosa diseñada, presentada y difundida para causar de forma intencionada un daño público o por lucro – y no las noticias falsas (conocidas según la terminología anglosajona como fake news), término que no consideran adecuado porque no captura la complejidad del problema y porque, además, está siendo utilizado por determinados políticos y sus seguidores para calificar todo aquello que no les resulta favorable.

Volviendo a la desinformación en nuestro ámbito nacional, y según indica un reciente y acertado documento del Centro Criptológico Nacional (CCN)[43], cada vez se utiliza más el ciberespacio como plataforma para lanzar ataques contra uno de los principales elementos que, perteneciente al ámbito cognitivo, configura una democracia liberal y un Estado-Nación moderno: la opinión pública. Los responsables de estos ataques suelen ser gobiernos y/o grupos organizados que tienen como objetivo erosionar y debilitar la cohesión interna de un Estado o alianza de estados considerados como adversarios. Tomando como punto de partida el referido documento se pueden considerar varios factores que contribuyen a impulsar el uso cada vez mayor de acciones hostiles basadas en campañas de desinformación que aprovechan las siguientes peculiaridades de los canales y redes digitales:

1. **Elevada efectividad**: resulta relativamente sencillo y asequible producir mensajes multimedia y difundirlos de manera directa y eficaz a través de canales digitales a las audiencias que, en función de los objetivos a alcanzar, se consideren las más adecuadas.

2. **Difícil atribución**: las plataformas digitales y la propia naturaleza de la red favorecen la aparición de actores que actúen de forma anónima. Hoy en día

41 ALEJANDRE, Fernando,«El papel de las Fuerzas Armadas ante la amenaza híbrida», XXX Seminario Internacional de Seguridad y Defensa,«La guerra hibrida: La mentira como arma y la verdad como víctima», Asociación de Periodistas Europeos, Toledo, 2018, pág. 220, disponible en: http://www.apeuropeos.org/descargas/XXX-Seminario-Defensa.pdf, fecha de la consulta 26.06.2019.

42 COMISIÓN EUROPEA. 2018,«Final report of the High Level Expert Group on Fake News and Online Disinformation», Unión Europea, 2018, disponible en: https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation, fecha de la consulta 26.06.2019.

43 CCN-CERT,«BP/13. Desinformación en el Ciberespacio», CNI, Madrid, 2019, disponible en:https://www.ccn-cert.cni.es/informes/informes-ccn-cert-publicos/3552-ccn-cert-bp-13-desinformacion-en-el-ciberespacio-1/file.html, fecha de la consulta 26.06.2019.

prácticamente cualquier usuario de Internet tiene a su al alcance las herramientas y procedimientos necesarios para utilizar las redes sociales de manera discreta, e incluso de crear sus propias redes de comunicación. Todo ello dificulta en gran medida la trazabilidad, fuente de origen y atribución de la información utilizada como arma.

3. **Compleja o inexistente regulación**: a diferencia de las acciones ofensivas tradicionales, las acciones de desinformación y de manipulación de la opinión pública no resultan sencillas de combatir desde la perspectiva legal propia de las democracias liberales

4. **Limitación para establecer una relación de causalidad**: resulta muy difícil poder probar una relación de causalidad entre los intentos por alterar la opinión pública y los cambios en el comportamiento de los ciudadanos.

5. **Aprovechamiento de vulnerabilidades sociales ya existentes**: en las fases iniciales de una operación de desinformación contra un Estado, los agentes responsables comienzan detectando vulnerabilidades sociales y políticas reales y espontáneas que se están produciendo en el debate público de dicho Estado para después centrarse en aumentar y polarizar ese debate.

6. **Infiltración de la desinformación ilegítima en los medios legítimos de comunicación social y política**: para distribuir sus propios mensajes y contenidos, los responsables de las operaciones de desinformación aprovechan el marco de utilización legítima que actores políticos y sociales hacen de las nuevas plataformas tecnológicas de difusión masiva de información.

Las acciones ofensivas en el campo de la información son fenómenos complejos que, además de las noticias falsas, utilizan diferentes herramientas y procedimientos para hacer llegar a los ciudadanos mensajes que causen el caos y la confusión en la opinión pública de un país considerado adversario. Precisamente la Inteligencia Artificial ha sido una de las últimas armas en incorporarse a la amplia panoplia de este tipo de acciones.

La Inteligencia Artificial en el campo de la información

Existen diversos tipos de acciones llevados a cabo en el campo de la información en las que ya se está utilizando con éxito la IA, y para dotar de cierta coherencia al análisis se han considerado tres grandes categorías para agrupar las distintas iniciativas, asociándolas a determinadas fases de una acción ofensiva tradicional:

1. Obtención de datos personales de posibles objetivos (Reconocimiento)
2. Generación de contenidos (Preparación del ataque)
3. Difusión de contenidos (Ejecución del ataque)

I - Empleo de la IA para la obtención de datos personales de posibles objetivos

El factor más relevante en este apartado es la utilización abusiva o improcedente de los algoritmos de redes sociales y motores de búsqueda. Los algoritmos son un conjunto ordenado de operaciones que permite realizar un cálculo y hallar la solución a un tipo concreto de problemas. Tanto las principales redes sociales – *Twitter, Facebook, Instagram,* etc - como los motores de búsqueda – *Google, Bing, Yahoo, Ask, AOL* y *Baidu* son los más populares en la actualidad - utilizan tales algoritmos para predecir lo que los usuarios están interesados en ver o buscar y poder así generar una posterior interacción del usuario. Basándose en los hábitos personales, el historial de lugares visitados, los enlaces activados, los recursos compartidos y las preferencias de un usuario, los algoritmos filtran y priorizan el contenido que dicho usuario recibe, incluyendo los anuncios que va a visualizar, que se cargan en las páginas visitadas en función de los parámetros mencionados. Los datos generados por los usuarios de las plataformas digitales ofrecen una imagen muy precisa de su comportamiento y patrones de consumo, siendo práctica habitual la utilización de dichos datos por las empresas para generar una publicidad dirigida.

En el caso de ser utilizados de forma maliciosa, los algoritmos emplearán los datos obtenidos de usuarios concretos para promover cierto tipo de información y eliminar otra, pudiendo ser utilizados para amplificar el impacto de las campañas de desinformación de una manera más precisa y eficaz. Un ejemplo de cómo se utilizan estas técnicas para diseñar mensajes políticos a medida lo tenemos en las actividades llevadas a cabo por la empresa Cambridge Analytica para la campaña presidencial de los Estados Unidos de 2016, en la que se realizaron estudios de personalidad de gran número de usuarios – considerando las posteriores ramificaciones llegaron a disponer de casi 50 millones de perfiles - a partir de sus datos en la red social *Facebook* y de la huella dejada por sus distintas actividades digitales. Estos análisis proporcionan los que se conoce como perfiles psicográficos, una categoría particular de los estudios de segmentación de los mercados. La segmentación psicográfica consiste en delimitar el público objetivo de un determinado producto o servicio en base a sus actitudes, aficiones, estilos de vida, a su personalidad en definitiva, pues son estos factores los que determinan los hábitos de compras. Al final, su principal objetivo es que las marcas, o los grupos políticos, puedan obtener una idea de por qué alguien podría comprar un producto específico, apoyar una causa dada o votar de cierta manera, siendo así capaces de promover sus productos como si fueran expresiones de estilo de vida.

La aplicación de herramientas de IA es la que hace posible el análisis de las cantidades masivas de datos obtenidos de los consumidores y la elaboración de los perfiles adecuados para poder aplicar posteriormente campañas específicas y personalizadas sobre los distintos grupos identificados.

II - Empleo de la IA para la generación de contenidos

En este apartado, que se ha querido asociar con la fase de preparación de un ataque, se van a enumerar una serie de técnicas que, aprovechando las capacidades ofrecidas por la Inteligencia Artificial, van a proporcionar las herramientas – que también podíamos calificar como armas – con las que se pueden preparar las actuales campañas de desinformación.

El Procesamiento de Lenguaje Natural (NLP) es el mecanismo utilizado por los programas de IA para entender e interpretar el lenguaje humano que reciben y, a su vez, para generar como respuesta sus propios mensajes personalizados. Un ejemplo de esta categoría sería la recopilación y análisis automatizado de las publicaciones de un usuario en la red social Twitter para encontrar información personal o temas clave en los que esta persona manifiesta interés; así, un programa de IA es capaz de generar correos electrónicos en los que se incluya esa información de interés, bien para convencer al usuario analizado para que abra un correo utilizado como vector de ataque en una campaña de spear-phishing[44] o bien, en el caso más peligroso, para suplantarlo.

Curiosamente, se ha demostrado que la IA suele tener más éxito que los propios humanos generando mensajes que engañen a sus víctimas. Además, mediante el empleo de IA un actor de la amenaza puede conseguir una tasa de éxito más elevada ya que puede hacer llegar sus mensajes a un número mayor de víctimas potenciales en un período de tiempo más corto.

Una evolución de elemento anterior para la creación de contenidos son los programas generadores de texto. Como ejemplo podemos mencionar OpenAI, un grupo de investigación de inteligencia artificial que recientemente, ha revelado un algoritmo de aprendizaje automático que es capaz de generar texto coherente, incluyendo artículos completos de noticias falsas, después de haber recibido una pequeña muestra de texto a partir del cual construir su propio mensaje.

Una de las principales características es que el algoritmo puede ser ajustado para imitar el estilo de escritura del autor del texto de muestra y para ello el modelo, denominado GPT-2, ha sido diseñado con la función de predecir las siguientes palabras de un texto a partir de todas las anteriormente utilizadas. Utiliza un elevado número de parámetros y ha sido instruido tomando como referencia millones de páginas web.

44 Phishing es un método de ataque que busca obtener información personal o confidencial de los usuarios por medio del engaño recurriendo a la suplantación de la identidad digital de una entidad de confianza en el ciberespacio. El spear-phishing es un phishing dirigido de forma que se maximiza la probabilidad de engañar a la persona objeto del ataque, CCN-STIC,«Guía 401 Glosario y abreviaturas», CCN-CERT, Madrid, 2015.

Según sus desarrolladores, con todo este respaldo el GPT-2 es capaz de generar un texto sintético utilizando el mismo estilo utilizado en el texto de la referencia.[45]

Pero estas capacidades no han dejado indiferentes a nadie y se ha desarrollado cierta controversia[46] en cuanto a su posible uso. Debido a su potencial para ser utilizado para la creación de campañas de desinformación, OpenAI ha optado por no hacer disponibles para uso público las versiones más potentes de su algoritmo, y tampoco publicará el código fuente ni los datos de formación utilizados para mejorar el software. No obstante, similares desarrollos pueden estar ya en marcha y ser utilizados por actores de la amenaza que tengan la suficiente capacidad como para utilizarlos.

Empleo de la IA para la generación de audio, imágenes y video falsificados. Otra aplicación de la IA es el aprovechamiento de su capacidad para generar imágenes, audio y video creíbles pero falsos. Existen aplicaciones prácticas de IA de carácter beneficioso utilizadas para la creación de contenidos de videojuegos o efectos especiales de películas; sin embargo, estos mismos recursos pueden ser fácilmente utilizados para engañar a audiencias específicas.

Respecto a aplicaciones de audio, a partir de una muestra de voz del tamaño y calidad suficientes determinados programas de IA son capaces de crear una voz artificial que resulte idéntica a la de la persona a la que pertenece la muestra utilizada, resultando posible incluir palabras que no estaban incluidas en la muestra inicial. Esta tecnología todavía está siendo perfeccionada y continuará mejorando progresivamente, si bien resulta preocupante el hecho de que la técnica lleva siendo utilizada desde algunos años en estafas de phishing basadas en voz conocidas como vishing, y otros engaños llevados a cabo a través de llamadas telefónicas[47]. Basándose en esta misma tecnología se han desarrollado también chatbots de IA que se hacen pasar de manera convincente por personal de soporte al cliente y que ya se han utilizado para engañar a víctimas para que proporcionen información personal y financiera.

En lo que a imágenes se refiere, una reciente aplicación de IA permite generar caras que no pertenecen a ningún ser humano real tomando como base una selección de imágenes faciales preseleccionadas. Estas caras generadas pueden parecer totalmente reales para las personas que las ven, y algunas pueden incluso engañar a otros programas de IA dedicados a detectar imágenes faciales falsas.

45 RADFORD, Alec y WU, Jeffrey,«Language Models are Unsupervised Multitask Learners», disponible en: https://d4mucfpksywv.cloudfront.net/better-language-models/language_models_are_unsupervised_multitask_learners.pdf, fecha de la consulta 26.06.2019.

46 LOWE, Ryan,«OpenAI's GPT-2: the model, the hype, and the controversy», Towards Data Science, disponible en: https://towardsdatascience.com/openais-gpt-2-the-model-the-hype-and-the-controversy-1109f4bfd5e8?gi=9f071a48fc01, fecha de la consulta 26.06.2019.

47 YEBOAH-BOATENG, Ezer y MATEKO, Priscilla,«Phishing, SMiShing & Vishing: An Assessment of Threats against Mobile Devices», Journal of Emerging Trends in Computing and Information Sciences, abril 2014, disponible en: https://pdfs.semanticscholar.org/7a27/1a3ff90b2a19d6b4f4ecc800e0aebdcda063.pdf, fecha de la consulta 26.06.2019.

La IA también puede manipular muestras de vídeo, intercambiando la cara de una persona con el cuerpo de otra y añadiendo expresiones faciales y movimientos realistas. Y si las modificaciones de vídeo realizadas mediante IA se combinan con un audio preseleccionado resulta factible crear vídeos creíbles de un político dando un discurso imaginario, como el famoso caso del falso discurso del expresidente de los Estados Unidos Barak Obama[48]. Esto es lo que se ha venido a llamar noticias falsas profundas, o deep fake news y en ellas las imágenes y los vídeos manipulados son utilizados para alimentar falsas historias en los medios de comunicación y promocionar o atacar a un individuo o grupo.

III - Empleo de la IA en la difusión de contenidos

Corresponde ahora presentar una serie de actividades que, directamente dependientes de técnicas de Inteligencia Artificial, pueden considerarse parte integral de la ejecución de una campaña de desinformación, es decir, de un ataque en el campo de la información.

Empleo de bots sociales (social bots) ilegítimos: un social bot (en español, bot social) es un tipo particular de programa utilizado en las redes sociales que simula ser un usuario humano, siendo capaz de proporcionar respuestas automáticas creíbles. Existen usos beneficiosos de estos programas, y así una red social como Facebook ofrece la posibilidad, dentro de su plataforma de mensajería asociada Facebook Messenger, de que empresas o particulares creen sus propios bots, normalmente como herramienta de atención al cliente. Pero por otro lado, en el marco de campañas de desinformación, los bots sociales ilegítimos se pueden utilizar - casi siempre de forma coordinada - para apoyar ciertas narrativas, amplificar mensajes engañosos, distorsionar determinados discursos e incluso funcionar como una cuenta asociada a un individuo con la finalidad de acumular seguidores. Para hacernos una idea de la magnitud del problema, según un estudio de la Universidades de Indiana y Southern California[49], en 2017 entre un 9 y un 15% de las cuentas activas de la red social Twitter podrían ser considerados bots sociales, si bien con posterioridad la propia red inició una intensa campaña para suspender todas aquellas cuentas sospechosas. Otras fuentes llegan a afirmar que un 24% de los mensajes publicados en esta red son generados por perfiles automatizados. A modo de curiosidad, existen cuentas conocidas dentro de la citada red Twitter que pueden incluirse dentro de ésta categoría. Un buen ejemplo es @DroptheIBot - creado por los periodistas Jorge Rivas and Patrick Hogan de

48 BBC,«Fake Obama created using AI video tool», BBC News, 19 julio 2017, disponible en: https://www.youtube.com/watch?v=AmUC4m6w1wo, fecha de la consulta 26.06.2019.

49 VAROL, Onur; FERRARA Emilio y otros,«Online Human-Bot Interactions: Detection, Estimation, and Characterization», disponible en: https://arxiv.org/pdf/1703.03107.pdf, fecha de la consulta 26.06.2019.

American Fusion.net – que corrige a todos aquellos usuarios de Twitter que hayan enviado un mensaje conteniendo la frase«illegal immigrant», remitiéndoles de manera automática el mensaje«People aren›t illegal. Try saying ‹undocumented immigrant› or ‹unauthorized immigrant› instead»[50].

Otro dato de interés es la existencia de programas comerciales, y por tanto de fácil acceso, cuyo objeto es gestionar multitud de perfiles en redes sociales posibilitando la coordinación de todos aquellos que un mismo individuo o grupo hayan creado en distintas redes para que se generen mensajes de forma simultánea y coordinada.

Utilización de perfiles digitales falsificados: La difusión de noticias falsas o maliciosas no sólo se produce a través de nuevos medios de comunicación de escasa credibilidad. También lo hacen a través de la manipulación o falsificación de perfiles digitales en redes sociales de personajes o instituciones reales con el objetivo de hacer creer a la opinión pública que han realizado unas declaraciones que, en realidad, nunca llegaron a hacer.

La manipulación de estos perfiles digitales puede adoptar diferentes aspectos o modalidades, en los que se puede hacer un uso extensivo de la IA:

- Simulación maliciosa de cuentas y perfiles reales: el método más efectivo de suplantar el perfil digital de una persona o institución es mediante la utilización de software o sitios web que permiten recrear el aspecto de la cuenta de una red social real, pero añadiendo contenido inventado. Una vez hecha la recreación, se realiza una fotografía o una captura de pantalla (screenshot) de la publicación y se comparte a través de redes sociales como WhatsApp entre contactos cercanos, hasta que se viraliza el mensaje falso.

- Creación de perfiles digitales falsos o parodias: una posibilidad de difundir informaciones maliciosas a través de redes sociales es mediante la creación de un perfil digital de una persona o institución sin contar con su consentimiento y suplantando su identidad. Para evitar este tipo de situaciones, algunas plataformas digitales ofrecen la posibilidad de verificar y autentificar que el perfil digital de un individuo coincide con su verdadera identidad.

- Hackeo de perfiles digitales: una forma más sofisticada de manipulación es acceder sin autorización – *hackear* - a las contraseñas del usuario de un perfil digital para controlarlo de manera maliciosa durante un periodo de tiempo. Algunas de estas acciones han llegado a causar pérdidas y daños no sólo en la reputación de empresas, sino en la economía global.

Un ejemplo de acción perjudicial que se ha llevado a cabo utilizando esta metodología ocurrió el 23 de abril de 2013, cuando el grupo autodenominado«Ejército Electrónico

50 JUDAH, Sam,«The Twitter bot that ‹corrects› people who say ‹illegal immigrant›», BBC News 3 de agosto de 2015, disponible en: https://www.bbc.com/news/blogs-trending-33735177, fecha de la consulta 26.06.2019.

Sirio» se hizo con el control de la cuenta en Twitter de la agencia de noticias Associated Press (AP) y publicó información acerca de supuestas explosiones en la Casa Blanca que habían herido al presidente Obama[51]. En apenas minutos, este mensaje provocó un desplome en el índice Dow Jones la bolsa de Estados Unidos de más de 140 puntos que, afortunadamente solo duro cinco minutos.

A continuación trataremos de analizar dos casos posibles de uso de la IA en las acciones de desinformación: a favor o en contra de las mismas

Caso I: utilización de la IA en apoyo a la desinformación

Las posibilidades ofrecidas por la IA se convierten en muchos casos en herramientas de gran utilidad para las operaciones de desinformación. Buen ejemplo de ello es como, apoyadas en gran medida por IA, algunos Estados han sido, supuestamente, capaces de influir en determinados procesos electorales nacionales, convirtiéndose en un recurso más de los utilizados por las amenazas híbridas de este siglo. La gama de métodos utilizados incluye la propaganda, el engaño, la desinformación y otras tácticas que han sido utilizadas desde siempre para conseguir la desestabilización del adversario en el espacio físico. Lo que resulta novedoso en los ataques vistos en los últimos años es su transposición al ciberespacio, gracias a lo cual han conseguido una velocidad, escala e intensidad sin precedentes, facilitada por el rápido cambio tecnológico y la interconectividad global.

Resulta de especial interés para esta categoría de operaciones el hecho señalado en un reciente trabajo[52], de que tanto los modelos teóricos como los datos empíricos analizados demuestran que la información de baja calidad tiene la misma posibilidad de hacerse viral en las redes sociales que la información de elevada calidad. Además, en entornos experimentales se verifica que una mayor presencia en medios sociales de ciertos actores reduce la propensión a verificar los datos por ellos publicados. Es por ello que las autoridades nacionales, los periodistas y los propios responsables de dichas redes mantienen un intenso debate acerca de cómo combatir la amenaza de la desinformación.

Los robots de la Web, o«bots», son el tipo más común de agente autónomo utilizado en la propaganda computacional. Las capacidades de un *bot* se limitan a proporcionar

51 DOMM, Patti,«False Rumor of Explosion at White House Causes Stocks to Briefly Plunge», CNBC, 23 abril 2013, disponible en: https://www.cnbc.com/id/100646197, fecha de la consulta 26.06.2019.

52 CHESSEN, Matt,«The MADCOM future: how Artificial Intelligence will enhance computational propaganda, reprogram human culture, and threaten democracy... and what can be done about it», The Atlantic Council of the United States, Washington, 2017, disponible en: https://www.atlanticcouncil.org/images/publications/The_MADCOM_Future_RW_0926.pdf, fecha de la consulta 26.06.2019.

respuestas básicas a preguntas sencillas, publicar contenido en una programación o difundir contenido en respuesta a determinados criterios desencadenantes. Sin embargo, los *bots* pueden tener también un impacto desproporcionado porque es fácil crear muchos de ellos y poder publicar así grandes volúmenes de contenido con una elevada frecuencia, y sus perfiles están diseñados típicamente para imitar a los seres humanos que componen su población objetivo. Un individuo puede operar fácilmente cientos de *bots* de *Twitter* con poco conocimiento técnico, usando hardware y software fácilmente disponibles. Los *bots* son actualmente utilizados por naciones, corporaciones, políticos, hackers, individuos, grupos patrocinados por el estado, ONGs y organizaciones terroristas en sus esfuerzos por influir en las conversaciones en línea. Se pueden identificar los siguientes tipos:

- ***Bots* de propaganda:** intentan persuadir e influir difundiendo un gran volumen de información en la que se mezclan verdades, medias verdades y desinformación.

- ***Bots* de seguidores:** falsean la aparición de un amplio acuerdo o consenso en apoyo a una idea o persona imitando el apoyo de las bases en un proceso basado en la técnica de marketing conocida como«astroturfing». En dicha técnica se oculta al verdadero emisor de un mensaje publicitario o propagandístico y se hace pasar la campaña de difusión como si fuera una expresión popular y espontánea. A través de este controvertido método se crea una popularidad ficticia para que otras personas sean más proclives a aceptar la idea, marca o producto que se desea promover. Se pueden así manipular o«secuestrar» los algoritmos que determinan las tendencias de noticias o de personas, generando numerosos«me gusta» respecto a los contenidos o siguiendo en masa a determinados usuarios.

- ***Bots* obstaculizadores (*roadblocks*):** socavan ciertos discursos desviando las conversaciones. Esto podría ser relativamente benigno, actuando como un animoso nacionalista, o mediante distracciones sencillas del tipo como«mira este divertido video del gatito.» El comportamiento del *bot* obstaculizador puede y suele ser más insidioso, como los *hashtags* de spam utilizados por los activistas, de manera que sus conversaciones y la posibilidad de coordinación se vean abrumadas con términos incomprensibles. En los casos más extremos, los *bots* obstaculizadores se utilizan para boicotear, provocar controversias o intimidar a periodistas, activistas o cualquier otra persona, bombardeándolos con miles de mensajes amenazantes u odiosos.

- ***Bots* de IA**: son cada vez más capaces de entablar conversaciones creíbles sobre temas complejos. Por ejemplo, el chatbot de IA de Microsoft en mandarín Xiaoice tiene una elevada sofisticación, empatía y flexibilidad conversacional que hacen que«ella» sea extremadamente popular. Según ha sido definido por algunos autores, principal misión es la de convertirse en un chatbot de IA con el que los usuarios sean capaces de establecer conexiones emocionales a largo plazo, una característica que no sólo distingue a XiaoIce de los primeros chatbots sociales (como Eliza), sino también de otros asistentes

como Cortana. En sus cuatro primeros años de funcionamiento Xiaoice ha establecido contacto con 660 millones de consumidores, el usuario medio interactúa con ella 60 veces un mes, y fue clasificada como la mejor influencer de Weibo en 2015. Pero también hubo fiascos famosos como Tay, un chatbot lanzado por Microsoft en 2016 y que se volvió racista tras mantener múltiples conversaciones con trolls[53].

Caso II: utilización de la IA en contra a la desinformación

¿Y cómo nos podemos defender de estas campañas de desinformación que utilizan recursos tan sofisticados? Pues aplicando contramedidas igual de sofisticadas, empleando en un primer momento lo que siempre se ha conocido como sentido común, es decir, utilizando nuestra capacidad de discernimiento y de análisis acudiendo a fuentes de información fiables para comprobar la veracidad de las noticias que nos llegan a través de medios de fiabilidad desconocida o dudosa.

Además de aplicar desde un primer momento el sentido común siempre será adecuado utilizar otras medidas. Un primer paso sería la detección de perfiles digitales automatizados, ya que el éxito de una campaña de desinformación depende en gran medida del nivel de sofisticación empleado en la creación de tales perfiles. Así, resulta posible definir algunas características para determinar si detrás de una cuenta digital existe una persona real, o si bien estamos ante una herramienta automatizada utilizada para difundir de forma masiva mensajes seleccionados. Dichas características son las siguientes:

- **No correspondencia con una persona real verificable:** el primer rasgo de un perfil digital automatizado es que, ni el nombre, ni la fotografía, ni la información de la cuenta coinciden con una persona real identificable en otras fuentes. Asimismo, en el historial de contenidos del perfil no se observa ningún detalle ni comentario relativo a la vida personal de un individuo o una institución real.

- **Alta o inusual actividad diaria: l**as cuentas automatizadas destacan por su inusual comportamiento temporal y así publican un número de mensajes diarios inusualmente elevado (más de cincuenta) o generan cientos de mensajes sobre un mismo tema en un periodo de tiempo muy corto. También es un buen indicativo que los mensajes sean publicados durante las 24 horas del día los siete días de la semana.

- **Ausencia de seguidores o seguidores sospechosos de ser cuentas automáticas:**

53 GONZALEZ, María,«Microsoft retira su bot de IA después de que éste aprendiera y publicara mensajes racistas», disponible en: https://www.xataka.com/robotica-e-ia/microsoft-retira-su-bot-de-ia-despues-de-que-este-aprendiera-y-publicara-mensajes-racistas, fecha de la consulta 26.06.2019.

los perfiles digitales empleados en campañas de desinformación destacan por el escaso número de seguidores que tienen en su perfil o, alternativamente, por tener un elevado número de seguidores que también podrían ser cuentas automáticas tipo *bot*.

- **Unilateralidad:** los perfiles digitales automatizados no suelen dialogar en las redes sociales, sólo se limitan a difundir mensajes en conversaciones donde se encuentran sus audiencias potenciales.

- **Ausencia de contenido original:** otra de las principales características de las cuentas digitales de comportamiento no humano es la ausencia de contenido original. La mayoría de estos perfiles redifunden o interactúan con contenido creado por otros perfiles.

- **Poca variedad temática:** los perfiles digitales automatizados se centran en publicar y dar la mayor difusión a aquellos mensajes políticos o sociales para los que han sido creados.

- **Escasa variedad de fuentes:** de igual manera, las cuentas propias de campañas de desinformación utilizan de manera exclusiva aquellas fuentes que forman parte de la misma estrategia y crean y difunden mensajes similares con los mismos enfoques.

Un buen ejemplo de la utilización automatizada de perfiles en redes sociales lo encontramos en el grupo terrorista DAESH, que durante años utilizó esta la técnica de manera habitual para difundir sus campañas. En ellas, los expertos en comunicación del grupo terrorista generaban una media de cien perfiles digitales nuevos, que apenas contaban con seguidores, y que eran gestionados mediante *bots* que automatizaban la distribución de los mensajes. Además, la *viralización* de los contenidos se realizaba haciendo un uso parásito de los *hashtags* – etiquetas utilizadas en redes sociales formadas por una o más palabras clave para que los usuarios compartan contenidos de un tema en particular - más populares entre las conversaciones de sus audiencias potenciales.

Existen diversas páginas web - conocidas con el nombre genérico de *fact checkers* - dedicadas a verificar informaciones y actuar como fuente de validación de noticias. Las más conocidas, pertenecientes al ámbito anglosajón, son *FactCheck.org*, *FlackCheck*, *Politifact*, *Poynter*, *Snopes* (leyendas urbanas) y son utilizadas como fuente de validación de historias de procedencia incierta -, rumores de internet, cadenas de mensajes - principalmente estadounidenses. En España podemos enumerar a *CazaHoax, Maldito Bulo*, *La Chistera* (blog del diario *El Confidencial*), *El Tragabulos* (Sección del suplemento *Verne* del diario *El País*), canales de *Vost Spain* (plataforma de voluntarios de emergencias que, entre sus misiones principales incluye detectar y neutralizar bulos y rumores sobre esta clase de incidentes) o *Salud sin Bulos,*

Otros recursos disponibles para automatizar esta detección son las extensiones para navegadores de internet – por ejemplo, *Fake News Detector* para los navegadores *Firefox* y *Chrome* o *B.S. Detector* para Chrome – que buscan en una base de datos todos los enlaces no fiables y a continuación proporcionan advertencias visuales sobre

la presencia de enlaces cuestionables o que se está navegando en sitios de dudosa reputación.

También se pueden encontrar herramientas web sencillas que, como *Botometer*, permiten verificar los datos de cualquier cuenta de la red social *Twitter* para determinar la probabilidad de que se trate de un *bot* o programa diseñado para promover y difundir ciertos contenidos de forma automática. A modo de ejemplo, diremos que *Botometer* funciona mediante un algoritmo de aprendizaje automático desarrollado con la finalidad de discernir si una cuenta está controlada por un humano o si se trata de un *bot*, basándose en miles de ejemplos etiquetados. Cuando se desea comprobar una cuenta en una red social, el navegador solicita información de cientos de cuentas públicas, así como de sus menciones y reenvíos de mensajes utilizando la plataforma de *Twitter*. Estos datos son estudiados por el algoritmo, que utiliza 1200 características para perfilar la cuenta en cuestión, sus entornos, lenguaje utilizado y otros patrones de actividad. Así, en relación con el estudio de una cuenta concreta puede tomar en consideración características específicas del lenguaje y el sentimiento con el que se escribieron, así como el estudio de amigos, red, publicaciones temporales y usuarios que interactuaron con la cuenta en cuestión.

Conclusiones

Estamos asistiendo a cambios significativos en la forma en que se desarrolla el enfrentamiento entre estados. En primer lugar, se refuerza la importancia de dos escenarios, el ciberespacio y el de la información; en segundo lugar se otorga un mayor protagonismo, en algunos casos un papel central, a dos categorías de objetivos, las personas y, sobre todo, a sus percepciones.

El concepto de amenaza híbrida incorpora de forma simultánea tácticas y técnicas de todo tipo, desde acciones no convencionales, que incluyen habitualmente actos terroristas y criminales, hasta ciberataques, guerra psicológica, con otras de carácter puramente convencional, lo que obliga a aproximaciones globales para resolver las crisis, de baja o alta intensidad, en los espacios físico, virtual o de la información.

Aunque de momento no existe una solución técnica que contrarreste totalmente el impacto de la desinformación, el uso de la Inteligencia Artificial puede contribuir a mitigar sus efectos.

En cuanto a la posible expansión de las amenazas existentes, los costes de los ataques pueden reducirse mediante el uso escalable de los sistemas de IA para completar tareas que normalmente requerirían trabajo humano, inteligencia y experiencia. Esto supondría un incremento del conjunto de actores que pueden llevar a cabo determinados ataques, la velocidad a la que pueden llevar a cabo dichos ataques, y del conjunto de objetivos potenciales.

La generación de información a gran escala mediante *bots* controlados por herramientas de IA pueden apoyar los ataques lanzados a través de los canales de información para inundarlos con ruido (información falsa o simplemente distracción), haciéndolo más difícil la adquisición de información real.

Los sistemas que buscan humanos a los que para influir terminarán, inevitablemente, tratando de persuadir a otros perfiles controlados por máquinas que, a su vez, se hacen pasar por humanos. Se puede dar la paradoja de que *bots* y otros perfiles controlados por máquinas hablen entre ellos ahogando las conversaciones humanas con una marea de voces y contenidos generados artificiales, viéndose abrumado el entorno de información en línea por el discurso generado por máquinas diseñadas para persuadir y vender.

Las máquinas han llegado, y quieren hablar con nosotros. La forma en que abordemos el proceso de adaptación a la cacofonía de su discurso determinará nuestra percepción de la realidad

Analizadas estas técnicas no me queda más remedio regresar a la ya sugerida aplicación del sentido común. Hoy en día se da la paradoja de que tenemos a nuestro alcance una cantidad de canales de información inusitada respecto a lo que era habitual hace unos años; sin embargo; la población conforma en gran medida sus opiniones acerca de los temas más complejos a partir de los mensajes que les llegan a través de sus redes de conocidos, convirtiéndose ésta proximidad o afinidad con el remitente en la garantía de la veracidad de la información.

Confiemos en que se mantengan el análisis, la reflexión y la observación como herramientas fundamentales para contrarrestar las operaciones de influencia, y para ello volvemos a recurrir a la 13ª sentencia de la obra de Baltasar Gracián, citada al principio de este trabajo:

> *«Acude la observación entendiendo su perspicacia, y descubre las tinieblas revestidas de la luz; descifra la intención, más solapada cuanto más sencilla».*

Bibliografía.

ALEJANDRE, Fernando,«El papel de las Fuerzas Armadas ante la amenaza híbrida», XXX Seminario Internacional de Seguridad y Defensa,«La guerra hibrida: La mentira como arma y la verdad como víctima», Asociación de Periodistas Europeos, Toledo, 2018, disponible en: http://www.apeuropeos.org/descargas/XXX-Seminario-Defensa.pdf, fecha de la consulta 26.06.2019.

BBC,«Fake Obama created using AI video tool», BBC News, 19 julio 2017, disponible en: https://www.youtube.com/watch?v=AmUC4m6w1wo, fecha de la

consulta 26.06.2019.

CCN-CERT,«BP/13. Desinformación en el Ciberespacio», CNI, Madrid, 2019, disponible en: https://www.ccn-cert.cni.es/informes/informes-ccn-cert-publicos/3552-ccn-cert-bp-13-desinformacion-en-el-ciberespacio-1/file.html, fecha de la consulta 26.06.2019.

CCN-STIC,«Guía 401 Glosario y abreviaturas», CCN-CERT, Madrid, 2015.

COMISIÓN EUROPEA. 2018, «Final report of the High Level Expert Group on Fake News and Online Disinformation», Unión Europea, 2018 disponible en https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation, fecha de la consulta 26.06.2019.

CHESSEN, Matt,«The MADCOM future: how Artificial Intelligence will enhance computational propaganda, reprogram human culture, and threaten democracy... and what can be done about it», The Atlantic Council of the United States, Washington, 2017, disponible en: https://www.atlanticcouncil.org/images/publications/The_MADCOM_Future_RW_0926.pdf, fecha de la consulta 26.06.2019.

DOMM, Patti,«False Rumor of Explosion at White House Causes Stocks to Briefly Plunge», CNBC, 23 abril 2013, disponible en: https://www.cnbc.com/id/100646197, fecha de la consulta 26.06.2019.

DSN (Departamento de Seguridad Nacional),«Estrategia de Seguridad Nacional», Madrid, 2017, disponible en https://www.dsn.gob.es/es/estrategias-publicaciones/estrategias/estrategia-seguridad-nacional-2017, fecha de la consulta 26.06.2019.

EMAD (ESTADO MAYOR DE LA DEFENSA),«Concepto de empleo de las Fuerzas Armadas, cambio 2», Madrid, 30 de mayo de 2018, disponible en: http://www.emad.mde.es/Galerias/EMAD/files/CEFAS-cambio-2.pdf, fecha de la consulta 26.06.2019.

EMAD, Centro Conjunto de Desarrollo de Conceptos,«PDC-01 (A). Doctrina para el empleo de las Fuerzas Armadas», CESEDEN, Madrid, 2018, disponible en: https://publicaciones.defensa.gob.es/pdc-01-a-doctrina-para-el-empleo-de-las-fas-libros-papel.html, fecha de la consulta 26.06.2019.

GONZALEZ, María,«Microsoft retira su bot de IA después de que éste aprendiera y publicara mensajes racistas», disponible en: https://www.xataka.com/robotica-e-ia/microsoft-retira-su-bot-de-ia-despues-de-que-este-aprendiera-y-publicara-mensajes-racistas, fecha de la consulta 26.06.2019.

GRACIÁN, Baltasar,«Oráculo manual y Arte de Prudencia», Biblioteca Virtual Miguel de Cervantes, Alicante, 1999, disponible en: *http://www.cervantesvirtual.com/obra/oraculo-manual-y-arte-de-prudencia--0/*, fecha de la consulta 26.06.2019.

HOFFMAN, Frank y MATTIS, James«Future Warfare: The Rise of Hybrid Wars», Revista *Proceedings*, nov 2005, US Naval Institute, Maryland, 2005.

JUDAH, Sam,«The Twitter bot that 'corrects' people who say 'illegal immigrant'», BBC News 3 de agosto de 2015, disponible en: https://www.bbc.com/news/blogs-trending-33735177, fecha de la consulta 26.06.2019.

LOWE, Ryan,«OpenAI's GPT-2: the model, the hype, and the controversy», Towards Data Science, disponible en: https://towardsdatascience.com/openais-gpt-2-the-model-the-hype-and-the-controversy-1109f4bfd5e8?gi=9f071a48fc01, fecha de la consulta 26.06.2019.

RADFORD, Alec y WU, Jeffrey,«Language Models are Unsupervised Multitask Learners», disponible en: https://d4mucfpksywv.cloudfront.net/better-language-models/language_models_are_unsupervised_multitask_learners.pdf, fecha de la consulta 26.06.2019.

RAE (Real Academia Española de la Lengua),«Diccionario de la lengua española», disponible en: http://dle.rae.es/?id=KIg05mN, fecha de la consulta 26.06.2019.

SIMPSON, Erin.«Thinking about Modern Conflict: Hybrid Wars, Strategy, and War Aims», comunicación del 9 de abril de 2005 en el *Annual Meeting of the Midwest Political Science Association* (Panel 13-10 Strategies for Modern War).

US DoD (Department of Defense),«The National Defense Strategy of the United States of America», Washington, 2005, disponible en: http://www.au.af.mil/au/awc/awcgate/nds/nds2005.pdf, fecha de la consulta 26.06.2019.

VAROL, Onur; FERRARA Emilio y otros,«Online Human-Bot Interactions: Detection, Estimation, and Characterization», disponible en: https://arxiv.org/pdf/1703.03107.pdf, fecha de la consulta 26.06.2019.

YEBOAH-BOATENG, Ezer y MATEKO, Priscilla,«Phishing, SMiShing & Vishing: An Assessment of Threats against Mobile Devices», Journal of Emerging Trends in Computing and Information Sciences, **a**bril 2014, disponible en: https://pdfs.semanticscholar.org/7a27/1a3ff90b2a19d6b4f4ecc800e0aebdcda063.pdf, fecha de la consulta 26.06.2019.

Capitulo IV

Inteligencia Artificial para la seguridad y defensa del Ciberespacio

Enrique Cubeiro Cabello

Resumen

Inteligencia Artificial (IA) y Ciberseguridad son dos de las disciplinas asociadas a las nuevas tecnologías que mayor interés y preocupación suscitan en la actualidad. El artículo dedica una primera parte a argumentar los motivos por los que la Ciberseguridad es uno de los campos en los que más determinante puede ser la irrupción de la IA; a continuación, y a partir del conocimiento de las vulnerabilidades y de las ciberamenazas que tratan de explotarlas, identifica y analiza posibles aplicaciones concretas de la IA en Ciberseguridad y Ciberdefensa; finalmente, esboza algunos posibles nuevos riesgos y problemas que para la Ciberseguridad y la Ciberdefensa pueden suponer tanto las imperfecciones de la IA como su utilización por parte de la amenaza.

Palabras clave

Inteligencia Artificial (IA), Ciberseguridad, Seguridad CIS, Ciberdefensa, Aprendizaje Automático, Aprendizaje Profundo, Big Data, Computación, Ciberamenazas, Ciberataques, Malware, Amenazas Persistentes Avanzadas, Ciclo O-O-D-A, Centro de Operaciones de Seguridad (COS), Centro de Operaciones Ciberespaciales (CyOC), Búsqueda de ciberamenazas, Gestión de Información y Eventos de Seguridad (SIEM).

Artificial Intelligence for the security and defense of Cybersecurity

Abstract

Artificial Intelligence (AI) and Cybersecurity are two of the disciplines associated with the new technologies that are of greatest interest and concern today. The article devotes a first part to arguing the reasons why Cybersecurity is one of the fields in which the irruption of the AI can be most decisive. Then, and from the knowledge of the vulnerabilities and the cyber threats that try to exploit them, possible specific applications of AI in Cybersecurity/Cyberdefense are identified and analyzed. Finally, it outlines some possible new risks and problems that the imperfections of the AI and its use by the threat can pose for Cybersecurity/Cybedefense.

Keywords

Artificial Intelligence, Cybersecurity, Computer Security, Cyberdefense, Machine Learning, Deep Learning, Big Data, Computing, Cyberthreats, Cyber Attacks, Malware, Advanced Persistent Threats, O-O-D-A Loop, Security Operation Center (SOC), Cyberspace Operations Center (CyOC), Threat Hunting, Security Information and Event Management (SIEM)

Introducción

IA y Ciberseguridad son dos áreas de conocimiento que tienen mucho en común: ambas están en la cresta de la ola, sin haber alcanzado aún la madurez; evolucionan con un ritmo de crecimiento exponencial y a las dos se les adivina un exitoso porvenir; comienzan a influir y modificar nuestros hábitos y parece que en un futuro no muy lejano comenzarán a tener un gran impacto en nuestras vidas; ambas tienen un profundo y muy complejo sustrato tecnológico; muy pocos seres humanos, si es que hay alguno, son capaces de abarcar el conocimiento necesario para su comprensión completa. Y hay más, que irán saliendo a lo largo de este artículo.

A priori, la IA tiene un sinfín de potenciales aplicaciones. Podemos vislumbrar las ventajas que puede aportar a campos y áreas tales como la conducción de vehículos, los diagnósticos médicos o el entretenimiento. Pero, obviamente, los campos en los que será mayor su utilidad e impacto serán todos aquellos en los que es necesario obtener una respuesta inmediata ante situaciones complejas y altamente cambiantes, lo que abre también un amplísimo abanico de posibilidades.

Por motivos obvios, entre esos campos están la Ciberseguridad y la Ciberdefensa, que aglutinan todos esos condicionantes y en las que no basta con un elevado porcentaje de éxito, en tanto una sola amenaza no detectada o una decisión errónea pueden tener muy graves repercusiones. Por lo tanto, para la seguridad y defensa en y del Ciberespacio, la integración de la IA no ha de verse simplemente como una línea de investigación interesante, sino que se ha convertido ya en una imperiosa necesidad.

Adentrarse en este terreno, cuando el que escribe pertenece a la generación de los «baby-boomers» produce bastante vértigo y hasta cierto desasosiego. Y, además, implica el elevado riesgo de equivocarse en cuantas conclusiones se expongan o vaticinios se hagan, motivo por el cual resulta muy conveniente, antes de seguir, ampararse en las Leyes de Clarke[54]:

1) Cuando un científico veterano afirma que algo es posible, es casi seguro que tiene razón. Cuando afirma que algo es imposible, muy probablemente está equivocado. (A lo que se suma el que el autor de este artículo, si bien bastante veterano, no es científico).

2) La única manera de descubrir los límites de lo posible es aventurarse un poco más allá, hacia lo imposible.

3) Cualquier tecnología lo suficientemente avanzada es completamente indistinguible de la magia.

54 La primera ley aparece en el ensayo «Hazards of prophecy: the failure of imagination» que forma parte del libro Profiles of the future (1962). En 1973, en una edición revisada, Clarke desarrolló la segunda y propuso la tercera.

La Ciberseguridad

Una nueva dimensión de la seguridad

En noviembre de 1988, un joven estudiante de informática, Robert Morris, desarrolló como divertimento un sencillo programita de solo 99 líneas. El gusano, al que se bautizó con el apellido de su creador, acabó causando graves destrozos en la ARPANET[55], la precursora de Internet. La «travesura» de Morris puede considerarse el primer ciberataque de la Historia. Hasta entonces, la seguridad de los sistemas se había planteado fundamentalmente en el marco de la seguridad física. Pero quedaba claro que había que contemplar una nueva dimensión. A raíz de este incidente, se creó el primer Computer Emergency Response Team (CERT) en la Universidad de Carneggie-Melon y comenzaron a utilizarse con profusión los términos «seguridad informática» y «ciberseguridad», entendidos ambos como la protección de los sistemas informáticos frente a agresiones, robo o uso no autorizado.

La Ciberseguridad, por tanto, ronda los 30 años de existencia. A pesar de esa considerable edad, parece sufrir del mismo mal que el macho de la especie humana: le cuesta mucho madurar. Y es que, desde sus orígenes, la Ciberseguridad ha ido siempre por detrás de la amenaza. Y, a pesar de los ingentes recursos dedicados a ella, la distancia no se ha reducido. Más bien, todo lo contrario.

Para entender bien la Ciberseguridad hay que conocer la naturaleza tanto del ciberespacio, y las vulnerabilidades que de ella derivan, como la de esas ciberamenazas que tratan de explotarlas, por lo que vamos a dedicar a ello los siguientes apartados.

La naturaleza del ciberespacio

Si al principio de la década de los años 80 el ciberespacio se limitaba a varios millares de computadores, muchos de ellos aislados, las cosas son muy diferentes hoy en día. Se estima que el número de dispositivos conectados a Internet supera ya los treinta mil millones. Y ya no se trata de ordenadores, sino de todo tipo de dispositivos, la mayor parte de ellos englobados bajo el nombre de Internet de las Cosas (*Internet of Things, IoT*), muy heterogéneo grupo que abarca desde electrodomésticos a marcapasos.

55 ARPANET (Advanced Research Projects Agency Network) fue una red de computadoras creada a finales de la década de los 60 por encargo del Departamento de Defensa de los Estados Unidos con el fin de enlazar diferentes instituciones académicas y estatales de investigación.

Como consecuencia, el ciberespacio es hoy un elemento clave para casi todo. La información y los sistemas informáticos que la manejan, almacenan o transmiten resultan cada vez más imprescindibles para infinidad de procesos que afectan a todos los sectores de actividad. Nuevas tecnologías que utilizan el ciberespacio están penetrando rápida y profundamente en todos los aspectos de nuestras vidas. La economía, la comunicación, el transporte, la energía, la sanidad, la seguridad y defensa y hasta las relaciones sociales o el ocio dependen de este nuevo espacio que se convierte en imprescindible.

Pero este escenario también nos condiciona. Así, la creciente conectividad y la ciber-dependencia suponen serias vulnerabilidades y dificultan considerablemente la protección tanto de la información como de los propios sistemas. La superficie de exposición de las potenciales víctimas se agiganta y se complica, incorporando las redes sociales, webs corporativas, la telefonía móvil y ese difuso almacén que es «la nube», hasta el punto que la gran mayoría de las organizaciones desconocen su perímetro en el ciberespacio.

Por otra parte, y a pesar de esa naturaleza virtual que se le atribuye, el ciberespacio se sustenta en elementos físicos, lógicos y humanos, todos ellos imperfectos y salpicados de vulnerabilidades intrínsecas.

Los elementos físicos y lógicos están sujetos tanto al riesgo de fallos que alteren su correcto funcionamiento como a vulnerabilidades que pueden ser explotadas con fines malintencionados. Esta situación se agrava por el uso masivo de elementos comerciales, tanto hardware como software, en cuyo diseño y funcionamiento generalmente prevalecen los criterios de funcionalidad frente a los de seguridad. Lo efímero de estos productos, fabricados muchos de ellos en condiciones (y países) que no ofrecen las debidas garantías, resulta un obstáculo prácticamente insuperable para su adecuada certificación y dificulta enormemente la debida seguridad de sus cadenas de suministro.

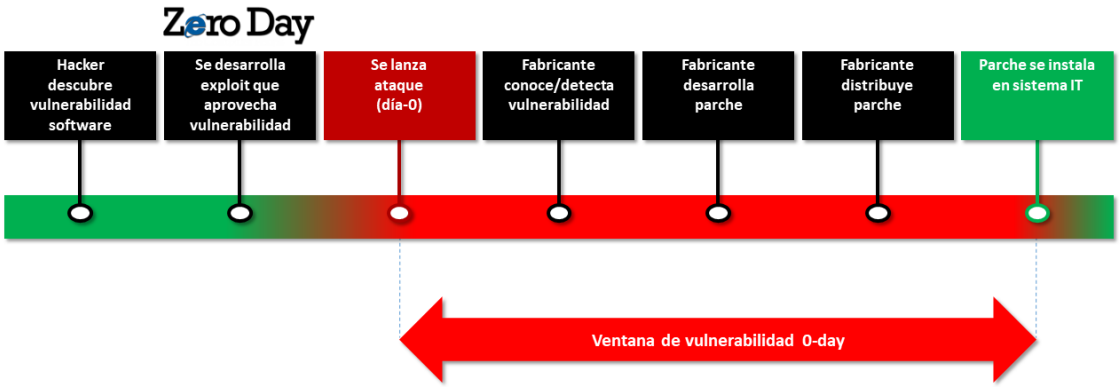
Limitaciones estructurales, resistencia inadecuada, errores de fabricación o de diseño, o emanaciones electromagnéticas no deseadas o son algunas de las imperfecciones inherentes a los elementos físicos.

En lo que se refiere a vulnerabilidades de los elementos lógicos, la mayoría tiene que ver con deficiencias en la programación o en la configuración. Estas deficiencias pueden permitir, por ejemplo, que los atacantes ejecuten acciones maliciosas a través de funcionalidades que tienen una finalidad legítima, o que puedan abrir «puertas» que no han sido tenidas en cuenta por el programador o los administradores de seguridad. Los resultados, según la intencionalidad del atacante, pueden propiciar desde el acceso no autorizado a la caída del servicio. Para que el lector lego lo entienda, recurramos a un símil. Supongamos una cámara acorazada en la que se guarda algo especialmente valioso. La cámara está protegida con una puerta de acero de gran espesor, que solo puede abrirse con un sistema de triple llave. Los muros son de hormigón, impenetrable.

Pero el arquitecto ha introducido en el diseño un conducto de ventilación que comunica el exterior con la cámara, error que ha pasado inadvertido para todos los responsables. Y por ese conducto resulta que puede penetrar un hombre.

En Ciberseguridad es necesario tapar todos los resquicios por los que un intruso puede entrar en el sistema; y, en el plano lógico, ese «intruso» puede tener forma, en muchas ocasiones, de unas simples líneas de código con las que el atacante atente contra los sistemas operativos, los accesos al sistema o las aplicaciones.

Entre todas las modalidades de ataque a vulnerabilidades lógicas, merecen una especial atención aquellas que explotan vulnerabilidades desconocidas y que se conocen como «ataques de día 0».



Para el sistema objetivo, se abre una ventana de vulnerabilidad que se inicia en el momento en que se lanza un ataque que aprovecha la brecha de seguridad y que finaliza el día en que se instala el parche que la elimina. Volviendo a nuestro símil, sería el período que va desde que alguien descubre la entrada al conducto de ventilación que lleva a la cámara acorazada hasta que esa entrada se cubre con barrotes. En Ciberseguridad, no es raro que este período de grave exposición al riesgo sea de varios meses, incluso para vulnerabilidades críticas.

La demanda de vulnerabilidades «día cero » ha crecido mucho en los últimos años, hasta el punto de que por el descubrimiento de algunas de ellas puede llegar a pagarse cantidades en dólares o euros de hasta siete cifras[56]. Es el caso, por ejemplo, de las vulnerabilidades críticas asociadas a sistemas operativos Windows para servidores. Por otra parte, existen estadísticas que demuestran que la gran mayoría de los ataques exitosos aprovechan vulnerabilidades para las que ya existe un parche que las solventa, pero sin que éste haya sido instalado por los responsables de seguridad del sistema (y aquí enlazamos con las vulnerabilidades humanas).

El elemento humano que interviene activamente en los procesos también introduce nuevos riesgos. Es muy conocido el dicho que una cadena es tan resistente como lo es su eslabón más débil. Pues bien, existe un consenso generalizado en identificar

56 https://www.zerodium.com/program.html, fecha de la consulta 07.05.2019.

al elemento humano como el eslabón más débil de la cadena de la Ciberseguridad. Se estima que las personas (administradores, mantenedores, usuarios) intervienen de manera decisiva en torno al ochenta por ciento de los ciberincidentes[57], bien por desborde de trabajo (mal común entre los administradores), desconocimiento, falta de concienciación o de manera deliberada (trabajadores descontentos o agentes hostiles con acceso al sistema «desde dentro»).

Además, el traslado al ciberespacio de muchas actividades, incluidas las relaciones sociales, provoca una sobreexposición que incrementa ese riesgo, facilitando la identificación y análisis de los objetivos, las actividades de ingeniería social y posibilitando nuevos vectores de ataque.

Persiguiendo sombras

Uno de los pilares de la seguridad es la disuasión. La gran dificultad que supone actualmente la atribución y señalamiento de los atacantes provoca que una de las dimensiones de la disuasión, la disuasión por represalia, resulte todavía inefectiva. Ello obliga a concentrar los esfuerzos disuasorios en la negación, en tratar de dificultar al máximo el éxito de los atacantes. Atacantes que se aprovechan de la enorme facilidad que el ciberespacio concede al anonimato, la suplantación y el acceso remoto desde cualquier lugar del mundo. Y eso sin entrar en el débil y confuso marco jurídico que rodean a todas las actividades en y a través del ciberespacio; si la legislación suele ir siempre bastante por detrás de los acontecimientos, este retraso alcanza su máximo grado cuando en el asunto a regular intervienen las nuevas tecnologías.

Este escenario obliga a las defensas a evolucionar continuamente, para ir adaptándose a una amenaza que ha llevado siempre la iniciativa y que se multiplica por el efecto llamada que genera su alto grado de impunidad. Todo ello, mientras la superficie a defender crece y se complica cada día, a medida que los sistemas digitales van apoderándose de procesos asociados a infraestructuras y servicios y se incrementa el número de dispositivos, personas y aplicaciones conectados a la red (telefonía móvil, Internet de las Cosas, la nube, redes sociales).

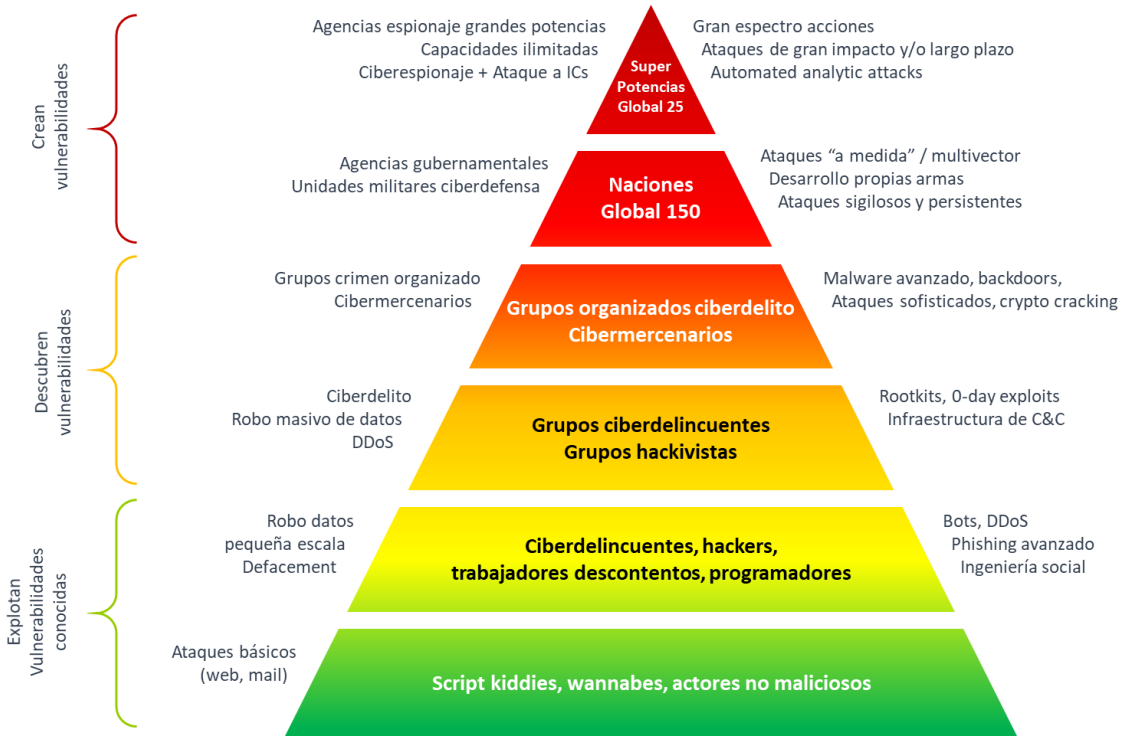
Las ciberamenazas

El «ecosistema» de las ciberamenazas resulta de lo más heterogéneo, tanto en lo que se refiere a formas de actuación como a capacidades y motivaciones. En la cúspide de

57 https://www.abc.es/tecnologia/redes/abci-80-por-ciento-ciberataques-responden-fallos-humanos-seguridad-201605121355_noticia.html, fecha de la consulta 09.05.2019.

ese ecosistema estarían las todopoderosas organizaciones de inteligencia de las grandes potencias, a través de lo que se conoce como Amenazas Persistentes Avanzadas (APT) y cuya principal orientación es el ciberespionaje. Estos grupos cuentan con grandes recursos de todo tipo, que se traducen en importantes capacidades para el desarrollo de ataques quirúrgicos y sigilosos, dirigidos y diseñados a medida del objetivo. Sin olvidar que muchos estados cuentan ya con unidades militares especializadas para las operaciones en el ciberespacio, incluyendo las ofensivas.

Pero también encontramos ciberdelincuentes que exploran incesantemente modelos de negocio más lucrativos, con esquemas propios del crimen organizado; grupos terroristas con intención de realizar ciberataques o que aprovechan el ciberespacio para labores de adoctrinamiento, captación, financiación y propaganda; así como grupos hacktivistas que buscan la notoriedad a través de acciones en la red, y organizaciones privadas que buscan la manera de perjudicar a sus competidores.



Tampoco se puede menospreciar la muy seria amenaza que entrañan los elementos descontentos o infiltrados en las propias organizaciones, conocidos como insiders, o el riesgo potencial que suponen actores individuales menos relevantes, como los black hat hackers, y hasta los wanabees o script-kiddies, especialmente frente a sistemas o información indebidamente protegidos[58].

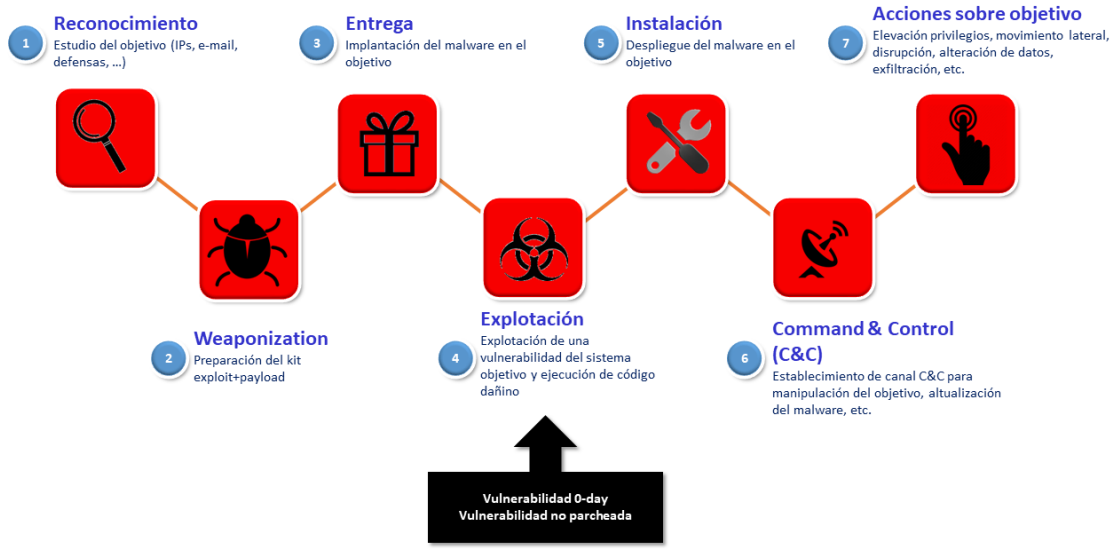
58 https://www.eleconomista.es/tecnologia/noticias/7511916/04/16/Casi-la-mitad-de-los-

A ello se une el incremento continuado del «crime as a service», lo que implica que gran parte de las amenazas se constituye a partir de la subcontratación de especialistas.

Y llegados a este punto, es necesario resaltar el cada vez más profuso empleo que del ciberespacio está haciendo un creciente número de actores, recogidos bajo el título genérico de «amenaza híbrida», entre cuyas principales formas de actuación se encuentran los ciberataques, la desinformación y la propaganda.

A partir de aquí, existen infinidad de modalidades de ataque a la confidencialidad, disponibilidad e integridad de la información y de los propios sistemas que la gestionan. Entre las más habituales, podemos destacar el ransomware[59], la denegación de servicios, la intrusión, el robo de datos o información, la alteración de contenidos, el malware dirigido a sabotear el normal funcionamiento de servicios o infraestructuras o ataques dirigidos a atentar contra la privacidad de los ciudadanos.

A pesar de esa gran variedad de modalidades de ataque, prácticamente todas siguen un patrón similar, conocido como killchain. Los ataques más sofisticados (por ejemplo, los desarrollados por grupos APT con fines de ciberespionaje) cuentan con todas las fases, siendo posible que ataques menos elaborados obvien alguna de ellas.



Así, por ejemplo, un grupo APT estudiará concienzudamente a su objetivo, con el fin de lanzar un ataque perfectamente «a medida» de la víctima para asegurarse el éxito y el sigilo. Por el contrario, otros actores lanzarán ataques masivos e indiscriminados, en la seguridad de que conseguirán siempre un suficiente porcentaje de éxito que rentabilizará con creces el esfuerzo y la inversión.

---

ciberataques-a-la-industria-espanola-en-manos-de-hackers-amateurs.html, fecha de la consulta 14.06.2019.

59 Cifrado de los archivos del sistema víctima y requerimiento de un pago, generalmente en criptomoneda, para su descifrado.

Reconocido y analizado el objetivo, un actor amenaza de categoría preparará las «ciberarmas a medida» (exploit + payload) con las que llevará a cabo el ataque y, fruto de ese mismo esfuerzo de investigación, preparará una o varias maneras de implantar ese malware en el sistema objetivo. Por ejemplo, mediante un correo electrónico enviado a algún usuario concreto de ese sistema, que trate sobre un tema de interés para éste (phishing e-mail), habiendo suplantado previamente la identidad del remitente (spoofing). Y es asombrosa la imaginación y perfección «artística» de algunas de las formas con las que atacantes han conseguido penetrar en sistemas casi inexpugnables. Como en tantos otros aspectos de la vida, la realidad supera con frecuencia a la ficción. Animo a los lectores a consultar el famoso «caso Stuxnet» (2010)[60], que es, entre los ciberataques conocidos, uno de los más extraordinarios tanto por la forma en que se consiguió penetrar un sistema completamente aislado y altamente protegido como por la manera de obtener los efectos perseguidos sobre el objetivo, que no era otro que la planta de centrifugadoras para la obtención de uranio enriquecido en la central iraní de Natanz.

En un siguiente paso, el malware se instalará en la red objetivo (de la misma forma en que se instala un software legítimo, pero de forma completamente oculta), estableciéndose a continuación algún tipo de canal con el atacante que permita su actuación remota. Por lo general, los siguientes pasos son el desplazamiento lateral y la escalada progresiva de privilegios, siempre de forma sigilosa para no hacer saltar las alertas, llegando en muchas ocasiones a alcanzar privilegios de administrador de dominio o de la red completa. Los atacantes pueden ir desplegando progresivamente funcionalidades de forma simultánea a su expansión por la red, en busca de los activos de mayor interés según el objetivo que persigan. Por ejemplo, pueden instalar módulos que permitan capturas de pantalla o pulsaciones de teclado; llevar a cabo la activación de micrófonos, cámaras o geolocalización; búsqueda y recopilación de contraseñas, direcciones de correo, documentos o archivos con una extensión determinada; empaquetado y codificación de toda la información recopilada; y exfiltración mediante múltiples vías, enmascarando ese tráfico entre las conexiones legítimas.

En los ataques más desarrollados y complejos, los atacantes incorporan técnicas de evasión avanzadas en cada una de las fases en las que corren algún riesgo: para penetrar defensa perimetral sin ser detectados (por ejemplo, para evitar a los antivirus o firewalls), para la exfiltración de datos o información (empaquetado, esteganografía[61], navegación web), para evitar la ingeniería inversa sobre el malware (cifrado, ofuscación) y para eliminar cualquier tipo de huella y desaparecer sin dejar rastro ni evidencias de

60 https://es.wikipedia.org/wiki/Stuxnet, fecha de la consulta 01.06.2019.

61 La esteganografía (del griego στεγανος steganos, «cubierto» u «oculto», y γραφος graphos, «escritura») trata el estudio y aplicación de técnicas que permiten ocultar mensajes dentro de otros, llamados portadores, de modo que no se perciba su existencia (Ribagorna, G. A., Estévez-Tapiador, J. y Hernández, J., «Esteganografía, esteganoálisis e Internet. Descubriendo el reverso de Internet: web mining, mensajes ocultos y secretos aparentes»). Puede utilizarse, por ejemplo, para ocultar información en un archivo informático de formato tipo imagen.

las que pueda derivar una atribución. Algunas campañas de ciberespionaje conocidas han permanecido activas en las redes objetivo un gran número de años antes de ser detectadas (Red October[62], Careto[63]). Y, en la gran mayoría de los casos, no han podido obtenerse evidencias suficientes para una atribución con fundamento.

Ciberseguridad y Ciberdefensa

Existe una confusión generalizada sobre el significado de los términos Ciberseguridad y Ciberdefensa. En opinión del que esto escribe, la Ciberdefensa es la traslación de la Ciberseguridad al ámbito militar. De forma muy simplista, podemos decir que la Ciberdefensa engloba a la Ciberseguridad (que equivaldría a la Ciberdefensa defensiva) y va más allá, en tanto supone integrar capacidades de ciberinteligencia y ofensivas enfocadas a la acción contra un oponente.

La mayoría de naciones del mundo cuentan ya con capacidades de Ciberdefensa en el seno de sus Fuerzas Armadas. El ciberespacio está ya ampliamente reconocido como el 5º ámbito de las operaciones militares y así se contempla en la doctrina nacional y de la OTAN. Sin embargo, aún queda mucho por madurar; en particular, en el plano doctrinal.

Uno de los aspectos más controvertidos y que más quebraderos de cabeza origina en las organizaciones es la definición de la línea que separa lo que tradicionalmente se ha conocido como Seguridad CIS de la Ciberseguridad (y, por extensión, de la Ciberdefensa en su vertiente defensiva), disciplina que se ha incorporado con posterioridad pero que ha de convivir en estrecha coordinación con la primera.

En el ámbito de las organizaciones de Defensa, existe una corriente doctrinal, con la que el que esto escribe está completamente de acuerdo, que responsabiliza a la Seguridad CIS de lo que se denominan CIS Infraestructure Operations, actividades que se desarrollan siempre en el ámbito de la infraestructura propia y están orientadas fundamentalmente a la provisión del servicio (empleo, gestión, mantenimiento), lo que por lógica implica ciertos aspectos de la seguridad. En ese modelo, la Ciberdefensa se encargaría de la ejecución de las actividades defensivas, de reconocimiento y ofensivas, que por su naturaleza implican siempre la interacción con un adversario. Las doctrinas de la OTAN y de los Estados Unidos se han desarrollado conforme a este modelo.

62 https://securelist.com/red-october-diplomatic-cyber-attacks-investigation/36740/, fecha de la consulta 09.06.2019.

63 https://securelist.com/the-caretomask-apt-frequently-asked-questions/58254/, fecha de la consulta 09.06.2019.

En concreto, la taxonomía de la OTAN[64] distingue cuatro tipos de actividades operacionales en el ciberespacio:

a. CIS Infrastructure Operations.

b. Defensive Cyberspace Operations.

c. Cyberspace Intelligence, Surveillance and Reconnaissance.

d. Offensive Cyberspace Operations.

No existiendo dudas de que las c) y d) están constreñidas al ámbito de la Ciberdefensa, sí que se plantean en cuanto a dónde se establece la línea que divide a las dos primeras.

En este sentido, la taxonomía de la OTAN entiende las primeras[65] como las medidas del día a día enfocadas a garantizar un grado razonable de seguridad de los activos de la red o sistema CIS frente a una amenaza no definida, en tanto que las segundas[66] están orientadas a la misión y buscan contrarrestar una amenaza concreta. Es decir, las primeras tienen carácter eminentemente preventivo, en tanto las segundas son de carácter proactivo y reactivo e implican siempre la existencia de un adversario concreto.

64 IMSM-0123-2018 HIGH LEVEL TAXONOMY OF CYBERSPACE OPERATIONS, de 15 de marzo de 2018.

65 CIS Infrastructure Operations: Actions taken to employ, secure, operate and maintain CIS in a way that creates and preserves data availability, integrity, and confidentiality, as well as user/entity authentication and non-repudiation. CIS Infrastructure Operation may be thought of as traditional J6 activities, to include the building or deployment of networks and systems, the associated threat-agnostic and routine security measures required to maintain a baseline of cyber security, and the day-to-day operation and maintenance of that infrastructure. (Operaciones de infraestructura CIS: acciones tomadas para utilizar, asegurar, operar y mantener los CIS de forma que asegure y preserve la disponibilidad, integridad y confidencialidad de los datos, así como la autenticación y no repudio de usuarios / entidades. Las Operaciones de Infraestructura CIS pueden considerarse como actividades tradicionales J6, que incluyen el levantamiento o despliegue de redes y sistemas, las medidas de seguridad rutinarias y contra amenazas genéricas requeridas para mantener una línea de base de ciberseguridad y la operación diaria y mantenimiento de esa infraestructura). IMSM-0123-2018, «HIGH LEVEL TAXONOMY OF CYBERSPACE OPERATIONS», de 15.03.2018.

66 Defensive Cyberspace Operations: Active and passive measures to preserve the ability to use cyberspace. Defensive Cyberspace Operations involve the mission-focused and threat-specific activities taken to mitigate known risks and defend against adversaries who are executing or are about to execute offensive actions. (Operaciones Defensivas en el Ciberespacio: medidas activas y pasivas para preservar la capacidad de utilizar el ciberespacio. Las Operaciones Defensivas en el Ciberespacio engloban las actividades orientadas a la misión y enfocadas a amenazas específicas para mitigar los riesgos conocidos y defenderse de los adversarios que están ejecutando o pretenden ejecutar acciones ofensivas. IMSM-0123-2018, «HIGH LEVEL TAXONOMY OF CYBERSPACE OPERATIONS», de 15.03.2018.

En cualquier caso, por razones obvias, resulta condición absolutamente necesaria para el éxito que exista una clara delimitación de responsabilidades y una estrecha coordinación entre los elementos responsables de unas y otras.

El OODA loop en Ciberdefensa

En todo proceso en el que hay dos voluntades en contienda, la actividad de cada parte se desarrolla a través de un proceso cíclico que se inicia con la percepción de lo que hace el adversario, de la que se genera una orientación, de la que a su vez nacen decisiones que se transforman posteriormente en acciones, repitiéndose nuevamente el proceso. Este ciclo se conoce como O-O-D-A loop (Observación, Orientación, Decisión, Acción) y su invención se debe a un coronel de la USAF, John Boyd, a partir de su experiencia en la campaña aérea de la guerra de Corea. La idea que preside su teoría es que aquel de los contendientes que ejecute el ciclo de forma más rápida cuenta con ventaja. La rapidez del ciclo, obviamente, depende de la suma de los tiempos requeridos por cada una de las fases, por lo que puede reducirse a partir de un más exacto y reciente conocimiento de la situación, un más rápido análisis de lo observado o perfeccionando los procesos de toma de decisiones y transmisión de órdenes.

Este modelo es perfectamente aplicable a la Ciberseguridad/Ciberdefensa.

Al igual que ocurre en otros campos en los que impera la tecnología, en ambos se ha ido incorporando la automatización a la práctica totalidad de sus procesos. No obstante, aún queda muchísimo por avanzar.

Lo curioso es que la progresiva automatización no ha ido eliminando al elemento humano que, paradójicamente, resulta cada vez más necesario, tanto en términos cuantitativos como cualitativos. De momento, las máquinas no son suficientes por sí solas; solo ayudan a hacer el trabajo. Eso sí, sin ellas no habría nada que hacer.

Comencemos por ver qué son capaces de aportar a la fase de Observación.

Hasta hace unos años, el modelo que predominaba en la protección de activos en el ciberespacio era el de la defensa por capas. Para ello, resultaba fundamental conocer el perímetro del sistema a defender, su arquitectura y sus conexiones, así como la composición y naturaleza de sus activos (*hardware* y *software*).

En ese modelo «clásico» había dos elementos de seguridad fundamentales: antivirus y cortafuegos o firewalls.

En un principio, cuando todavía el malware evolucionaba lentamente, su detección se fundamentaba en la comparación de firmas. Pero esa etapa «feliz», que coincidió con el boom de algunas firmas antivirus, duró tan solo unos años. En octubre de 2017, la compañía española Panda informaba de que cada día se descubrían 285.000 nuevas muestras de malware. No es descabellado pensar que mientras escribo estas líneas (mayo de 2019) ese número haya crecido hasta el medio millón. Y es que

hace años que se abandonó la producción de malware «genérico», y éste comenzó a preparase y ajustarse específicamente para cada campaña o ataque, muchas veces a medida del objetivo concreto. Por tal motivo, el modelo de detección ha tenido que ir adaptándose. Primero, sobre procesos heurísticos: los archivos se analizaban buscando patrones de código que se asemejaran a los utilizados por el malware conocido. Con la incorporación de técnicas de cifrado y ofuscación en la producción de malware, este modelo comenzó a resultar insuficiente y dejó paso a otras formas de detección, como la basada en el comportamiento, y a la inclusión de nuevos elementos en el proceso, como las cajas de arena (sandbox). Y, en los últimos años, una nueva modalidad de malware que rompe con todo lo anterior y que requiere soluciones completamente nuevas: el malware sin fichero, del que hablaré más adelante.

Los cortafuegos y antivirus continúan siendo necesarios. Pero los elementos de ciberseguridad han tenido que ir incrementándose y especializándose para hacer frente a la evolución de la amenaza.

Hoy en día, la protección de un sistema recae sobre infinidad de dispositivos, que ejecutan procesos muy específicos y complementarios. IDS, IPS, NAC, DLP, MDM, … son algunas de las siglas en inglés por las que se conocen estos dispositivos (Intrusion Detection System, Intrusion Prevention, System, Net Access Control, Data Loss Prevention, Mobile Device Management, …). Elementos que es necesario programar y afinar a medida de la red que protegen y, además, alimentar continuamente con la información de inteligencia que obtiene y facilita (muchas veces, previo pago) la comunidad de ciberseguridad: nuevas vulnerabilidades conocidas, listas negras de direcciones IP, dominios maliciosos, firmas de malware, patrones de ataque, etc., que constituyen una ingente masa de datos e información diaria que hay que introducir en ellos, cada una con su formato.

En la fase de Orientación resulta fundamental contar con herramientas capaces de procesar e integrar toda la información que esos elementos desplegados en el sistema proporcionan.

A día de hoy, por lo general, en los Centros de Operaciones de Seguridad (COS, o SOC, por sus siglas en ingles), un número insuficiente de operadores y analistas trabajan con varias decenas de herramientas diferentes, que en muchas ocasiones no se hablan entre sí y que se operan desde distintas consolas, investigando los cientos, cuando no miles, de alertas que cada día se generan, muchas de ellos falsos positivos, desarrollando de manera manual infinidad de tareas tediosas y repetitivas y sin tener una visión clara de conjunto.

Elemento fundamental en un COS es un sistema de gestión eventos e información de seguridad (SIEM, del inglés Security Information and Event Management). La tecnología SIEM resulta de la fusión de dos categorías de productos: los gestores de eventos de seguridad (SEM) los gestores de información de seguridad (SEM). Un SIEM centraliza el almacenamiento y la interpretación de los grandes volúmenes de

datos (big data) en forma de logs[67], permitiendo, con el apoyo de reglas y algoritmos, un análisis unificado de la situación a partir de la generación de alertas. Estas alertas se originan a partir de la detección de tendencias y patrones anómalos que puedan corresponder a un incidente de ciberseguridad.

Como el resto de elementos de cibersegruridad, un SIEM requiere afinamiento y actualización continua, de forma que permita detectar nuevas formas y patrones de ataque. En este proceso, como ocurría anteriormente, es imprescindible contar con numerosas y eficaces fuentes de inteligencia sobre ciberamenazas.

La visión que estas herramientas proporcionan es, no obstante, insuficiente, en tanto solo es la instantánea que corresponde al sistema o sistemas protegidos en un momento concreto. Para ser rigurosos, esto no es exactamente así, en tanto son capaces de echar la mirada atrás, analizando logs pretéritos a partir de indicadores nuevos. En cualquier caso, no proporcionan una visión más allá, ni en lo temporal, hacia el futuro, ni en lo espacial, más allá del perímetro protegido. Por ello, es necesario tener una consciencia situacional de lo que ocurre en el conjunto del ciberespacio para, en la medida de lo posible, anticiparse a la acción del adversario.

Para ello, algunas compañías llevan tiempo trabajando en el desarrollo de soluciones que, con el apoyo de la IA, permitan poner orden y claridad en esa torre de Babel que es hoy en día un COS. Entre ellas, cabe destacar los orquestadores y las soluciones inteligentes de threat hunting, de las que se hablará más adelante.

Pero, a pesar de las graves carencias ya señaladas, es en las fases de Decisión y Actuación donde más queda por conseguir. El Comandante de una operación en el ciberespacio necesita disponer en su centro de mando operacional de múltiples elementos que le permitan, por un lado, tener un adecuado conocimiento de la situación sobre el que tomar decisiones y, por otro, un adecuado flujo de información con sus unidades subordinadas, otros mandos componentes y el nivel superior de mando. Esta tarea se desarrolla desde lo que se conoce como CyOC (Cyberspace Operational Center). Uno de los problemas más complejos de resolver actualmente es el de alcanzar un adecuado conocimiento de la situación en el área de operaciones ciberespacial; conocimiento que se sustenta fundamentalmente en el dibujo operacional (operational picture) en un entorno operativo en el que, a diferencia de los ámbitos tradicionales, la cinemática y la geolocalización del adversario resultan prácticamente intrascendentes.

Por otra parte, el ciclo O-O-D-A en el ciberespacio no se mide en los mismos términos que en el resto de ámbitos. En el ciberespacio, unos minutos, quizás segundos, pueden ser la diferencia entre el éxito y el fracaso. De ahí que, como en ningún otro ámbito de la seguridad, se requieran sistemas automáticos de apoyo para la toma de decisiones y

67 En informática, se usa el término log, historial de log o registro para referirse a la grabación secuencial en un archivo o en una base de datos de los eventos o acciones que afectan a un proceso (aplicación, actividad de una red informática, etc.), constituyendo una evidencia del comportamiento del sistema.

para la ejecución de lo decidido. Esta situación viene derivada, en buena medida, de la complejidad de integrar todos los elementos, factores y procesos que actúan sobre un espacio de naturaleza especialmente opaca, volátil, confusa e impredecible.

Las oportunidades

Tras esta rápida visión, tan superficial como incompleta, de la Ciberseguridad y la Ciberdefensa y el entorno en el que se desarrollan, ha llegado el momento de analizar cómo puede contribuir la IA a mejorar el estado de las cosas.

Y ya anticipo que hay quién cree que, lejos de conseguirlo, su efecto puede muy bien ser el contrario. Pero a esta controvertida cuestión dedicaré un apartado específico.

A grandes rasgos, son varias las áreas en las que la IA comienza a ser imprescindible, y que en parte se han esbozado en el apartado anterior; en algunas de ellas ya se trabaja sobre soluciones que, en mayor o menor medida, incorporan técnicas asociadas a la IA. Entre estas áreas se encuentran, principalmente y sin ánimo de ser exhaustivo:

- La identificación de usuarios.
- La detección y mitigación de vulnerabilidades.
- El desarrollo de *software* seguro.
- La detección de *malware*.
- La detección de ataques.
- La reacción ante ataques.
- La restauración de sistemas.
- La consciencia situacional.
- La toma de decisiones.

En algunas de ellas ya existen avances suficientes como para posibilitar el desarrollo de soluciones comerciales. En otras, aún es escaso el recorrido.

Veámoslo en mayor detalle.

IA para la identificación de usuarios

A medida que la IA penetra en otros campos, es posible ir incorporando algunos de sus avances al mundo de la Ciberseguridad. Seguramente, muchos lectores cuentan ya en alguno de sus dispositivos con sistemas de inicio de sesión basados en IA: huella digital, voz, escaneo de retina, reconocimiento facial. Estos sistemas facilitan la autenticación y suponen un salto de calidad respecto a las engorrosas y siempre inseguras contraseñas.

No obstante, «el lado oscuro» se las ingenia siempre para encontrar la forma de ir sorteando todas estas protecciones. Resulta relativamente sencillo conseguir una huella dactilar o un patrón de voz; y seguro que casi todos recuerdan cómo el agente protagonizado por Tom Cruise en la película *Minority report* consigue engañar al sistema de seguimiento de ciudadanos basado en el reconocimiento de retina.

IA para la detección y mitigación de vulnerabilidades

Como ya hemos visto, algo que resulta clave para proteger un sistema es el conocimiento de sus vulnerabilidades. En este sentido, hemos de distinguir tres grupos, como se explicó anteriormente: a) las vulnerabilidades asociadas a determinados elementos y de las que la comunidad está advertida y para las que existen medidas mitigadoras, b) las conocidas pero para las que no existe aún solución y c) las aún no descubiertas.

Obviamente, el mayor reto los constituyen estas últimas. He de confesar que, tras consultas a expertos del «mundillo» de la Ciberseguridad, no he sido capaz de obtener referencia alguna a estudios o productos concretos enfocados a este campo, que se antoja del mayor interés pero en el que avanzar se adivina extremadamente complicado.

Más sencillo resulta, obviamente, conocer cómo afectan a un sistema los dos primeros grupos. En la obtención de ese conocimiento resultan fundamentales las inspecciones y auditorías. El problema es que éstas, por muy frecuentes que sean, solo muestran el estado del sistema en el momento en que se llevan a cabo. Sin embargo, cada día se publican decenas de nuevas vulnerabilidades, muchas de ellas asociadas precisamente al *software* y *hardware* más común (sistemas operativos, navegadores, paquetes de ofimática, servidores, *switches*, *routers* y hasta los propios dispositivos de seguridad). Para su corrección, los administradores de seguridad han de cruzar la información de los boletines de nuevas vulnerabilidades con las listas de inventario de su red, aplicando las medidas de mitigación que correspondan (si es que estas existen, que no es siempre el caso). Esta situación convierte la labor de los responsables de seguridad de los sistemas en un auténtico infierno, tanto mayor cuanto más complejo y heterogéneo sea el sistema.

Así mismo, la aplicación de parches requiere un especial cuidado, pues la solución a una vulnerabilidad puede crear efectos indeseados sobre otros procesos o aplicaciones, por lo que lo correcto es realizar pruebas en entorno de preproducción antes de aplicarlas a la red. Estas tareas pueden resultar muy complicadas y ocupar varias semanas.

Por lo tanto, la automatización resulta desde hace mucho tiempo absolutamente imprescindible y, para sistemas y redes complejas, la aplicación de soluciones apoyadas en IA se adivina como la única forma de poder acometer estas tareas de forma eficiente en un futuro cercano.

En resumidas cuentas, la IA aplicada a este aspecto podría proporcionar un conocimiento de la situación de seguridad del sistema que fuera dinámico y en tiempo real, apoyaría la evaluación de las medidas de mitigación, reduciendo los tiempos (que son tiempos de exposición a la amenaza), y facilitaría su despliegue.

Por otra parte, ya se ha comentado que los usuarios son una de las principales vulnerabilidades de los sistemas. En este sentido, y dentro de este apartado enfocado fundamentalmente a la prevención, podemos citar como ejemplo de soluciones apoyadas en IA un proyecto en el que trabaja desde hace algún tiempo la compañía española SIA para su empleo en el sector bancario: la aplicación de técnicas de aprendizaje supervisado al perfilado (*profiling*) de usuarios con fin de aplicar medidas de protección personalizadas. Así, mediante el análisis de amplias bases de datos, se han desarrollado algoritmos específicos que permiten asociar un factor de riesgo (*score)* a cada cliente en función de un amplio abanico de parámetros circunstanciales. Este valor puede enfocarse, por ejemplo, a calcular la probabilidad de que un usuario concreto sea objeto de ataques de phishing, permitiendo adoptar las medidas preventivas adecuadas (por ejemplo, incrementando su nivel de concienciación).

IA para el desarrollo de software seguro

El número de vulnerabilidades y correspondientes parches y actualizaciones se reduciría muy considerablemente si el software comercial se diseñara siguiendo los protocolos adecuados. Por desgracia, la seguridad por diseño no ha calado todavía lo suficiente en un campo en el que, a pesar de todo, continúa primando la funcionalidad y en el que la carrera por lanzar productos al mercado antes que la competencia da lugar a que los fabricantes no sometan a sus productos a unas pruebas tan exhaustivas como debieran.

La IA podría auxiliar tanto en el propio desarrollo del producto como en esa necesaria fase de pruebas, propiciando entornos lo más realistas posibles. Entre la batería de pruebas, resultan especialmente importantes los análisis de código en busca de vulnerabilidades, para lo cual existen ya diferentes herramientas que lo hacen de forma automática, pero que la IA podría mejorar.

En cuanto a los entornos, deberían incorporar la actuación de usuarios virtuales, modelados conforme a patrones de comportamiento reales (basados en estudios estadísticos, en los que también tendría cabida la IA), incluyendo el porcentaje adecuado de usuarios torpes – denominados habitualmente «rubias» (blondies), designación claramente sexista y nada apropiada en estos tiempos que corren – y de malintencionados. La IA permitiría, además, completar esta fase en unos tiempos compatibles con el ritmo frenético impuesto por el mercado de estos productos.

IA para la detección de malware

Este es, probablemente, el campo al que más se está enfocando la investigación de IA en Ciberseguridad, junto con el de detección de ataques, íntimamente relacionado. Y es que la detección de código dañino antes de que éste traspase el perímetro ha sido objeto prioritario de atención desde los primeros tiempos de la Ciberseguridad.

Ya se ha explicado el estado de las cosas en cuanto a producción de malware y los distintos modelos aplicados sucesivamente a su detección: firmas, heurística, comportamiento.

Durante algún tiempo se confió en poder obtener soluciones basadas en IA (o, para ser más precisos, en «machine learning» (ML)) que, a través de un entrenamiento exhaustivo, fueran capaces de etiquetar como «limpia» o «potencialmente dañina», y con una tasa de error mínima, cualquier muestra de código que se presentara a la entrada de la red. A medida que ha avanzado la investigación en este sentido, se ha podido ir viendo cada vez con mayor nitidez lo utópico de esa idea. Los complejos problemas y limitaciones a los que se enfrenta esta línea tecnológica se explican más adelante.

En cualquier caso, aceptando de antemano la imposibilidad de que los sistemas detectores de malware alcancen algún día la condición de infalibles, resultan absolutamente imprescindibles como elemento esencial de la primera línea de protección perimetral y su perfeccionamiento de la mano de la IA se presenta como la única manera de mejorar sustancialmente lo existente.

IA para detección de ataques

Este es, sin lugar a dudas, el campo de mayor interés, en tanto afronta el punto clave de la Ciberseguridad. La detección de ataques se sustenta, fundamentalmente, en el descubrimiento de anomalías o de concatenaciones de sucesos que puedan responder a un patrón de ataque. El problema, lógicamente, se presenta ante una nueva modalidad ofensiva, cuya secuencia de eventos no es seguro que sea detectada al no encontrarse adecuadamente etiquetada en el sistema generador de alertas.

En cualquier caso, podemos diferenciar dos niveles de aplicación de la IA para la detección de ataques en curso. Por una parte, y dada su especialización y complementariedad, deberíamos considerar soluciones basadas en IA que fueran aplicables a los diferentes elementos de seguridad de la red. Así, podríamos hablar, por ejemplo, de firewalls o de sistemas de detección/prevención de Intrusiones (IDS/IPS) inteligentes; esto es, gobernados por algoritmos complejos y dotados de capacidad de aprendizaje. De hecho, ya existen en el mercado productos entre cuyas prestaciones ya se señala esa capacidad. En octubre de 2018, la compañía china Huawei anunciaba[68] el lanzamiento del, según la empresa, primer firewall perimetral basado en IA, específicamente diseñado para la detección avanzada de amenazas; en el comunicado, la compañía presumía de un 99 por ciento de éxito en la detección.

En un plano superior, también apoyados en soluciones IA, estarían los elementos integradores de la información procedente de todos esos elementos inteligentes de seguridad, papel que hoy en día corresponde fundamentalmente a los SIEM, dotando al sistema en su conjunto de una capacidad superior de detección y mejorando considerablemente la eficacia de lo hasta hace poco existente. Así, ya es posible encontrar fabricantes que atribuyen esa capacidad a alguno de sus productos. Splunk o IBM QRadar son, según la consultora Gartner[69], los SIEM más avanzados del mercado y sus fabricantes los califican en sus páginas web como SIEM inteligentes.

También existen diversos productos ya maduros enfocados a mejorar la visión de conjunto y a la integración de herramientas especializadas, como es el caso de los orquestadores, e incluso soluciones supuestamente inteligentes que integran orquestadores y SIEM, como es el caso de Helix[70] de la compañía FireEye.

No obstante, aún queda un amplio margen de mejora.

Tradicionalmente, y ante la evidencia de que toda la panoplia de elementos de seguridad resultaba insuficiente ante las amenazas avanzadas, los analistas han tendido a desarrollar sus propios procesos, generalmente manuales, utilizando su particular conocimiento y familiaridad con la red que protegen para crear hipótesis sobre amenazas potenciales. Esta técnica, de carácter marcadamente proactivo e iterativo y casi podríamos decir que artesanal, conocida como *threat hunting*, mejora su eficiencia cuando la búsqueda es parcialmente automatizada o asistida por una máquina. En este sentido, hay numerosas compañías investigando la forma de conseguir el salto cualitativo que supondría una «caza de ciberamenazas» dirigida por inteligencia artificial, por lo que en los próximos años parece probable que vayan apareciendo en

68 https://www.huawei.com/en/press-events/news/2018/10/Huawei-Industry-AI-Based-Firewal, fecha de la consulta 01.08.2019.

69 https://championsg.com/gartner-reveals-the-2018-magic-quadrant-for-siem, fecha de la consulta 01.08.2019.

70 https://cybersecuritynews.es/combinando-la-proxima-generacion-de-siem-con-orquestacion-avanzada-y-seguridad-en-la-nube/, fecha de la consulta 22.06.2019.

el mercado propuestas cada vez más interesantes.

Por otra parte, una adecuada detección de anomalías requiere, y en ello también tendrá mucho que aportar la IA, parametrizar el máximo número de factores que permitan establecer los estados de la red correspondientes a la situación «normal», a fin de detectar como anomalías los que no estén dentro de esos parámetros. Ello implica, por ejemplo, ser capaces de conocer y etiquetar convenientemente los comportamientos típicos de cada usuario (correo, navegación, horas de arranque y apagado de equipos,…), tráfico habitual en la red en cada instante del día, etc., lo cual se antoja harto complicado para redes complejas, habida cuenta de la enorme casuística a la que están sometidas.

Sin embargo (y aquí invoco la 1ª Ley de Clarke), ya existen soluciones comerciales que aseguran haberlo conseguido. Es el caso de Magnifier, de la empresa Palo Alto, que aspira a ser el siguiente estadio en la evolución. Magnifier, según se señala en la web de la compañía californiana[71], es una aplicación de inteligencia artificial basada en la nube que analiza los datos recopilados desde una plataforma de seguridad de nueva generación, realiza perfiles de comportamiento de los usuarios y los dispositivos en la red y detecta anomalías de comportamiento que puedan sugerir la existencia de un ataque en curso.

Con ese objetivo cada vez más generalizado entre los fabricantes de productos de Ciberseguridad de facilitar la labor de operadores y analistas, IBM trabaja intensamente en otra línea interesante como es la integración de su popular SIEM QRadar con otro de sus productos estrella: Watson[72], una inteligencia artificial capaz de responder preguntas en lenguaje natural[73].

IA para la reacción ante ataques

Este apartado se encuentra muy ligado al anterior, en cuanto la detección del ataque ha de ser siempre el paso previo a la reacción.

Ya se cuenta con sistemas comerciales especializados en contrarrestar algunos tipos de ataque, como es el caso de los ataques distribuidos de denegación de servicios (DDoS), que son capaces de, a través del aprendizaje, de discriminar entre el tráfico legítimo y malicioso. También se cuenta ya hace tiempo con eficaces soluciones anti-spam.

71 https://researchcenter.paloaltonetworks.com/2018/01/magnifier-para-el-analisis-de-comportamiento-caza-y-detiene-rapidamente-las-amenazas-de-red-mas-sigilosas/?lang=es, fecha de la consulta 22.06.2019.

72 https://www.ibm.com/watson, fecha de la consulta 29.06.2019.

73 https://www.ibm.com/es-es/marketplace/cognitive-security-analytics, fecha de la consulta 01.07.2019.

No obstante, se trata en ambos casos de modalidades de ataque que responden a patrones relativamente simples, que no es lo habitual en algunas de las amenazas potencialmente más serias.

Se puede argumentar que ya existen fabricantes que incorporan a sus SIEM capacidades automatizadas de respuesta basada en la aplicación de «cuadernos de jugadas» (*playbooks*) que la solución incorpora por defecto, permitiendo a cada equipo modificar los existentes o diseñar otros nuevos en función de su experiencia. No obstante, existe una diferencia considerable entre una respuesta automatizada y una inteligente.

Un sistema verdaderamente completo debería incorporar, además, aspectos como el conocimiento de la situación (*situational awareness*) y posibilidad de presentar la imagen operativa (*operational picture*) a los que hayan de tomar decisiones. También en estos terrenos la IA podría de aportar soluciones eficaces.

IA para la restauración de sistemas

Los argumentos para el empleo de la IA en esta faceta son muy similares a los referidos en el área de detección y mitigación de vulnerabilidades.

Permitiría, como poco, descargar de trabajo a los administradores, especialmente en situaciones críticas como puede ser un ciberataque serio o un fallo catastrófico que obligara a restaurar un gran número de elementos de la red a las condiciones previas.

Mediante soluciones apoyadas en IA se podría, por ejemplo, gestionar de forma flexible los *back-ups* periódicos, incrementando su frecuencia en función del estado de alerta, o determinar el orden de restauración de elementos más adecuado y automatizarlo, reduciendo de forma significativa la posibilidad de fallo y los tiempos de recuperación.

IA para la consciencia situacional

Ya se han esbozado algunos de los obstáculos que plantean las peculiaridades del ciberespacio para la obtención de una adecuada visión del campo de batalla; en particular, para la representación del dibujo operacional y su superposición al del resto de ámbitos operativos, que se sustentan en factores completamente diferentes.

Para paliar esta situación, existen hoy en día diferentes proyectos de investigación que tratan de resolver estas carencias y que se basan en la utilización profusa de nuevas tecnologías, entre las que se encuentra la IA.

Ejemplos de estos proyectos son CySAP (*Cyber Situational Awareness Package*) [74] y el Desarrollo de una Herramienta de Visualización y Trazado de un Ciberataque. El primero de ellos, que integra soluciones ML, se desarrolla en el marco de la Agencia Europea de Defensa (EDA) y está liderado por España a través del Mando Conjunto de Ciberdefensa (MCCD) y la Dirección General de Armamento y Material (DGAM). El segundo fue desarrollado en un marco de colaboración entre el MCCD y la Universidad Politécnica de Madrid con el objetivo primario de apoyar a la consecución de la Consciencia Situacional en el Ciberespacio, utilizando técnicas de ML para predecir (con cierto nivel de certeza) los siguientes pasos que se van a dar en un ciberataque.

Ambos proyectos suponen un considerable avance, junto con otras iniciativas, como pudiera ser CyCOP (*Cyber Common Operational Picture*)[75] desarrollado por la Universidad Politécnica de Valencia en el marco del programa COINCIDENTE, en el que ha colaborado estrechamente el MCCD, y que afronta el problema de una forma innovadora, mediante tecnologías de visualización 3D. Sin embargo, los resultados resultan aún insuficientes. En cualquier caso, se trata de buenos puntos de partida, sobre los que habrá que seguir construyendo (con el apoyo, obviamente, de tecnologías IA/ML).

IA para la toma de decisiones

Íntimamente ligado al anterior, es, probablemente, el campo en el que más queda por hacer y en el que más difíciles se presentan las cosas. La verdadera inteligencia se aprecia por la calidad y prontitud en la adaptación a entornos cambiantes y, muy en particular, ante situaciones nuevas.

Ya existen modelos para la toma de decisiones que están posicionándose en diversos sectores profesionales a través de distintas técnicas de IA (redes neuronales, lógica difusa, técnicas bio-inspiradas basadas en el comportamiento de las hormigas,…). Pero, hasta ahora, se trata de terrenos muy acotados y en los que se cuenta con un sólido y abundante sustrato de datos y modelos de comportamiento.

En el campo de la Ciberseguridad, la complejidad de cada uno de los procesos por separado y de sus múltiples interconexiones lleva a una casi imposible visión del conjunto y de las posibles consecuencias que una decisión que afecta a un proceso o elemento concreto puede tener sobre otros. La transversalidad del ciberespacio complica aún más las cosas, en tanto cualquier suceso que en él ocurra puede tener

74 https://www.eda.europa.eu/info-hub/press-centre/latest-news/2019/01/11/cyber-situation-awareness-package-(cysap)-project-launched-by-three-member-states, fecha de la consulta 23.06.2019.

75 https://www.ccn-cert.cni.es/pdf/documentos-publicos/xi-jornadas-stic-ccn-cert/2614-m31-07-cyber-situational-awaraness/file.html, fecha de la consulta 25/06/2019.

profundas repercusiones en otros muchos ámbitos más allá de la propia red que se protege y llegar a producir efectos colaterales en cascada. Por lo tanto, cada vez es más evidente que en Ciberseguridad no puede hacerse descansar en mentes humanas una eficaz y oportuna toma de soluciones y que es imperioso un salto de calidad que solamente puede venir de una decidida apuesta por la IA. No obstante, todas esas peculiaridades nos llevan a anticipar que este será uno de los campos de actividad en los que más se tarde en contar con soluciones adecuadamente eficientes y completas.

Los obstáculos y riesgos

La imperfección de la IA está escrita en su ADN

Hemos de asumir que la imperfección forma parte del ADN humano. Y esa imperfección se traslada a todo lo que el ser humano produce.

Por lo tanto, los sistemas basados en IA serán, en mayor o menor grado, imperfectos. Quizás logremos una tendencia asintótica a la perfección, pero ésta estará siempre reservada a los dioses.

Desde hace tiempo, hay voces que alertan de lo que puede suponer la irrupción de la IA en nuestras vidas. Voces, fundamentalmente, de pensadores y filósofos, pero también de científicos y especialistas en diferentes campos que advierten de la necesidad, cada vez más urgente, de poner ciertos límites, tanto éticos como legales, a ese avance que parece no tener fin.

También se comienza a alertar de la falsa sensación de seguridad que puede originarse en los niveles de decisión de empresas y organizaciones a partir de un exceso de confianza en los sistemas basados en IA. Confianza que puede traducirse en el abandono de otras tecnologías más tradicionales e, incluso, en la reducción del personal técnico especializado. Y esto podría ser un gran error, como se explicará en los siguientes apartados.

Por otra parte, también parece lógico inferir que un fallo en una súper IA que controle un gran sistema puede tener unas consecuencias muchísimo más graves que las que pueda ocasionar un fallo de orden similar en algún sistema más básico.

Y no hay que olvidar que la IA estará sustentada por sistemas que harán profuso empleo del ciberespacio, por lo que también serán vulnerables a las ciberamenazas. Es decir, habrá que dedicar mecanismos de ciberseguridad para proteger estos sistemas.

Pero, además de estas importantes cuestiones de carácter general, existen otras más específicas.

Las limitaciones en el aprendizaje

Para entrenar a una máquina de clasificación es necesario contar con un suficiente número de muestras. Supongamos que queremos construir un dispositivo que sea capaz de distinguir entre perros y gatos. Hacer algo así exclusivamente con algoritmos no es tan sencillo como pudiera parecer a priori, en tanto no es posible aplicar marcadores infalibles asociados a color, tamaño, pelaje, forma, número de uñas, patas o longitud de los bigotes, por citar algunos de los factores.

Para asegurar la infalibilidad de la máquina (o un porcentaje aceptable de aciertos), sería preciso mostrarle una cantidad ingente de imágenes, categorizadas como perro o gato, que incluyeran todas las razas posibles, incluyendo cruces, y desde diferentes perspectivas y diversas condiciones de iluminación, contraste, intensidad del color, etc. Además, sería necesario que esa categorización fuera perfecta, pues un error en los datos de aprendizaje podría tener consecuencias impredecibles. A mayor y más variado número de muestras confiables, mejor será la capacidad de la máquina para inferir; es decir, para determinar si una imagen nueva corresponde a uno y otro grupo. Pero, aunque entrenáramos la máquina con millones de muestras, sería absolutamente imposible conseguir una tasa de error del 0 por ciento. Para un ser humano, sin embargo, sería relativamente sencillo no cometer ningún error en la clasificación.

Si aplicamos este modelo a la detección de *malware*, nos encontramos con que no se dispone de un número de muestras excesivamente grande con relación al conjunto total; muestras que, además, no siempre están bien etiquetadas. Por otra parte, aunque aparentemente tratamos de distinguir entre dos especies (código bueno o malo), no partimos de dos especies «estables» como ocurre con el perro y el gato, sino en continua y frenética evolución de la mano de un adversario inteligente y con gran capacidad de adaptación (casi podríamos decir que en estos 30 años el *malware* ha evolucionado lo correspondiente al salto protozoo-mamífero).

Ya se han construido máquinas tan inteligentes como para derrotar a un campeón del mundo de ajedrez y a otros juegos de estrategia; pero en estos casos los «combates» tienen unas reglas vinculantes. En el ciberespacio, los atacantes no siguen las pautas ni o aceptan reglas. Incluso pueden cambiar el «tablero» a su conveniencia. Y no estamos hablando solo de reglas «técnicas», sino también éticas y morales.

Por tal motivo, se vaticina que en ciberseguridad las máquinas siempre estarán aprendiendo y su tasa de falsos positivos nunca será lo suficientemente baja como para considerarlos confiables, especialmente ante nuevas modalidades de desarrollo de malware.

La única forma posible de reducir los falsos positivos es incrementando el umbral de detección. Y ello siempre será a costa de aceptar un aumento de la cuota de malware que atraviese la barrera sin ser detectada. Y tan grave puede resultar etiquetar como «limpio» un código malicioso como bloquear el acceso a la red de un tráfico legítimo.

Por otra parte, un etiquetado incorrecto en una muestra podría producir un efecto «bola de nieve», con resultados impredecibles. Y tampoco puede asegurarse que una muestra etiquetada como limpia en un momento dado no se convierta en potencialmente dañina en un futuro.

Por todos estos motivos, se anticipa que el aprendizaje supervisado por humanos deberá mantenerse aún mucho tiempo en este campo y no se descarta que estos esquemas no puedan liberarse nunca de cierto grado de supervisión e interacción humana.

Un adversario inteligente y adaptativo

Ya se ha mencionado que las ciberamenazas han ido siempre por delante de las ciberdefensas, que han ido jugando siempre un papel muy reactivo. Y es que los atacantes encuentran siempre nuevas formas de evitar cada uno de los obstáculos y barreras que la Ciberseguridad ha ido incorporando a su arsenal defensivo.

El *malware* sin fichero es un buen ejemplo de lo anterior.

En las campañas de concienciación, se suele alertar a los usuarios de los riesgos de ejecutar adjuntos recibidos por correo o hacer clic sobre vínculos. Y es que el *phishing* vía e-mail ha sido siempre uno de los vectores de ataque más utilizados.

Pero, en los últimos años, se ha visto un importante incremento de lo que se denomina «*malware* sin fichero», hasta el punto que más de un tercio de los ataques exitosos en el 2018 emplearon esta técnica[76]. En este caso, el *malware* no penetra en el dispositivo víctima a través de un documento específico, sino que en realidad se instala dentro de la memoria RAM del propio equipo y se desarrolla con distintos procesos.

Con esta técnica, el *fileless malware* consigue pasar inadvertido tanto para el usuario como para la gran mayoría de las soluciones de ciberseguridad que no estén específicamente preparadas para detectar este tipo de intrusiones.

76 https://cdn2.hubspot.net/hubfs/468115/Campaigns/2017-Ponemon-Report/2017-ponemon-report-key-findings.pdf, fecha de la consulta 03/07/2019.

La IA en el «lado oscuro»

En esa lucha perpetua y constante entre defensas y ataques, no podemos esperar que la IA quede restringida solo al ámbito de los primeros.

Así, los atacantes también tendrían un muy amplio abanico de posibilidades para la aplicación de tecnologías IA (o ML) en su beneficio. Y, de hecho, ya han comenzado a hacerlo, haciendo gala de esa extraordinaria capacidad de adaptación antes mencionada.

Por ejemplo, para generar y distribuir correos electrónicos de *phishing* o *spam* de alta calidad[77], incluso en lenguas minoritarias. Y, por supuesto, para detectar las vulnerabilidades (tanto las propias como las de las potenciales víctimas), para crear y depurar nuevo *malware*, para diseñar técnicas de ataque a medida de los diferentes objetivos y que pasen desapercibidos para los sistemas de defensa, o para incorporar técnicas de evasión avanzadas.

De todo esto podemos deducir el comienzo de una especie de «carrera armamentística» de IA contra IA.

Y en esa carrera, los ciberdefensores tienen la desventaja de ser como los porteros de fútbol: han de tener éxito en todas sus intervenciones. A los atacantes, en cambio, les basta con transformar alguna de las ocasiones. Y en un partido que nunca tiene fin.

Por tal motivo, son muchos los que advierten que la irrupción de la IA en la Ciberseguridad, como mucho, tan solo servirá para mantener las cosas como hasta ahora, y que la ventaja que los atacantes han llevado siempre sobre las defensas seguirá existiendo.

Conclusión

La IA en apoyo a la Ciberseguridad/Ciberdefensa es ya una necesidad manifiesta. La automatización en muchas de las tareas de seguridad de los sistemas lleva años siendo imprescindible, pero ya no es suficiente para garantizar el éxito en la gran mayoría de los procesos: gestión de las vulnerabilidades, bloqueo de *malware*, detección y reacción ante incidentes, recuperación de los sistemas, toma de decisiones,… Es necesario un salto cualitativo que, ahora mismo, solo puede venir de la mano de la IA.

No obstante, lo variable, opaco y complejo del ámbito ciberespacial, con una amenaza en continuo crecimiento - tanto en número como en sofisticación - y extremadamente

77 Adam Varney. Analisys of the Impact of Artificial Intelligence to Cybersecurity and Protected Digital Ecosystems. Utica College, ProQuest Dissertations Publishing. 2019.

adaptativa, complica mucho las cosas. Por todo ello, amparado siempre en la primera Ley de Clarke, me atrevo a vaticinar que aún durante muchos años la intervención humana seguirá siendo tan imprescindible como decisiva.

Y del mismo modo que las defensas pueden obtener grandes ventajas a partir de la IA, resulta evidente que también las obtendrán los atacantes.

Por lo tanto, no parece descabellado pensar que esa interminable partida de ajedrez entre defensas y ataques, que comenzó hace ya tres décadas, continúe jugándose indefinidamente en un tablero cada vez más controlado por la IA, pero nunca de forma completa.

Resulta, pues, evidente que la relación entre IA y Ciberseguridad/Ciberdefensa no es en absoluto perfecta. Pero sí como la de una de esas parejas en las que, a pesar de todos los problemas y desavenencias, uno no puede vivir sin el otro.

Bibliografía:

ALONSO LECUIT, Javier, «Implicaciones sobre el uso de la inteligencia artificial en el campo de la Ciberseguridad», CIBERelcano No.44, Real Instituto Elcano, mayo 2019, disponible en: http://www.realinstitutoelcano.org/wps/portal/rielcano_es/contenido?WCM_GLOBAL_CONTEXT=/elcano/elcano_es/zonas_es/ciberseguridad/ari50-2019-alonsolecuit-implicaciones-uso-inteligencia-artificial-campo-ciberseguridad, fecha de la consulta 12.06.2019.

ESTEVE, Manuel, PÉREZ, Israel, PALAU, Carlos, CARVAJAL, Federico, HINGANT, Javier, FRESNEDA, Miguel A., SIERRA, Juan P. «Cyber Common Operational Picture: A Tool for Cyber Hybrid Situational Awareness Improvement», documento OTAN STO-MP-IST-148, 2018.

GROTH, Olaf, NITZBERG, Mark, ESPOSITO, Mark, «Rules for Robots. Why We Need a Digital Magna Carta for the Age of Intelligent Machines», Revista The Digital Future, 2018.

GUARINO, Alessandro, «Autonomous Intelligent Agents in Cyber Offence», NATO CCD COE Publications, Tallinn, 2013, disponible en: https://ccdcoe.org/uploads/2018/10/2_d1r1s9_guarino.pdf, fecha de la consulta 01.08.2019.

HOROWITZ Michael C., ALLEN, Gregory C. «Artificial Intelligence and International Security», Center for a New American Security, julio 2018.

http://www.ieee.es/Galerias/fichero/docs_trabajo/2019/DIEEET0-2018La_inteligencia_artificial.pdf, fecha de la consulta 24.06.2019.

KNIGHT, Will, «The Dark Secret at the Heart of AI», MIT Technology Review,

abril 2017, disponible en: https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/, fecha de la consulta 27.06.2019.

KOSTOPOULOS, Lydia, «Perspective: 6 Considerations to Prepare for Artificial Intelligence Surprises», diciembre 2018, disponible en: https://www.hstoday.us/subject-matter-areas/cybersecurity/perspective-6-considerations-to-prepare-for-artificial-intelligence-surprises/, fecha de la consulta 11.07.2019.

KUBOVIČ, Ondrej, «Can Artificial Intelligence power future malware?», ESET White Paper, 2018, disponible en: https://www.welivesecurity.com/wp-content/uploads/2018/08/Can_AI_Power_Future_Malware.pdf, fecha de la consulta 26.07.2019.

LAHOTI, Sugandha, «Malware Detection With Convolutional Neural Networks in Python», 2018, disponible en: https://dzone.com/articles/malware-detection-with-convolutional-neural-networ, fecha de la consulta 29.06.2019.

MrHouston.net, ¿Es la Inteligencia Artificial el futuro de la Ciberseguridad?, junio 2018, disponible en: https://mrhouston.net/blog/inteligencia-artificial-y-futuro-de-ciberseguridad/, fecha de la consulta 21.06.2019.

SCHARRE, Paul, FISH, Lauren, KIDDER, Katherine, and SCHAFER, Amy, «Emerging Technologies», Super Soldier Series, octubre 2018.

SCHARRE, Paul, FISH, Lauren, «Human Performance Enhancement», Super Soldier Series, octubre 2018.

SCHARRE, Paul, Horowitz Michael C. «Artificial Intelligence. What Every Policymaker Needs to Know», Center for a New American Security, junio 2018.

TYUGU, Enn, «Artificial Intelligence in Cyber Defense», © CCD COE Publications, Tallinn, Estonia, 2011.

Varios autores, «La inteligencia artificial aplicada a la defensa», Instituto Español de Estudios Estratégicos, 2018, disponible en: http://www.ieee.es/Galerias/fichero/docs_trabajo/2019/DIEEET0-2018La_inteligencia_artificial.pdf, fecha de la consulta 27.06.2019.

VARNEY, ADAM, «Analisys of the Impact of Artificial Intelligence to Cybersecurity and Protected Digital Ecosystems», Utica College, ProQuest Dissertations Publishing, 2019.

VILNA, Giovani, «How AI will help in the fight against malware», disponible en: https://techbeacon.com/security/how-ai-will-help-fight-against-malware, fecha de la consulta 12.07.2019.

WANG, Qinglong, GUO, Wenbo, ZHANG, Kaixuan, Ororbia II, Alexander G., XING, Xinyu, GILES, C. Lee, LIU, Xue, «Adversary Resistant Deep Neural Networks with an Application to Malware Detection», In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, CA,

Aug. 2017 (KDD'17), 9 pages.

Wikipedia, diversos artículos.

YUAN, Zhenlong, LU, Yongqiang, WANG, Zhaoguo y XUE, Yibo, «Droid-Sec: Deep Learning in Android Malware Detection», agosto 2014.

Capitulo V

La inteligencia artificial en el campo de batalla

Ángel Gómez de Ágreda

Resumen

Hace ya tiempo que la inteligencia artificial tiene presencia en el campo de batalla. Ésta va a ser cada vez mayor, tanto en labores propias de combate como en aquellas de apoyo al mismo. La tendencia actual apunta a sistemas de armas cada vez más autónomos pero, sobre todo, a una mayor interconexión y compartición de la información recopilada por múltiples sensores entre todos los actores implicados. Entre estos no se van a encontrar únicamente máquinas, robots, sino también los mismos seres humanos, que estarán vinculados a aquellas a través de interfaces que permitan fusionar las capacidades de la inteligencia del carbono, la nuestra, y la del silicio.

Palabras clave:

sistemas autónomos letales, integración de capacidades, sensorización, conectividad, derechos.

Taking AI into the Battleground

Abstract

Artificial Intelligence has been around for some time in modern battlefields. Its presence will be ever larger, both in its capacity as a combat tool and in combat support operations. Current trends show ever more autonomous weapons systems but, specially, a greater connectivity among all actors and sharing of the information gathered by multiple sensors. These actors will not only comprise machines –robots– but also humans, linked to the former through interfaces which will allow for a fusion of intelligences based on carbon -ours- and silicon.

Keywords:

Autonomous weapons systems, integration of capabilities, sensorization, connectivity, rights.

Introducción

El aspecto más sensible de la inteligencia artificial para uso militar –al menos de cara a la opinión pública– es el de la utilización de la robótica en el campo de batalla. La literatura y el cine han influido mucho en esta percepción apocalíptica de inteligencias artificiales generales, totalmente autónomas y conscientes de ellas mismas rebelándose contra sus creadores humanos. Sin embargo, más allá de visiones noveladas de realidades que no son, ni mucho menos, inminentes, la utilización de sistemas de armas autónomos letales (SALAS) –o «robots asesinos», como algunos prefieren denominarlos no de forma desinteresada– es una realidad que ya está presente en el campo de batalla y en la retaguardia logística.

Apenas es necesario un vistazo a la correlación de fuerzas militares entre los distintos países, a sus intereses geopolíticos y a sus inversiones tanto en el capítulo específico de Defensa como en la investigación y desarrollo de aplicaciones de carácter militar o de uso dual. El examen deja muy claro que las capacidades militares convencionales no han dejado de ser importantes ni es previsible que dejen de serlo. Pero también demuestra que todos los países con posibilidades de hacerlo recurren a formas de guerra consideradas hasta hace poco heterodoxas.

De ahí la proliferación de las estrategias asimétricas, no solo entre los países con menores fortalezas convencionales, sino entre todos los actores. La «guerra de los Toyota»[78] no es exclusiva de terroristas o grupos insurgentes, sino que presenta unas posibilidades tácticas y estratégicas que complementan perfectamente formas de acción más tradicionales. Los conflictos se vuelven híbridos y tienen como escenario una «zona gris»[79] que se mueve a caballo entre la paz y la guerra en una inestabilidad provocada para permitir la «pesca en río revuelto»[80].

En buena medida, la adopción de este tipo de tácticas y de estos medios viene condicionada por los cambios en la sociedad. Desde luego, existe una menor tolerancia

78 FRÍAS, Carlos, «La guerra de los Toyota», Ejército de Tierra español, número 906, octubre de 2016, págs. 32-38, ISSN: 1696-7178. Disponible en: http://www.ejercito.mde.es/Galerias/Descarga_pdf/EjercitoTierra/revista_ejercito/Primer_Premio_2017_LA_GUERRA_DE_LOS_TOYOTA.pdf, fecha de la consulta 01.09.2019.

79 Zona del espectro de los conflictos donde predominan las actuaciones situadas al margen del principio de buena fe entre estados (bona fide) que pese a alterar notablemente la paz no cruzan los umbrales que permitirían o exigirían una respuesta armada.» Estado Mayor de la Defensa, CCDC, PDC-01, «Doctrina para el Empleo de las FAS», disponible en: http://www.defensa.gob.es/ceseden/Galerias/ccdc/documentos/02_PDC-01_xAx_Doctrina_empleo_FAS.pdf, fecha de la consulta 03.09.2019.

80 El Center for Strategic and International Studies (CSIS) mantiene una página web con abundante documentación sobre el asunto en https://www.csis.org/grayzone, fecha de la consulta 01.09.2019.

a las bajas, a los féretros o las bolsas negras –como las que llegaban al aeropuerto de Dover, en Estados Unidos, procedentes de Irak o Afganistán– entre los países del primer mundo. El contraste no puede ser más marcado cuando se compara con la utilización de suicidas como vector de ataque por parte de los grupos terroristas. Y dicho contraste provoca frustración y confusión en cuanto a los medios para acometer una amenaza que se sitúa en un marco moral tan alejado del propio.

Por mucho que se sigue diferenciando entre las víctimas civiles y militares como si estos últimos fuesen elementos fungibles privados de dignidad humana, cualquier baja evitable provoca el rechazo de la opinión pública. El debate ha alcanzado a todos los rincones de la doctrina militar. Todas las operaciones se planifican y ejecutan con ojo y medio puesto en las posibles repercusiones jurídicas y mediáticas que puedan llegar a tener.

Esta aversión a las bajas –en especial a las propias, pero también a las del adversario– favorece la implantación de sistemas robóticos en el campo de batalla. Por un lado, para reducir la presencia humana en el mismo y, por lo tanto, la exposición de los soldados. En segundo lugar, para incrementar los factores de protección con que cuenten las fuerzas todavía presentes en el frente. Las labores de exploración, defensa avanzada y otras con grandes dosis de riesgo se externalizan en sistemas teledirigidos o autónomos en la medida de lo posible.

Finalmente, para aprovechar la precisión de los sistemas de guía automática para minimizar los daños colaterales entre los objetivos a alcanzar en el transcurso de la operación.

En este último aspecto, no obstante, se entra de lleno dentro del terreno de la ética, de las leyes, y de los usos y costumbres de la guerra. Algunos países –encabezados por Estados Unidos– utilizan el argumento de la mayor precisión y capacidad de discriminación de los sistemas digitales para apoyar el uso de los SALAS en el campo de batalla.[81] Otros, la mayoría, mantienen sus reservas respecto a esa capacidad de las máquinas de distinguir entre un combatiente y alguien que no lo es, o de valorar adecuadamente la oportunidad de una acción y sus repercusiones más allá del corto plazo de la ganancia táctica.[82] Es, en cualquier caso, una cuestión de grados en cuanto al nivel de autonomía[83] que se le debe conceder al armamento (o, aunque no vayamos

81 Delegación de Estados Unidos en el Group of Governmental Experts of the High Contracting Parties to the CCW, «Humanitarian benefits of emerging technologies in the area of lethal autonomous weapon systems», 2018. Disponible en: https://www.unog.ch/80256EDD006B8954/(httpAssets)/7C177AE5BC10B588C125825F004B06BE/$file/CCW_GGE.1_2018_WP.4.pdf, fecha de la consulta 01.09.2019.

82 Comité Internacional de la Cruz Roja, «Ethics and autonomous weapon systems: An ethical basis for human control?», Group of Governmental Experts of the High Contracting Parties to the CCW, 2018. Disponible en: https://www.icrc.org/en/download/file/69961/icrc_ethics_and_autonomous_weapon_systems_report_3_april_2018.pdf, fecha de la consulta 01.09.2019.

83 Puede resultar llamativo que, cuando se está demonizando el uso de las minas antipersonal, se

a entrar en el tema, a las máquinas en general)[84].

Los SALAS aportan numerosas ventajas en el combate y en su preparación. Su evolución es continua y las aplicaciones a que se les dedica también cambian constantemente. La única certeza que podemos tener en estos momentos es que, con mayor o menor grado de autonomía, van a formar una parte muy significativa de los ejércitos de los próximos años. Nadie va a renunciar a su utilización, incluso antes de que se haya terminado de regular su uso o su misma definición en las reuniones que están teniendo lugar en la sede de Naciones Unidas en Ginebra dos veces por año.[85] Incluso antes de que estén resueltos los problemas de seguridad asociados a su posible pérdida de control por inmadurez de la tecnología o debido a la acción del enemigo digital o electrónicamente.

Misiones de apoyo al combate

Mucho menos controvertido que el uso puramente agresivo –o defensivo– de la robótica en el campo de batalla será su implantación en labores logísticas o de apoyo. Como relata la profesora Rocío Barragán en su capítulo de este volumen, la integración de toda la información procedente de sensores ubicados en y alrededor del ejército permitirá una gestión muy eficiente de los medios disponibles. La robotización de muchas de las funciones logísticas liberará a buena parte del personal de esas tareas y permitirá un suministro similar al modelo industrial «just in time»[86].

Más allá de las líneas de producción, de los talleres y maestranzas de mantenimiento, o del transporte de materiales en la retaguardia, la presencia de robots en el frente ya empieza a ser habitual en forma de mulas autónomas capaces de acarrear parte del material que requieren los infantes o los miembros de operaciones especiales.

---

pretenda desplegar armamento sobre el que puede llegar a perderse el control en un modo similar. Esta circunstancia no ha pasado desapercibida para los estudiosos del tema, que asimilan ambos sistemas en cuanto a su reglamentación se refiere: GUBRUD, Mark A. «The Ottawa Definition of Landmines as a Start to Defining LAWS», según enviado a la Convention on Conventional Weapons Group of Governmental Experts Meeting on lethal autonomous weapons systems de Naciones Unidas en Ginebra que tuvo lugar entre el 9 y el 13 de abril de 2018. Disponible en: http://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2018/gge/documents/Landmines-and-LAWS.pdf, fecha de la consulta 01.09.2019.

84 Gómez de Ágreda, Ángel «Ethics Of Autonomous Weapons Systems And Its Applicability To Any AI Systems», Telecommunications Policy, pendiente de publicación en 2020.

85 Disponible en: https://www.unog.ch/80256EE600585943/(httpPages)/5535B644C2AE8F28C1258433002BBF14?OpenDocument, fecha de la consulta 01.09.2019.

86 La página web de GEINFOR explica en qué consiste el sistema «just-in-time», Disponible en: https://geinfor.com/business/que-es-el-sistema-just-in-time/, fecha de la consulta 01.09.2019.

Es perfectamente concebible un sistema autónomo de reaprovisionamiento o municionamiento en alta mar que pueda acudir allá donde se vaya a requerir su presencia en función de los datos sobre reservas disponibles.

Labores de protección o, incluso, de mantenimiento de infraestructuras automatizadas están también en las mesas de diseño, cuando no ya plenamente operativas. Se trata de funciones de apoyo a las operaciones que no requieren necesariamente del uso de la fuerza y que son susceptibles de tener un desarrollo más temprano en base a su uso dual en las empresas civiles o en funciones policiales.

Una de las grandes ventajas del rol defensivo de instalaciones propias es la posibilidad, muchas veces, de segmentar el escenario de actuación. Esto es, se puede constreñir el campo de acción del robot a un entorno cerrado predefinido al cual no tenga acceso (legal) ninguna posible víctima inocente. Se trata de un equivalente a la doble valla que sirve de demarcación para la actuación de los perros de vigilancia. Esta circunstancia permite una actuación mucho más autónoma del robot toda vez que las posibilidades de un error trágico se han minimizado.

Precisamente, esa es una de las razones por las cuáles sería mucho más viable la implantación de los primeros robots de combate en labores no letales o logísticas, o en actuaciones dentro de entornos en los que la presencia humana fuera marginal. Por ejemplo, la probabilidad de encontrar civiles no combatientes a partir de determinada cota bajo el mar o en el espacio exterior es ínfima. La segmentación también puede realizarse en función de otros parámetros, como la velocidad del objetivo, su temperatura, etc. Tampoco parece que un ataque autónomo contra un vehículo hipersónico, o uno que sigue una trayectoria parabólica, sea susceptible de causar bajas humanas.

El problema del armamento autónomo no es de tecnología. Los ingenieros irán encontrando soluciones a los retos que se planteen. El verdadero dilema es respecto del uso que se hace de él. Así, mientras que es fácilmente aceptable desde el punto de vista ético que una batería de misiles antiaéreos acometa de forma autónoma aquellos objetivos que se cataloguen inequívocamente como municiones, la continuación lógica natural de la acción, el fuego de contrabatería tendente a destruir el origen del ataque, supone una probabilidad muy alta de afectar a personal ajeno a la contienda y, por lo tanto, debería recibir el visto bueno de un operador humano.

Algunos analistas argumentan como posibles problemas de la automatización y robotización del campo de batalla que la falta de bajas puede reducir el umbral de entrada en el conflicto haciendo la decisión menos onerosa. Otros, hace tiempo que vienen avisando de que será la finalización de los conflictos lo que será más difícil de conseguir[87].

87 PRYER, Douglas A., «La sublevación de las máquinas ¿Por qué máquinas cada vez más perfectas contribuyen a perpetuar nuestras guerras y ponen en peligro a nuestra Nación?», Military Review, mayo-junio 2013. Disponible en: https://www.armyupress.army.mil/Portals/7/military-review/

Integración y conectividad

En cualquier caso, la aparición de las SALAS en el campo de batalla tiene un efecto más disruptor en cuanto a la conectividad que permite y los modos de operación que en cuanto a las capacidades que aporta. La mayor parte de las mismas están ya presentes a través de sistemas de armas operados por humanos por lo que no deberían resultarnos perturbadoras.

Al fin y a la postre, de una manera más o menos discreta, ya se están empleando estas tecnologías por parte de las principales potencias y de aquellos países de menor tamaño que han apostado por ellas tanto en aras de garantizar su seguridad como de potenciar su industria civil y militar. «Rusia está probando tanques autónomos en el campo de batalla de Siria, Estados Unidos ha soltado enjambres de drones en el cielo de California, el Reino Unido quiere usar escuadrones de drones en combate para finales de este mismo año y China está construyendo submarinos no tripulados que serían capaces de llevar a cabo ataques de estilo kamikaze contra buques enemigos»[88].

Si nos fijamos en los argumentos que esgrimen los detractores de este tipo de armamento letal -organismos como el Comité Internacional de la Cruz Roja[89] o la campaña Stop Killer Robots[90] encontramos siempre en el trasfondo la dignidad humana como el bien a proteger. Es la equiparación de la capacidad decisoria de máquinas y humanos lo que repele al instinto de supervivencia de la especie que quiere seguir viéndonos como el centro del universo, la culminación de la evolución animal o el ser elegido para dominar el universo. No se niega la potestad de matar, ni siquiera la de que lo hagan las máquinas, siempre que la decisión esté permanentemente en manos de un humano.

Es la despersonalización lo que preocupa a muchos de estos grupos. De hecho, en ocasiones mezclan sus argumentos con los de los aparatos pilotados remotamente, que también parecen alejar anímicamente al operador de la víctima (algo que también ocurre, si se quiere, en el moderno combate aéreo «más allá del horizonte»)[91]. Ni este

---

Archives/Spanish/MilitaryReview_20130630_art010SPA.pdf, fecha de la consulta 01.09.2019.

88 https://www.politico.eu/article/killer-robots-overran-united-nations-lethal-autonomous-weapons-systems/, fecha de la consulta 01.09.2019.

89 Comité Internacional de la Cruz Roja. Declaración de marzo de 2019 en el Group of Governmental Experts of the High Contracting Parties to the CCW, 2019. Disponible en: https://www.unog.ch/80256EDD006B8954/(httpAssets)/59013C15951CD355C12583CC002FDAFC/$file/CCW+GGE+LAWS+ICRC+statement+agenda+item+5e+27+03+2019.pdf, fecha de la consulta 01.09.2019.

90 https://www.stopkillerrobots.org/, fecha de la consulta 01.09.2019.

91 KORAC, Srdan T., «DEPERSONALISATION OF KILLING. Towards A 21st Century Use Of Force Beyond Good And Evil?», Philosophy and society, volumen 29, número 1, 1-152, enero de 2018. Disponible en: https://doi.org/10.2298/FID1801049K, http://www.doiserbia.nb.rs/ft.aspx?id=0353-57381801049K, fecha de la consulta 01.09.2019.

argumento, ni otros similares van a detener la puesta en funcionamiento de estas armas. Ejemplos históricos lo demuestran fehacientemente. Ya en 1139, el segundo Concilio de Letrán prohibió «en adelante, bajo pena de excomunión, el empleo contra cristianos y católicos de ese arte mortal, tan odioso a Dios, de los ballesteros y arqueros»[92] con el resultado ya conocido y esperable. Las SALAS se emplean y se emplearán cada vez más en función de la ventaja táctica o estratégica que proporcionen.

Como decíamos, la aportación más importante de los SALAS no es su autonomía, sino su capacidad de interconexión. Su «superpoder» es la omnisciencia, el conocimiento absoluto de todo lo que ocurre. Eso permite actuar allá donde se precisa cuando es más eficiente. Para ello no son necesarias capacidades letales distintas de las convencionales, simplemente un alto grado de sensorización del campo de batalla y una infraestructura lógica que permita integrar todos los datos recopilados.

De hecho, no estaremos normalmente hablando de sistemas puramente robóticos del mismo modo que ya es muy extraño hablar de combatientes humanos no asistidos de alguna manera en el campo de batalla. El futuro se presenta en forma de sistemas mixtos, de humanos de capacidades aumentadas o reforzadas por las máquinas.

Esta interconexión sí aporta, casi por sí misma, una capacidad que, sin ser específica de estas tecnologías, alcanza con la robotización un grado inaudito de sofisticación y poder: el swarming o actuación en enjambre. Michael Crichton ya exploraba esta posibilidad en su novela «Presa», en 2002. En la ficción, la combinación de la nanotecnología con la solución del problema del almacenamiento de energía para mover drones diminutos capaces de coordinarse entre ellos resulta de una letalidad aterradora.

Estados Unidos tiene, ya en la actualidad, varios proyectos relacionados con esta capacidad para operar «en enjambre»:

- El programa Gremlins de pequeños UAV (Unmanned Air Vehicle, vehículos aéreos no tripulados) lanzados desde otras aeronaves y dirigidos por medio de inteligencia artificial. Podrían recuperarse desde otro avión para su reutilización hasta 20 veces reduciendo los costes.[93]
- El LOCUST, un vehículo aéreo no tripulado con tecnología de «swarming», pretendería saturar las defensas aéreas enemigas con un gran número de blancos.[94]
- Perdix es un prototipo de UAV desarrollado por el prestigioso Instituto Tecnológico de Massachussets (MIT) que ya ha sido probado. Se trata de un

92 Literalmente: «Artem autem illam mortiferam et deo odibilem ballistariorum et sagittariorum adversus christianos et catholicos exerceri de cetero sub anathemate prohibemus.» Disponible en: http://www.clerus.org/bibliaclerusonline/es/index3.htm, fecha de la consulta 01.09.2019.

93 https://www.darpa.mil/program/gremlins, fecha de la consulta 01.09.2019.

94 https://www.youtube.com/watch?v=I47NaccMVnU, fecha de la consulta 01.09.2019.

pequeño aparato del tamaño de la mano que puede lanzarse desde el dispensador de bengalas de un caza. Protegido en un envoltorio, una vez se despliega conecta con el resto del enjambre y ejecuta las órdenes recibidas. Ya se han conseguido formaciones de más de cien aparatos lanzados desde tres F/A-18.[95]

- CICADA (Close-in Covert Autonomous Disposable Aircraft) es un veterano con más de una década en servicio. Un diminuto planeador con una trayectoria en espiral predefinida en su diseño aerodinámico, se trata básicamente de un circuito impreso en un avión «de juguete». Sus misiones típicas, de determinación de condiciones ambientales -meteorológicas o NBQ- y guerra electrónica son, sin embargo, de gran importancia para poblar las bases de datos sobre el terreno a cubrir. Lanzados desde cualquier plataforma, se consideran de un solo uso[96].

El antiguo piloto de bombarderos de la USAF y novelista Dale Brown describe en la mayor parte de sus obras algunos ejemplos -futuristas, pero realistas- de usos de aparatos similares.

Mucho más inminente y similar a la idea que describíamos antes es la Integrated Tactical Network (ITN) que ya está siendo integrada en la 1ª Brigada de la famosa 82 División Aerotransportada[97]. La idea de una red resiliente que combine todas las formas de comunicación entre sus integrantes y sea capaz de servir de base para la transmisión de los datos obtenidos por los sensores que porten sus componentes puede parecer simple, pero las posibilidades que ofrece son múltiples. La garantía de conocer en todo momento la situación de todas las fuerzas, su estado y cómo les afecta el entorno en el que están es solo una de ellas. Los combatientes podrían acceder a todo un catálogo de capacidades en una suerte de reach back inmediato que potenciaría exponencialmente sus posibilidades de actuación.

Cualquier miembro del equipo podría solicitar y recibir en tiempo real apoyo de todas las capacidades de la Brigada que estuvieran disponibles. El famoso «cabo estratégico»[98] del que hablaba el general del Cuerpo de Marines Charles Krulak adquiere, con estas competencias, un significado totalmente distinto. No es ya que su actuación pueda tener una repercusión estratégica o política desproporcionada a su puesto en la jerarquía, sino que la misma jerarquía se diluye para colocar a cada miembro del equipo en una posición de autoridad que, entre otras cosas, va a requerir de una formación exquisita de cada componenete humano del equipo.

95 https://www.youtube.com/watch?v=I47NaccMVnU, fecha de la consulta 01.09.2019.

96 https://defense-update.com/20160419_cicada.html, fecha de la consulta 01.09.2019.

97 https://www.foxnews.com/tech/army-details-future-tactical-war-network, fecha de la consulta 01.09.2019.

98 El concepto de cabo estratégico lo acuñó el general Krulak en el título de un artículo de 1999 de la revista «Marines Magazine», en el que trataba sobre la «guerra de los tres bloques», disponible en: https://apps.dtic.mil/docs/citations/ADA399413, fecha de la consulta 01.09.2019.

Un ejemplo de esta aproximación ya en desarrollo real es el Squad X que está coordinando la agencia DARPA (Defense Advanced Research Projects Agency) del Departamento de Defensa de los Estados Unidos. De nuevo, más que la incorporación de robots al campo de batalla -que también están incluidos- se trata de fusionar de forma rápida e intuitiva para el combatiente la miríada de datos que los sensores ponen a su disposición para acelerar el ciclo de inteligencia y toma de decisiones[99].

Siguiendo un modelo similar al de los modernos aviones de combate (con el F-35 JSF como icono), el Squad X pretende proporcionar información relevante al pelotón sobre la situación táctica en su entorno de relevancia. Así, sensores que pueden estar o no ubicados en la zona captan datos que son procesados para dar al soldado una conciencia situacional superior. Esto incluye un posicionamiento de alta precisión incluso en circunstancias en las que el acceso al GPS esté degradado, bien por cobertura, bien por interferencias causadas por el adversario o por las fuerzas propias. La integración de las referencias proporcionadas por todos los sensores permite un posicionamiento relativo entre ellos que habilita, por ejemplo, la dirección de tiro.

Aunque la decisión se mantiene en el humano, la imagen global del escenario facilita su recorrido por el ciclo de decisión, el famoso OODA Loop definido en su día por el coronel John Boyd. El ciclo, compuesto por la Observación, la Orientación (o contextualización), la Decisión y la Acción, se acelera cuando la primera etapa viene dada automáticamente por todos los sensores, y la segunda por la integración de los datos y la presentación del resultado después de su procesamiento. El jefe de pelotón se encuentra directamente requerido a efectuar la toma de decisiones y la ejecución de las mismas. El objetivo vuelve a ser llevar a cabo todas estas acciones mientras el adversario sigue recopilando información o intentando darle sentido.

El teniente general Deptula, decano de The Mitchell Institute for Aerospace Studies[100] califica esta capacidad como el siguiente gran paso: «La conclusión es que el siguiente gran paso que va a permitir a los Estados Unidos mantener su ventaja cualitativa es la compartición de información continua (sin costuras) y ubicua». El teniente general hace también referencia al papel que las aeronaves tendrán en lo que denomina el complejo de información, vigilancia y reconocimiento (ISR) formado por sensor-shooter-effector-maneuver[101].

Si bien el F-35, como hemos apuntado, se considera hoy el paradigma de esta integración de datos en tiempo real para facilitar la labor gestora del piloto, también ha sido noticia en los últimos meses por ilustrar uno de los mayores riesgos de la tecnología

99 https://www.foxnews.com/tech/soldiers-use-ai-to-fire-precision-grenades-guide-drone-attacks, fecha de la consulta 01.09.2019.

100 http://www.mitchellaerospacepower.org/, fecha de la consulta 01.09.2019.

101 DEPTULA, David A., «Evolving Technologies and Warfare in the 21st Century: Introducing the ´Combat Cloud´», The Mitchell Institute Policy Papers, volume 4, septiembre de 2016. Disponible en: http://docs.wixstatic.com/ugd/a2dd91_73faf7274e9c4e4ca605004dc6628a88.pdf, fecha de la consulta 01.09.2019.

y, sobre todo, de la dependencia tecnológica respecto de terceros, con independencia de si éstos son aliados o adversarios. La constatación por parte de Noruega de que sus JSF enviaban los datos recopilados durante cada misión a Estados Unidos una vez finalizada la misma supone una intolerable vulnerabilidad y pérdida de soberanía para los usuarios.[102]

Del mismo modo que esta acción está inserta en el «adn» del avión, la posibilidad de actuación sobre los mismos datos en el momento de su presentación al piloto distorsionándolos o denegándole el acceso a los mismos añade un flanco más que es necesario vigilar. En el caso de los vehículos autónomos ya se ha determinado el potencial que tienen para ser hackeados y utilizados contra sus pasajeros o contra el sistema de tráfico en general. Se calcula que, con solo un 10% de vehículos autónomos en las calles de una gran ciudad, un fallo o injerencia en su sistema de navegación que les dejase paralizados crearía un caos circulatorio de grandes proporciones. Evidentemente, el peligro es mucho mayor cuando se trata de vehículos militares armados.

En este sentido, ya se están elaborando las primeras guías para mitigar los efectos de las actividades cibernéticas y electromagnéticas (CEMA) del adversario sobre las fuerzas propias[103].

Carrera tecnológica

Desde luego, no cabe esperar que estas vulnerabilidades básicas estén resueltas antes de que se produzca el empleo masivo de robots en el campo de batalla. De hecho, Paul Scharre, autor del libro «Army of none» [104] y uno de los más reputados analistas de la materia, asegura que ya se ha producido ese despliegue, aunque todavía solamente en casos aislados. Scharre prevé una cierta cautela inicial en la puesta en funcionamiento de los SALAS pero un creciente uso de los mismos de forma indiscriminada si la situación estratégica lo demanda o si se presenta una oportunidad para aprovechar la ventaja tecnológica que suponen[105].

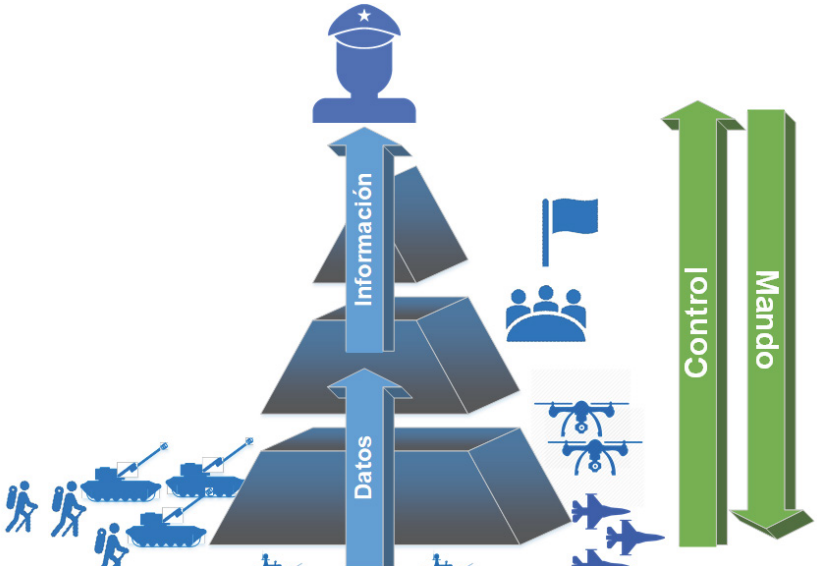
102 Probablemente parafraseando el título de la comedia protagonizada por Nicholas Cage «Army of one», publicada dos años antes. El título hace referencia a un ejército sin humanos, compuesto exclusivamente por robots. Scharre es autor de numerosos artículos sobre SALAS y es considerado uno de los principales referentes del tema.

103 Un ejemplo de estas guías sería ésta publicada por el US Army: https://asc.army.mil/web/wp-content/uploads/2018/08/Autonomous-Robotic-Systems-CEMA-Process-GuideAug2018.pdf, fecha de la consulta 27.09.2019

104 SCHARRE, Paul, «Army of None: Autonomous Weapons and the Future of War», W.W. Norton & Company, 2018, ISBN 978-0393608984.

105 SCHARRE, Paul. «Killer Apps: The Real Dangers of an AI Arms Race». Foreign Affairs, junio

El interés de los países queda claro en el grado de inversión en soluciones de inteligencia artificial y de drones (ver tabla). En ambos casos, Estados Unidos sigue siendo el líder indiscutible en cuanto a inversión con China y otros países como Israel realizando aportaciones importantes e incorporando sus tecnologías entre las más punteras del mercado. El carácter dual de la mayor parte de las tecnologías -que no de la aplicación de las mismas- permite una aportación cruzada entre la industria civil y el sector de la defensa que refuerza los avances en ambas. La modularidad de la mayor parte de los desarrollos permite incorporar estas mejoras incluso cuando no han sido diseñadas específicamente para el sistema al que se aplican.



Fuente: Elaboración propia sobre datos de Siemens recogidos en The Economist.

Otro aspecto relevante -e inquietante para los mandos militares- es la necesidad de mantener un carácter impredecible en las acciones de la inteligencia artificial aplicada a los sistemas de armas. Un comportamiento completamente previsible podría ser derrotado por otra inteligencia artificial o, incluso, por un humano con suficiente visibilidad sobre las actuaciones pretéritas del sistema. La optimización de la respuesta tiene que incluir un número tal de variables a considerar que resulte prácticamente imposible determinar de antemano cuál dará el robot.

Esta circunstancia, que no supone aleatoriedad en su comportamiento, sino una extrema complejidad en la determinación de la acción a tomar, resulta muy incómoda para el planificador humano, que está acostumbrado a que las fuerzas a su cargo cumplan las órdenes con una información contextual limitada. Como en cualquier otro caso, el grado de sofisticación del proceso de decisión del sistema determinará si vencen los ataques o las defensas y, por lo tanto, será preciso mantener un pulso constante en I+D y una férrea disciplina que evite el acceso del enemigo a ella.

de 2019, páginas 135–145. Disponible en: https://www.foreignaffairs.com/articles/2019-04-16/killer-apps, fecha de la consulta 01.09.2019.

Sistemas autónomos aeronáuticos

La alta visibilidad que tienen y la confusión que en ocasiones se produce entre sistemas autónomos y aquellos pilotados de forma remota hace que se relacione SALAS con UAS (*Unmanned Air Systems*) o RPAS (*Remotelly Piloted Air Systems*). Lo cierto es que la fusión de estos últimos con la inteligencia artificial aparece como un paso lógico para los casos en que la aeronave tenga que operar en entornos electrónicos no permisivos o cuando se requiera de una gran coordinación entre varios aparatos.

Desde hace años, vienen existiendo drones aéreos con un alto grado de autonomía. En muchos casos, las misiones que se les asignan son las que requieren largos periodos de vigilancia y objetivos muy concretos. El *Harpi* o el *Harop* (o *Harpi-2*) israelíes son dos ejemplos de *loitering munition*, municiones que se mantienen en espera. En este caso, se trata de aparatos diseñados para la supresión de defensas aéreas enemigas (SEAD). El dron se mantendrá orbitando una determinada zona hasta identificar la activación de un radar enemigo cuya emisión seguirá hasta explotar contra él. Se trata de operaciones de denegación del espacio electromagnético o de supresión de defensas con antelación a un ataque con aviones convencionales en la zona.

En el extremo opuesto del espectro estarían los proyectos para la fabricación de aviones de caza autónomos como el que está desarrollando el *Air Force Research Laboratory's Autonomy Capability Team 3* (ACT3). El escenario más probable es el de una combinación de aviones tripulados y no tripulados en el que estos últimos coordinarían sus acciones como escolta de los primeros o bien como vectores de vigilancia o ataque avanzados. La combinación de ambos tipos de aeronaves, igual que las interfaces entre las capacidades neuronales humanas y cibernéticas, parece estar en todos los estudios serios sobre el futuro de la guerra aérea.

Dentro del programa del Future Combat Air System (FCAS) europeo, se está desarrollando un concepto de colaboración hombre-máquina en el cual se están realizando pruebas para incorporar la capacidad del empleo remoto de drones desde un avión tripulado. Los drones tendrían una capacidad de navegación y guiado autónomo, actuación en enjambre y provisión de información al avión tripulado como si fueran una extensión del mismo. Además del armamento tradicional, estos grupos de combate de los años 40 de este siglo incluirán probablemente armamento hipersónico, cibernético, realidad aumentada y radares avanzados tanto pasivos como cognitivos[106].

106 RANDS, James; TORRUELLA, Anika; NURKIN, Tate y MAPLE, Derrick, «Artificial Intelligence: The Development of key technologies and the barriers to further progress», Jane´s IHS, 29.11.2018.

El Reino Unido está trabajando también en un proyecto denominado Tempest[107], un caza que tendría una escolta de drones inteligentes, mientras que, en Francia, el grupo Dassault está cooperando con Thales en el Plan d'Etude Amont «Man-Machine-Teaming»[108] que pretende incorporar también técnicas de machine learning para facilitar la toma de decisiones fundamentada. Los mismos sensores empleados tendrán capacidad de aprendizaje, de modo que sus observaciones se irán adaptando a las necesidades de la misión según ésta evolucione.

Este entorno, perfectamente reproducible en los ambientes terrestre y naval, implicará una enorme capacidad de adaptación y flexibilidad tanto por parte de las máquinas como de los operadores -en este caso, los pilotos- encargados de explotar su potencial.

Una característica cada vez más común a los sistemas en desarrollo es la capacidad para efectuar una computación «en la nube» que centralice la actividad y la distribuya desde ahí. Será la combinación de sistemas, de formas de obtención de la información y de capacidades cognitivas del silicio y del carbono (humanas) lo que mejor definirá el futuro, y no un campo de batalla uniformemente poblado por un solo tipo de plataforma.

Hemos visto cómo la mayor parte de la inversión seguirá realizándose en drones de gran tamaño y autonomía mientras que, al mismo tiempo, los enjambres de pequeños aparatos -muchas veces de un solo uso- se configuran como dos de las tendencias principales. En Estados Unidos, el Air Force Next-Generation ISR Flight Plan prevé, precisamente la utilización conjunta de todos estos instrumentos para la generación de una imagen global de inteligencia (como la que describe la profesora Barragán en su capítulo).

Otros países, como China, Israel, Corea o India, también están trabajando en proyectos basados en conceptos similares; en algunos casos en proyectos conjuntos en los que se combinan las capacidades de sus respectivas industrias nacionales. Todo ello alterará los modos de operación en el aire-espacio.[109]

En cualquier caso, es preciso enfatizar el hecho de que no todo el sistema de armas tiene que ser necesariamente autónomo. Cada vez más, veremos aplicaciones basadas en la inteligencia artificial formando parte de equipos pilotados por humanos, sea

107 JENNINGS, Gareth, «Tempest's unmanned 'loyal wingmen' to be carrier capable», Jane´s 360, 18 de febrero de 2019. Disponible en: https://www.janes.com/article/86417/tempest-s-unmanned-loyal-wingmen-to-be-carrier-capable, fecha de la consulta 01.09.2019.

108 CHANETZ, Bruno, «Lancement du Plan d'Études Amont Man-Machine-Teaming», 3AF, 21 de mayo de 2018. Disponible en: https://www.3af.fr/article/en-direct/lancement-du-plan-d-etudes-amont-man-machine-teaming, fecha de la consulta 01.09.2019.

109 OSBORN, Kris, «How AI changes attack missions for US fighter jets and bombers», Fox News, 26 de junio de 2019. Disponible en: https://www.foxnews.com/tech/how-ai-changes-attack-missions-for-us-fighter-jets-and-bombers, fecha de la consulta 01.09.2019.

directamente o de forma remota. OceanWatch, por ejemplo, es una carga de pago que puede ir instalada en una aeronave no tripulada del tipo S-100. Su misión es la detección de blancos de pequeño tamaño en la superficie del mar con capacidad para analizar en tiempo real las imágenes y clasificarlas.

Es evidente que muchas de estas aplicaciones pueden ser también útiles en el mundo de la seguridad civil, tanto en su versión de protección frente a amenazas (security) como de seguridad frente a eventos (safety). No sería de extrañar la presencia en las aeronaves de un futuro próximo de cargas de pago de muy pequeño tamaño, pero capaces de llevar a cabo tareas como la detección de incendios, la monitorización de corrientes marinas, el control de los niveles de contaminación, radiación o cualquier otra circunstancia. Su labor se efectuaría de forma absolutamente transparente para la tripulación y el pasaje, y supondría un consumo de energía despreciable para la aeronave nodriza (al tiempo que podrían suponer un servicio público, en el caso de aeronaves de Estado, o una forma adicional de financiación de las privadas).

Esta y otras formas de cooperación público-privada tienen que ser contempladas permanentemente como una realidad necesaria y conveniente. El desarrollo de una industria nacional propia, tanto en lo relativo a la inteligencia artificial y todas las tecnologías asociadas a la misma como en sistemas de seguridad que garanticen su integridad, confidencialidad y disponibilidad continuada será altamente relevante para garantizar la soberanía de los países.

Otros sistemas autónomos

Una cierta equivalencia en el ámbito naval a lo que era en el aéreo la munición de tipo «loitering», la que se mantiene a la espera para efectuar un ataque, serían las minas inteligentes. Diversos modelos adaptados a cada circunstancia están ya en uso en diversas marinas en todo el mundo. Se trata de artefactos explosivos configurables -y reconfigurables en muchos casos- para destruir blancos marítimos específicos.

En primer lugar, incorporan mejoras con respecto a los modelos clásicos en cuanto a los sensores que les proporcionan información. Con una mayor cantidad de datos, más son las posibilidades de definir el objetivo a batir de una forma diferencial.

Curiosamente, tanto en el caso de este armamento naval como de su versión terrestre, la incorporación de nuevas tecnologías viene a compensar el aspecto por el cual muchos países habían decidido su inutilización: la forma indiscriminada de matar sin limitaciones respecto del blanco o del tiempo desde el que fueran plantadas. Si bien las mismas discusiones sobre minas antipersonal se utilizaron en ocasiones como base de partida para argumentar contra los SALAS, la combinación de inteligencia artificial y aquellas puede llegar a resolver, al menos parcialmente, su permanencia activa más allá del periodo de hostilidades.

Junto a las minas inteligentes, el principal desarrollo naval dotado de inteligencia artificial es el UUV (Unmaned Underwater Vehicle), un submarino autónomo que actúa -en la mayor parte de los casos- como un torpedo desde el momento en que detecta su blanco. La autonomía en este tipo de aparatos es particularmente útil al eliminar la necesidad de establecer contacto con el submarino incrementando la probabilidad de que sea detectado. Eso, desde luego, da lugar a otras consideraciones éticas y operativas relativas a la imposibilidad -o dificultad- de revertir una orden una vez dada[110].

Algunos modelos han sido diseñados con propulsión y carga nuclear. El ruso Status-6, también conocido como Poseidón, estaría diseñado para atacar ciudades o instalaciones costeras, bien con la explosión directa, bien provocando un tsunami con su carga estimada entre 50 y 100 megatoneladas[111].

Como era de esperar, los desarrollos basados en la inteligencia artificial y la robótica también se emplean para contrarrestar estas novedosas formas de ataque. El Centre for Maritime Research and Experimentation (CMRE) de la OTAN publicó ya a finales de 2016 un informe sobre «Autonomía colaborativa para contramedidas en la guerra de minas».[112] En él se detallan aspectos del Mine-hunting UUV for Shallow-water Covert Littoral Expeditions (MUSCLE),[113] un futuro vehículo de alta autonomía diseñado para la exploración, detección y clasificación de amenazas y su posterior acometimiento. También se tratan algunos estudios preliminares sobre el REMUS 100,[114] una plataforma de readquisición de objetivos.

En un estado avanzado está también el desarrollo de Knifefish, de General Dynamics, otro sistema antiminas diseñado para ser desplegado desde los Litoral Combat Ships de la Marina de los Estados Unidos[115].

110 Straub, J. «Consideration of the use of autonomous, non-recallable unmanned vehicles and programs as a deterrent or threat by state actors and others». Technology in Society, 2016, número 44, páginas 39–47. Disponible en: https://doi.org/10.1016/J.TECHSOC.2015.12.003, fecha de la consulta 01.09.2019.

111 RAGHEB, Magdi, «Nuclear Ramjet and Scramjet Propulsion». Disponible en: www.mragheb.com, http://mragheb.com/NPRE%20402%20ME%20405%20Nuclear%20Power%20Engineering/Nuclear%20Ramjet%20and%20Scramjet%20Propulsion.pdf, fecha de la consulta 01.09.2019.

112 DUGELAY, Samantha; CONNORS, Warren; FURFARO, Thomas C. y BARALLI, Francesco. «Collaborative autonomy for mine countermeasures», CMRE, 21 de diciembre de 2016. Disponible en: https://www.cmre.nato.int/research/publications/technical-reports/formal-reports/1037-collaborative-autonomy-for-mine-countermeasures-1/file, fecha de la consulta 01.09.2019.

113 https://auvac.org/configurations/view/192, fecha de la consulta 01.09.2019.

114 https://www.kongsberg.com/maritime/products/marine-robotics/autonomous-underwater-vehicles/AUV-remus-100/, fecha de la consulta 01.09.2019.

115 https://www.gd.com/en/Articles/2018/06/04/general-dynamics-team-completes-test-us-navy-knifefish-unmanned-undersea-vehicle, fecha de la consulta 01.09.2019.

Para mayores detalles sobre los sistemas autónomos en el ámbito naval, el capitán de corbeta Jaime Boloix Tortosa analiza en su Trabajo de Fin de Máster del Máster Universitario de Política de Defensa y Seguridad Internacional de la Universidad Complutense de Madrid de 2018 «El impacto de la robótica y la inteligencia artificial en el empleo y la efectividad de la Fuerza Naval».

Mucho menos letales son otros desarrollos que ya están en servicio, como el Pathfinder de LEIDOS, un submarino autónomo o semiautónomo capaz de explorar tanto el lecho marino como las condiciones ambientales.[116]

E, incluso, se han desarrollado soluciones para uso militar -o dual- que incrementan las probabilidades de salvar vidas humanas. El proyecto ICARUS (Integrated Components for Assisted Rescue and Unmanned Search operations),[117] fue financiado por la Unión Europea bajo el paraguas del Séptimo Programa Marco de Investigación e Innovación (FP7) ya en 2014. El objetivo de este proyecto multinacional -en el que participan 10 países- fue el desarrollo de plataformas robóticas que ayuden a detectar, localizar y recatar tripulaciones o individuos en peligro. Las operaciones SAR (Search and Rescue, Búsqueda y Salvamento) -y, presumiblemente, también las CSAR, de SAR de combate- se verían beneficiadas por la integración de las informaciones de numerosos sensores y la alta disponibilidad de los sistemas no tripulados para efectuar el rescate o alguna de las acciones previas al mismo.

Todo ello se combina con grandes avances en cuanto a la sensorización y a la combinación de señales y datos digitales o digitalizados recibidos desde sonoboyas, radares, sonares o desde drones voladores o satélites.

Más allá de lo cinético

Hemos desarrollado la evolución del armamento inteligente de tipo cinético que se empleará, probablemente, en los campos de batalla de los próximos años. Sin embargo, muchos expertos apuntan a una conclusión de mayor calado y que no se debería ignorar. Parece incontestable que en los próximos 20 años la inteligencia artificial y la robótica van a tener un papel importante, incluso clave, en el desarrollo de los conflictos. Pero esta importancia no depende, necesariamente, del desarrollo de robots asesinos del tipo Terminator o de grandes máquinas de matar[118].

116 https://investors.leidos.com/default.aspx?SectionId=5cc5ecae-6c48-4521-a1ad-480e593e4835&LanguageId=1&PressReleaseId=d6121a31-ac43-4187-9de0-d98062c923d1, fecha de la consulta 01.09.2019.

117 https://www.cmre.nato.int/news-room/blog-news-archive/42-rokstories/300-cmre-plays-a-crucial-role-in-enhancing-autonomy-and-integration-between-unmanned-vehicles-as-part-of-the-icarus-sar-project, fecha de la consulta 01.09.2019.

118 O´HANLON, Michael E., «The role of AI in future warfare», Brookings, 29 de noviembre de

La potenciación de las capacidades humanas, la asociación hombre-máquina, la capacidad de integración de conocimiento o la sensorización de cada pequeño dispositivo para ofrecer una imagen altamente detallada de la realidad en cada momento van a alterar de una forma mucho más drástica la forma de hacer la guerra.

La aplicación de la inteligencia artificial a la misma estructura de las máquinas, las plataformas y los vectores también potenciará sus características de una forma asimétrica en función de quién sea capaz de desarrollar esas tecnologías. Actuadores hidráulicos que modifiquen perfiles alares o el blindaje de los carros de combate en función de los requisitos del momento activados por la información proporcionada por sensores y por una capacidad para procesarla de forma instantánea podrán hacer realidad lo que ahora está solamente en las mesas de diseño.

En esta evolución van a tener un papel fundamental los centros de investigación, desarrollo e innovación civiles -financiados o no por los departamentos de Defensa correspondientes- y las grandes corporaciones tecnológicas. Iniciativas como Project MAVEN[119] no van a dejar de producirse por la resistencia de una parte de los empleados de estas empresas a trabajar en proyectos militares.

Una guerra que, como ya hemos empezado a ver en los últimos años, cada vez se lucha menos en los frentes de combate y más en los entornos cognitivo y sensitivo. La misma acumulación de datos a la que acabamos de hacer referencia permitirá una todavía mayor precisión en los ataques al interior de cada individuo, combatiente o no. El principal campo de batalla será el interior de cada uno de ellos, las sensaciones, las emociones, los afectos que muevan las voluntades. Una guerra basada en los afectos o en los sentimientos para la que también será preciso desarrollar el armamento adecuado y en la que el factor humano -paradójicamente, tratándose de una guerra más tecnológica que nunca- será la clave al final.

El futuro

Un dicho moderno atribuido a numerosas fuentes dice que «la gente tiende a sobreestimar lo que se puede conseguir en un año y a infravalorar lo que puede hacerse en diez». Aplicable a numerosos asuntos es, desde luego, perfectamente válido para ilustrar el futuro de la inteligencia artificial y la robótica, y su uso militar en concreto.

2018. Disponible en: https://www.brookings.edu/research/ai-and-future-warfare/, fecha de la consulta 01.09.2019.

119 FRISK, Adam, «What is Project MAVEN? The Pentagon AI project Google employees want out of», Global News, 5 de abril de 2018. Disponible en: https://globalnews.ca/news/4125382/google-pentagon-ai-project-maven/, fecha de la consulta 01.09.2019.

El hecho de que la inteligencia artificial general, la singularidad tecnológica, no deba ser una preocupación inminente no significa que no se deba estar atento a la evolución de los grados de autonomía del armamento con la mente siempre puesta en el mantenimiento del control humano sobre sus actuaciones[120] y, como premisa para poder hacerlo, la capacidad para comprender el razonamiento de las máquinas.

Esta «explicabilidad» de los algoritmos es una preocupación para los ingenieros que desarrollan los sistemas, pero también para los responsables de operarlos o de ponerlos en funcionamiento. DARPA, la agencia del Departamento de Defensa para la Investigación Avanzada, hace años que tiene en marcha un programa en este sentido, XAI (Explainable AI)[121].

Por el momento, el futuro de los desarrollos militares pasa por la cooperación público-privada y por las soluciones comerciales off-the-shelf. Los avances en el desarrollo de microprocesadores especializados, supercomputación y computación cuántica resultarán transcendentales en los próximos años. El uso de redes neuronales y su integración en sistemas hombre-máquina, la nanotecnología y la misma biotecnología van a cambiar no solamente el campo de batalla, sino también a los combatientes.

La aparición de nuevos materiales -como el grafeno o los nanotubos de cobre- serán también altamente relevantes. Una de las principales limitaciones actuales de los sistemas autónomos -o de muchos de ellos- deriva de los problemas para proporcionarles la energía que requieren para incrementar sus tiempos de operación, una de las promesas en cuanto a las ventajas de la máquina respecto del hombre.

La computación en la nube y la computación distribuida se complementarán más que competirán para conseguir sistemas robustos y resilientes en un ambiente que se prevé constantemente degradado. La amenaza para los sistemas no procederá solamente del mundo físico, sino también del lógico, del cognitivo, del cibernético y del electromagnético. En un entorno tan difícil, pasar de la luz a la obscuridad, del control a la impotencia, puede ser cuestión de unos pocos nanosegundos.

La «higiene» de los datos será, por lo tanto, fundamental. Un dato contaminado puede suponer un cálculo fallido o una puerta trasera al conjunto del sistema. Sin embargo, en un campo de batalla que se moverá a velocidades hipersónicas, en el que las decisiones tienen que tomarse de forma instantánea y en el que las crisis pueden escalar en muy breve espacio de tiempo este proceso también tendrá que estar automatizado. En muchos casos, cometiendo «un millón de errores por segundo»[122].

120 LEWIS, L. «Redefining Human Control. Lessons from the Battlefield for Autonomous Weapons. Center for Autonomy and AI», CAN, 2018, páginas 1–23. Disponible en: https://www.cna.org/CAAI, fecha de la consulta 01.09.2019.

121 https://www.darpa.mil/program/explainable-artificial-intelligence, fecha de la consulta 01.09.2019.

122 SCHARRE, Paul. «A Million mistakes a second». Foreign Policy, otoño de 2018, páginas 23–27.

Por el momento, es más probable que encontremos incorporaciones de inteligencia artificial sobre plataformas convencionales (como en el caso ruso, en el que el carro de combate T-14 resulta desproporcionadamente caro en comparación con la incorporación de tecnologías de IA al T-72) que modelos pensados desde el principio para operar de forma autónoma.

En el campo táctico y operacional hay muchos deberes que ir completando. Al desarrollo de dispositivos propios y propietarios hay que añadir la capacidad para defenderse de los de posibles enemigos, adversarios, rivales o competidores.

En el ámbito estratégico y político -aunque extendiéndose también a los subordinados- la principal preocupación puede seguir estando en la utilización de los datos en los entornos sociales y cognitivos. Se trata de un ecosistema adicional que habrá que proteger y explotar, y en el que mantener la soberanía nacional es tan trascendental como en el mundo físico.

Bibliografía

CHANETZ, Bruno, «Lancement du Plan d'Études Amont Man-Machine-Teaming», 3AF, 21 de mayo de 2018. Disponible en: https://www.3af.fr/article/en-direct/lancement-du-plan-d-etudes-amont-man-machine-teaming, fecha de la consulta 01.09.2019.

COMITÉ INTERNACIONAL DE LA CRUZ ROJA, «Ethics and autonomous weapon systems: An ethical basis for human control?», Group of Governmental Experts of the High Contracting Parties to the CCW, 2018. Disponible en: https://www.icrc.org/en/download/file/69961/icrc_ethics_and_autonomous_weapon_systems_report_3_april_2018.pdf, fecha de la consulta 01.09.2019.

COMITÉ INTERNACIONAL DE LA CRUZ ROJA. Declaración de marzo de 2019 en el Group of Governmental Experts of the High Contracting Parties to the CCW, 2019. Disponible en: https://www.unog.ch/80256EDD006B8954/(httpAssets)/59013C15951CD355C12583CC002FDAFC/$file/CCW+GGE+LAWS+ICRC+statement+agenda+item+5e+27+03+2019.pdf, fecha de la consulta 01.09.2019.

Delegación de Estados Unidos en el Group of Governmental Experts of the High Contracting Parties to the CCW, «Humanitarian benefits of emerging technologies in the area of lethal autonomous weapon systems», 2018. Disponible en: https://www.unog.ch/80256EDD006B8954/(httpAssets)/7C177AE5BC10B588C125825F004B06BE/$file/CCW_GGE.1_2018_WP.4.pdf, fecha de la consulta 01.09.2019.

Disponible en: https://foreignpolicy.com/2018/09/12/a-million-mistakes-a-second-future-of-war/, fecha de la consulta 01.09.2019.

DEPTULA, David A., «Evolving Technologies and Warfare in the 21st Century: Introducing the ´Combat Cloud´», The Mitchell Institute Policy Papers, volume 4, septiembre de 2016. Disponible en: http://docs.wixstatic.com/ugd/a2dd91_73faf7274e9c4e4ca605004dc6628a88.pdf, fecha de la consulta 01.09.2019.

DUGELAY, Samantha; CONNORS, Warren; FURFARO, Thomas C. y BARALLI, Francesco. «Collaborative autonomy for mine countermeasures», CMRE, 21 de diciembre de 2016. Disponible en: https://www.cmre.nato.int/research/publications/technical-reports/formal-reports/1037-collaborative-autonomy-for-mine-countermeasures-1/file, fecha de la consulta 01.09.2019.

FRÍAS, Carlos, «La guerra de los Toyota», Ejército de Tierra español, número 906, octubre de 2016, págs. 32-38, ISSN: 1696-7178. Disponible en: http://www.ejercito.mde.es/Galerias/Descarga_pdf/EjercitoTierra/revista_ejercito/Primer_Premio_2017_LA_GUERRA_DE_LOS_TOYOTA.pdf, fecha de la consulta 01.09.2019.

FRISK, Adam, «What is Project MAVEN? The Pentagon AI project Google employees want out of», Global News, 5 de abril de 2018. Disponible en: https://globalnews.ca/news/4125382/google-pentagon-ai-project-maven/, fecha de la consulta 01.09.2019.

GÓMEZ DE ÁGREDA, Ángel «Ethics Of Autonomous Weapons Systems And Its Applicability To Any AI Systems», Telecommunications Policy, pendiente de publicación en 2020.

GUBRUD, Mark A. «The Ottawa Definition of Landmines as a Start to Defining LAWS», según enviado a la Convention on Conventional Weapons Group of Governmental Experts Meeting on lethal autonomous weapons systems de Naciones Unidas en Ginebra que tuvo lugar entre el 9 y el 13 de abril de 2018. Disponible en: http://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2018/gge/documents/Landmines-and-LAWS.pdf, fecha de la consulta 01.09.2019.

JENNINGS, Gareth, «Tempest's unmanned 'loyal wingmen' to be carrier capable», Jane´s 360, 18 de febrero de 2019. Disponible en: https://www.janes.com/article/86417/tempest-s-unmanned-loyal-wingmen-to-be-carrier-capable, fecha de la consulta 01.09.2019.

KORAC, Srdan T., «DEPERSONALISATION OF KILLING. Towards A 21st Century Use Of Force Beyond Good And Evil?», Philosophy and society, volumen 29, número 1, 1-152, enero de 2018. Disponible en: https://doi.org/10.2298/FID1801049K, y http://www.doiserbia.nb.rs/ft.aspx?id=0353-5738180104 9K, fecha de la consulta 01.09.2019.

LEWIS, L. «Redefining Human Control. Lessons from the Battlefield for Autonomous Weapons. Center for Autonomy and AI», CAN, 2018, páginas 1–23. Disponible en: https://www.cna.org/CAAI, fecha de la consulta 01.09.2019.

O´HANLON, Michael E., «The role of AI in future warfare», Brookings, 29 de

noviembre de 2018. Disponible en: https://www.brookings.edu/research/ai-and-future-warfare/, fecha de la consulta 01.09.2019.

OSBORN, Kris, «How AI changes attack missions for US fighter jets and bombers», Fox News, 26 de junio de 2019. Disponible en: https://www.foxnews.com/tech/how-ai-changes-attack-missions-for-us-fighter-jets-and-bombers, fecha de la consulta 01.09.2019.

PRYER, Douglas A., «La sublevación de las máquinas ¿Por qué máquinas cada vez más «perfectas» contribuyen a perpetuar nuestras guerras y ponen en peligro a nuestra Nación?», Military Review, mayo-junio 2013. Disponible en: https://www.armyupress.army.mil/Portals/7/military-review/Archives/Spanish/MilitaryReview_20130630_art010SPA.pdf, fecha de la consulta 01.09.2019.

RAGHEB, Magdi, «Nuclear Ramjet and Scramjet Propulsion», www.mragheb.com. Disponible en: http://mragheb.com/NPRE%20402%20ME%20405%20Nuclear%20Power%20Engineering/Nuclear%20Ramjet%20and%20Scramjet%20Propulsion.pdf, fecha de la consulta 01.09.2019.

RANDS, James; TORRUELLA, Anika; NURKIN, Tate y MAPLE, Derrick, «Artificial Intelligence: The Development of key technologies and the barriers to further progress», Jane´s IHS, 29.11.2018.

SCHARRE, Paul, «Army of None: Autonomous Weapons and the Future of War», W.W. Norton & Company, abril de 2018, ISBN 978-0393608984.

SCHARRE, Paul. «A Million mistakes a second». Foreign Policy, otoño de 2018, páginas 23–27. Disponible en: https://foreignpolicy.com/2018/09/12/a-million-mistakes-a-second-future-of-war/, fecha de la consulta 01.09.2019.

SCHARRE, Paul. «Killer Apps: The Real Dangers of an AI Arms Race». Foreign Affairs, junio de 2019, páginas 135–145. Disponible en: https://www.foreignaffairs.com/articles/2019-04-16/killer-apps, fecha de la consulta 01.09.2019.

STRAUB, J. «*Consideration of the use of autonomous, non-recallable unmanned vehicles and programs as a deterrent or threat by state actors and others*». Technology in Society, 2016, número 44, páginas 39–47. Disponible en: https://doi.org/10.1016/J.TECHSOC.2015.12.003, fecha de la consulta 01.09.2019.

US Army, «Autonomous and Robotic Systems CEMA Process Guide», agosto de 2018. Disponible en: https://asc.army.mil/web/wp-content/uploads/2018/08/Autonomous-Robotic-Systems-CEMA-Process-GuideAug2018.pdf, fecha de la consulta 01.09.2019.

https://www.darpa.mil/program/explainable-artificial-intelligence, fecha de la consulta 01.09.2019.

http://www.clerus.org/bibliaclerusonline/es/index3.htm, fecha de la consulta

01.09.2019.

https://www.darpa.mil/program/gremlins, fecha de la consulta 01.09.2019.

https://www.youtube.com/watch?v=I47NaccMVnU, fecha de la consulta 01.09.2019.

https://www.youtube.com/watch?v=I47NaccMVnU, fecha de la consulta 01.09.2019.

https://defense-update.com/20160419_cicada.html, fecha de la consulta 01.09.2019.

https://www.foxnews.com/tech/army-details-future-tactical-war-network, fecha de la consulta 01.09.2019.

https://www.foxnews.com/tech/soldiers-use-ai-to-fire-precision-grenades-guide-drone-attacks, fecha de la consulta 01.09.2019.

http://www.mitchellaerospacepower.org/, fecha de la consulta 01.09.2019.

https://www.politico.eu/article/killer-robots-overran-united-nations-lethal-autonomous-weapons-systems/, fecha de la consulta 01.09.2019.

https://www.stopkillerrobots.org/, fecha de la consulta 01.09.2019.

https://auvac.org/configurations/view/192, fecha de la consulta 01.09.2019.

https://www.kongsberg.com/maritime/products/marine-robotics/autonomous-underwater-vehicles/AUV-remus-100/, fecha de la consulta 01.09.2019.

https://www.gd.com/en/Articles/2018/06/04/general-dynamics-team-completes-test-us-navy-knifefish-unmanned-undersea-vehicle, fecha de la consulta 01.09.2019.

https://investors.leidos.com/default.aspx?SectionId=5cc5ecae-6c48-4521-a1ad-480e593e4835&LanguageId=1&PressReleaseId=d6121a31-ac43-4187-9de0-d98062c923d1, fecha de la consulta 01.09.2019.

https://www.cmre.nato.int/news-room/blog-news-archive/42-rokstories/300-cmre-plays-a-crucial-role-in-enhancing-autonomy-and-integration-between-unmanned-vehicles-as-part-of-the-icarus-sar-project, fecha de la consulta 01.09.2019.

https://www.csis.org/grayzone, fecha de la consulta 01.09.2019.

https://www.unog.ch/80256EE600585943/(httpPages)/5535B644C2AE8F28C1258433002BBF14?OpenDocument, fecha de la consulta 01.09.2019.

https://geinfor.com/business/que-es-el-sistema-just-in-time/, fecha de la consulta 01.09.2019.

A modo de reflexión final

Ángel Gómez de Ágreda

Nadie puede dudar a estas alturas de que el papel de la inteligencia artificial en la vida social y política está creciendo exponencialmente. Esta presencia se extiende a los aspectos bélicos de la política alterando también su esencia misma. El potencial que tienen los algoritmos para incrementar las capacidades humanas hace que no se pueda prescindir de ellos en un entorno tan competitivo como la guerra.

Los márgenes temporales para la toma de decisiones y para la ejecución de las órdenes se estrechan cada día en función de la automatización de procesos y de la asignación de funciones concretas a las máquinas. Un caza moderno obtiene, por ejemplo, su maniobrabilidad del hecho de ser intrínsecamente muy inestable. Su manejo por parte de un piloto humano sería prácticamente imposible sin el apoyo de las docenas o cientos de ordenadores que lleva a bordo.

Igual que en el caso del caza, la inmensa mayoría de las tareas que desempeñan las tecnologías basadas en la inteligencia artificial resultan transparentes para el usuario. Precisamente, su objetivo es conseguir llevar a cabo su labor con la mínima participación humana. El reconocimiento facial, de voz, de textos, de imágenes, están perfectamente integrados en aplicaciones más complejas y no hacen sino simplificar nuestra experiencia de las mismas. Sin embargo, se han convertido en algo ubicuo de lo que difícilmente estaríamos dispuestos a prescindir.

Estas tecnologías son, por otro lado, perfectamente neutras en cuanto a su caracterización ética. Un programa de reconocimiento de voz no hace más que comparar los registros de entrada con determinadas características que tiene almacenadas en su base de datos. Si el resultado sirve para desbloquear un teléfono móvil, para identificar el blanco de un bombardeo o para determinar las características de una persona con objeto de clasificarla social o laboralmente es algo que no tiene que ver con la función propia de ese sistema de algoritmos concreto.

De hecho, estas tecnologías relacionadas con la inteligencia artificial son casi siempre de uso dual. De alguna manera son piezas de un juego de construcción que se pueden ensamblar para que constituyan un tractor o un carro de combate. Es la aplicación que se extrae de ellas la que encierra un componente ético, no las tecnologías subyacentes.

Incluso se puede proseguir el argumento diciendo que es la forma en que actúa la aplicación la que encierra una calificación moral u otra. Aunque algunas personas argüirían en favor de la atribución de características éticas a una máquina, eso no deja de ser equivalente a la más que olvidada costumbre militar de «arrestar» una puerta, o un fusil, o un vehículo cuando habían estado implicados en un accidente o en un error.

Atribuir valores a los algoritmos no eleva su categoría a la de las personas, sino

que degrada la dignidad humana hasta llevarla a la equiparación con meras máquinas con programaciones más o menos complejas. La discusión no debería centrarse en el carácter ético del armamento autónomo sino en el grado de autonomía que se le puede conceder a cada una de las funciones de una máquina.

Especialmente cuando detrás de esa atribución se esconde la pretensión de permitir que los sistemas de armas autónomos puedan adoptar decisiones sin la participación obligatoria de un controlador humano. No tanto porque sea una máquina la que pueda ejecutar una acción de forma independiente cuanto porque la combinación de la asignación de valores y capacidad de toma de decisiones pretende exonerar al humano de la responsabilidad de dicha acción.

La responsabilidad va –o debería ir– asociada a la autoridad. Delegar la responsabilidad implica renegar de la autoridad, por mucho que no quiera verse así por los proponentes de esta visión.

Y lo que es aplicable a los sistemas de armas autónomos, sean letales o no, debería hacerse extensivo a otras aplicaciones basadas en la inteligencia artificial. No es la acción concreta que lleve a cabo una máquina lo que resulta censurable, sino el hecho de que lo haga de forma autónoma afectando a la vida o a la libertad de los seres humanos.

Pocos en este foro se cuestionarán, por ejemplo, que una maquina pueda recibir la orden de neutralizar un objetivo concreto y lleve a cabo la acción de forma independiente. Al fin y al cabo, eso es lo que hacen los misiles guiados *fire and forget*. Una vez fijado el blanco, identificado y asumida la responsabilidad por parte del operador, la máquina ejecuta eficazmente la tarea sin más necesidad de intervención humana.

Tampoco parece que el debate tenga que incluir a sistemas defensivos del tipo de la artillería contramisil. Resulta evidente que la participación de un operario, por hábil que pueda resultar, no hará más que entorpecer o dilatar el tiempo de reacción ante un ataque de un proyectil o una salva de ellos. Una máquina puede encargarse de identificar el ataque como constituido por un ser inanimado que procede a gran velocidad hacia un objetivo propio y tomar la decisión de abatir el vector.

Otra cuestión sería que esa misma máquina pretendiese adoptar decisiones de fuego de contra-batería disparando a su vez contra el emplazamiento del que ha partido el ataque. En ese caso, la probabilidad de provocar víctimas humanas y de que estas sean, en su totalidad o parcialmente, de las contempladas en los acuerdos internacionales como no combatientes es mucho más elevada y, por lo tanto, esa decisión reviste características morales que deben hacerla recaer en un ser humano.

Incluso atendiendo a argumentos –como los presentados por Estados Unidos o Rusia en las reuniones de Grupo de Expertos Gubernamentales de Naciones Unidas en la Convención sobre Ciertas Armas (CCW) en Ginebra– que indican que los sensores asociados a una aplicación basada en la inteligencia artificial puede llegar a tener mayor precisión en la discriminación entre combatientes y no combatientes, y en su acometimiento, resulta discutible que la decisión pueda obviar la participación

humana.

¿Acaso nos sentimos cómodos con la idea de que, con el fin de disminuir los accidentes de tráfico, sean los algoritmos los que decidan quién sobrevive y quién no cuando la colisión sea inevitable? Más allá del criterio y de la más que dudosa capacidad de los programadores para conseguir que las máquinas sigan el espíritu del paradigma establecido, delegar la decisión en las máquinas ofende a la dignidad de las personas implicadas.

Decía George Orwell que no se trata tanto de estar vivo como de seguir siendo humanos. Si valoramos nuestra libertad hasta el punto de dar la vida por mantenerla, incluso por contribuir a difundirla, quizás deberíamos también ampliar la misma protección que exigimos a las máquinas autónomas letales a aquellas que, sin serlo, sí eliminan o coartan nuestra libertad.

Valorar la letalidad como criterio para regular o prohibir la autonomía de los algoritmos supone infravalorar, bien la libertad misma, bien la capacidad de los mismos para afectar a nuestras decisiones como seres humanos. Lo primero menoscaba nuestra dignidad y nuestra esencia como personas, lo segundo es fruto de una injustificada soberbia como especie dominante en este momento histórico.

Si nos aproximamos a las tecnologías que sustentan la inteligencia artificial desde un punto de vista de la plataforma en la que se apoyan podríamos dividirlas entre las que tienen un soporte lógico, aquellas que tienen un soporte físico y las que lo tienen biológico. Es decir, tendríamos por un lado los algoritmos que no están vinculados a un dispositivo físico concreto –por ejemplo, un chatbot–, en segundo lugar, aquellos que sí lo están y proporcionan la base lógica de un robot, y en tercer lugar aquellos que interactúan directamente con la inteligencia humana a través de interfaces cerebro-ordenador, por ejemplo.

Por mor de darle un nombre a cada categoría, podríamos distinguir entre *soft-AI*, una inteligencia artificial que se mantiene en la esfera del software, *hard-AI*, la que controla robots y otros dispositivos, y una *bio-AI*, que interactúa con los humanos. Como una sub-categoría de la segunda, encontraríamos a los SALAS, los Sistemas de Armas Letales Autónomos, cuya regulación parece avanzar de forma separada a la del resto.

Sin embargo, las mismas prevenciones que parecen lógicas cuando se trata de *robots asesinos* se perciben como ajenas al campo del resto de las tecnologías asociadas a la autonomía de las máquinas. Es como si la lucha de voluntades que supone una guerra tuviera que contar con aspectos cinéticos exclusivamente y no pudiera resolverse actuando sobre los afectos o sobre las percepciones.

El entorno académico de la investigación en tecnologías asociadas a la inteligencia artificial ha producido ya docenas de códigos éticos en los últimos pocos años. Para la mayor parte de los mismos, un aspecto fundamental es el tener en cuenta las prevenciones que expresan en todas las fases de diseño de un sistema. Ya desde la misma concepción de la aplicación deben considerarse todas las posibles derivadas

éticas que pueda hacer surgir, así como las posibilidades de que su propósito inicial pueda verse subvertido debido a una acción exterior –un hackeo– o una falla en la programación.

Estas condiciones tienen resonancias que recuerdan las de industrias punteras con escaso margen de error, como la aeronáutica, en las que la seguridad –en su doble vertiente de *security* ante un acto externo y de *safety* ante un fallo interno– está presente en todas las fases del diseño, desarrollo, utilización y disposición del sistema.

Los usos que las Fuerzas Armadas o, más genéricamente, el Ministerio de Defensa, pretendan hacer de las posibilidades que ofrece la inteligencia artificial deberían partir de estas consideraciones éticas y de estas salvaguardas. En estas últimas tienen que estar, también, muy presentes los condicionantes jurídicos en aquellos aspectos que estén regulados o en los que exista ya una tendencia clara.

Los principios subyacentes a la mayor parte de estos códigos podrían resumirse en:

- La (relativa) beneficencia o bondad del uso de la tecnología
- El respeto por la dignidad humana
- El respeto por la privacidad de los datos
- La preservación de la autonomía de las personas
- La equidad y justicia
- La «explicabilidad» de los algoritmos

En cuanto al primero, resulta evidente que el objetivo marcado por la primera ley de la robótica enunciada por el científico y novelista Isaac Asimov –un robot no hará daño a un ser humano ni, por inacción, permitirá que lo sufra– no es de mucha aplicación cuando el objetivo del robot se convierte precisamente en causar ese daño. Sin embargo, la inclusión de la relatividad en el término permite incluir soluciones que sean menos letales que sus equivalentes no autónomos.

No se puede descartar la adopción de tecnologías autónomas según este criterio. Del mismo modo que los automóviles provocan más de un millón de muertos todos los años, pero también proporcionan enormes beneficios y salvan innumerables vidas, y resulta impensable prescindir de las ventajas a pesar del coste que acarrean.

El respeto por la dignidad humana ya ha sido tratado en estas páginas. Más allá de la categoría intelectual que pueda llegar a alcanzar un diseño humano, la asignación de características propias de las personas a las máquinas supone desvirtuar su esencia y, por lo tanto, privar a los humanos de la misma. Del mismo modo, permitir que los algoritmos tomen decisiones sobre la vida o la libertad de las personas, por mucho que esa decisión sea beneficiosa para el conjunto de la sociedad, es algo que resulta alarmante. ¿Queremos que sea una máquina la que decida si es el pasajero del coche que tripula o el peatón que está cruzando la calle el que muere si la colisión es inevitable?

Las soluciones asociadas a la inteligencia artificial se basan en la acumulación y el

análisis de miles de millones de datos para generar un conocimiento profundo de la realidad. Ese conocimiento tendría que poder limitarse a la función concreta que se va a desarrollar y no estar disponible para ninguna otra. Es preciso acotar el grado de control sobre las poblaciones, sobre las personas individuales, que permitiría que todo ese flujo de conocimiento podría propiciar usos ilegítimos del mismo por parte de gobiernos, empresas o particulares que pudieran acceder a él.

La anonimización de los datos, la posibilidad de establecer una identidad digital única protegida para permitir que el usuario sea el propietario de los datos que genere y otras iniciativas podrían mitigar este riesgo. Aun así, estas medidas solo son posibles si se adoptan desde el inicio del diseño del sistema. Se puede imaginar la gravedad de una situación en la que, para gestionar mejor los recursos de una base o de un buque, se coloquen sensores en todo el personal y el material del mismo, y alguna entidad ajena fuera capaz de acceder a ellos. Las enormes ventajas que se pudieran obtener de la visibilidad que se conseguiría apenas si serían comparables al daño al que expondríamos al grupo y a cada uno de los individuos.

La cuestión de la preservación de la autonomía, no ya de las máquinas, sino de las personas se solapa ligeramente con algunos de los anteriores puntos. La capacidad para ejercer la libertad desde un pensamiento independiente está muy relacionada con la privacidad y termina repercutiendo en los derechos de las personas y, por lo tanto, en su libertad. La tendencia humana a dejarse guiar por las sugerencias de lo que percibe como un agente imparcial hace que, incluso cuando no existe coacción alguna, se sigan las indicaciones de los algoritmos de una forma más o menos ciega.

La guerra siempre ha explotado las vulnerabilidades humanas para alcanzar los objetivos marcados. La transparencia que genera el tratamiento de sus datos en las personas implica que estas vulnerabilidades son mucho más evidentes y fáciles de utilizar. Del mismo modo que la ingeniería social se ha convertido en la técnica más empleada en el ciberespacio, muy por encima de cualquier otra de ingeniería, también la capa humana se convierte en el flanco más accesible para los ataques por parte de la inteligencia artificial.

La equidad y la justicia en el tratamiento de las personas por parte de los procesos automatizados no parecen tener gran relación con los objetivos de la Defensa. No se debe olvidar, no obstante, que las guerras del siglo XXI no tienen lugar en entornos ni en momentos separados del periodo de paz. Se trata de guerras que tienen lugar EN la gente y, muchas veces, utilizando a la población como campo de batalla, como arma y como objetivo.

En estos casos, la privacidad de los datos no es suficiente. El tratamiento de los mismos tiene que garantizar que, ni dentro de las operaciones ni fuera de ellas, no puedan ser empleados para discriminar a unas personas respecto de otras, o a unos grupos en relación con los demás. Igual que hay algoritmos que tienden a identificar como criminales a los miembros de determinados grupos étnicos o a privilegiar la contratación de un género respecto del otro, los mismos sesgos pueden producirse en la asignación de tareas en un buque, en las promociones al siguiente empleo en

una base o en la priorización en la atención médica a determinadas unidades sobre otras. Resulta, no obstante, complejo diseñar sistemas que permitan discriminar para diferenciar sin discriminar para marginar.

El problema de los sesgos, no obstante, no tiene tanto que ver con el hecho de que existan como con el de que se conozcan y se hayan colocado allí intencionalmente. No tenemos que renunciar a la discriminación positiva o a la selección en función de criterios concretos que no necesariamente generen la mayor eficiencia (especialmente, en labores militares en las que debe primar la eficacia sobre la eficiencia). Lo importante es que los sesgos que estén presentes en los procesos no sean el resultado de un error en la provisión de datos a los algoritmos o en la configuración de los mismos.

El último de los aspectos que aparecen recurrentemente en los códigos éticos académicos es el de la «explicabilidad» de los algoritmos, un término que podríamos traducir por inteligibilidad, aunque el anglicismo ha triunfado ya en el sector. Se trata de la capacidad por parte de los ingenieros de comprender la lógica que sigue la máquina y prever el proceso que va a seguir en la obtención de resultados. Al fin y al cabo, algo tan evidente como la capacidad de entender la forma de razonar de una máquina en la que estamos confiando la toma de decisiones o, al menos, el asesoramiento.

A pesar de que hay varios proyectos que están trabajando en este área –incluyendo uno de DARPA, la Agencia de Investigación Avanzada de Defensa de Estados Unidos-, no está claro que se pueda conseguir explicar el mecanismo de toma de decisiones y la ponderación que se asigna a cada dato en procesos tan complejos que supondrían años de trabajo para un ser humano.

Y, sin embargo, se antoja una temeridad arrojarse en manos de cajas negras en cuyo interior no se sabe qué ocurre. A pesar de todo, la transparencia asociada a la previsibilidad supone también una vulnerabilidad a la hora de contrarrestar las acciones de los algoritmos propios por parte del adversario. Si nuestro sistema es previsible, el del enemigo podrá actuar de modo que contrarreste su eficacia.

Este conjunto de características se resumen, finalmente, en la capacidad para comprender los procesos lógicos de las máquinas y mantener el control sobre los resultados en algún momento. La medida más deseable sería la inclusión de criterios éticos, de seguridad y jurídicos –cuando proceda– en todas las fases de diseño, desarrollo y uso de las aplicaciones.

Todo ello, llevado más allá del estrecho campo de los usos de la inteligencia artificial en los sistemas de armas autónomos capaces de ser letales, sino sobre cualquier técnica relacionada con este campo que pueda afectar de forma independiente a la vida o la libertad de las personas.

Igual que con el ciberespacio, todo ello se tiene que combinar con la potenciación de las aplicaciones civiles beneficiosas, evitando restringir la investigación o ralentizarla en función de sus posibles usos militares o ilícitos. Es una difícil tarea que excede el propósito de este trabajo.

Para los responsables de integrar las soluciones derivadas de la inteligencia artificial en la Defensa, la identificación de posibles aplicaciones expuestas en estas páginas, y las prevenciones éticas y jurídicas que se mencionan deberían constituir una buena base de partida. Todo ello, teniendo en cuenta que la naturaleza disruptiva de estas tecnologías obliga a llevar a cabo algo más que la integración de las mismas en el funcionamiento diario de las unidades y de las operaciones e incorporar también el desarrollo constante de soluciones propias que proporcionen un cierto margen de soberanía sobre este ámbito.

Cabe la tentación de limitarse a pensar en cómo se pueden seguir ejecutando las mismas tareas mejorando su eficiencia con el uso de la inteligencia artificial. El verdadero potencial, sin embargo, es identificar aquellas tareas que no podían llevarse a cabo –o resultaban insufriblemente costosas– antes y que ahora otorgan una ventaja en el campo de batalla de la información, del conocimiento y de los sentimientos en el que se ganan las guerras del siglo XXI. Hay que encontrar soluciones novedosas a los problemas y escenarios actuales, no solo optimizar los procesos que ya venían desarrollándose hasta ahora.

Composición del grupo de trabajo

Presidente

Ángel Gómez de Ágreda
Coronel de Aviación
Área de Análisis Geopolítico – DICOES/SEGENPOL

Coordinador

José Molino Martínez
Coronel de Aviación
Área de Análisis y Prospectiva – CCDC/CESEDEN

Vocales

Inmaculada Mohíno Herranz
Departamento de Teoría de la Señal y Comunicaciones
Escuela Politécnica Superior
Universidad de Alcalá – UAH

Rocío Barragán Montes
Departamento de Sistemas Aeroespaciales,
Transporte Aéreo y Aeropuertos.
Escuela Técnica Superior de Ingeniería Aeronáutica
y del Espacio – ETSIAE
Universidad Politécnica de Madrid – UPM

Francisco Antonio Marín Gutiérrez
Teniente Coronel del Ejército de Tierra
Strategic Employment Directorate.
NATO ACO (SHAPE)

Enrique Cubeiro Cabello
Capitán de Navío
Jefe del Estado Mayor
Mando Conjunto de Ciberdefensa

Jose Luis Aznar Lahoz
Teniente Coronel del Ejército de Tierra
Jefatura CIS de las FAS
EMAD


ieee.es
Instituto Español de Estudios Estratégicos