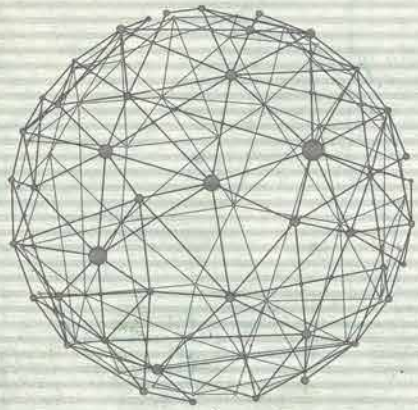
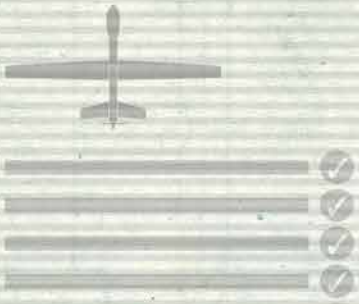
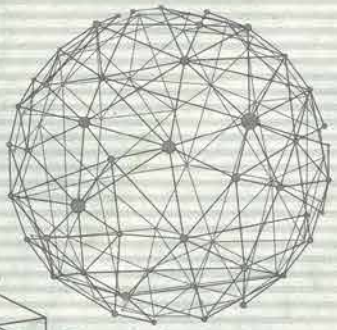
ROBOTS ASESINOS

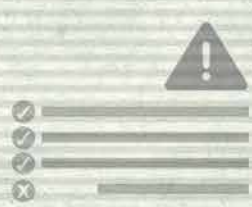
18 preguntas y respuestas

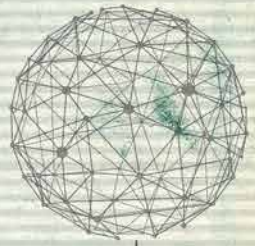




Autoría:
PERE BRUNET
TICA FONT
JOAQUÍN RODRÍGUEZ







CAMPAIGN TO STOP
KILLER ROBOTS

[01] POR QUÉ LA INTELIGENCIA ARTIFICIAL ES ALGO TAN FASCINANTE?

El buscador de Google encuentra 87 millones de páginas que hablan de inteligencia artificial[1] mientras que el número de páginas que tratan del cambio climático asciende a 45 millones. Prácticamente la mitad. ¿Por qué? Se han escrito miles de páginas sobre lo que pueden traernos estos nuevos sistemas de inteligencia artificial, muchas más de las que se han dedicado al cambio climático o, por ejemplo, a divulgar lo que permite que podamos hacer una foto y mandarla al instante a amigos que se encuentran en cualquier parte del planeta.

El tratamiento social que se está dando a estos tres temas, por citar sólo unos pocos, es cuanto menos sorprendente. Venimos hablando de inteligencia artificial desde hace 70 años, aunque sólo en las dos últimas décadas hemos empezado a ver algunos resultados concretos. De las nuevas posibilidades que ofrece internet para la comunicación de imágenes, vídeos y música se habla poco: es un regalo que el siglo XXI nos ha traído por sorpresa y sin aviso previo; algo que no entendemos, pero que rápidamente hemos integrado en nuestras vidas. Y el interés social por la crisis ambiental que causa el cambio climático, hasta hace poco era ínfimo. Este tema era objeto de debate únicamente en círculos científicos, a pesar de las catastróficas implicaciones que puede llegar a tener para todos.

Se considera que la inteligencia artificial (IA) empezó a definirse correctamente en 1950, año en el que Alan Turing propuso su conocido test para poder llegar a determinar si una máquina era inteligente o no.[2] El objetivo era claro, aunque tanto las técnicas informáticas como el incipiente desarrollo tecnológico eran demasiado rudimentarios. Las primeras soluciones no llegarían hasta entrado el siglo XXI. En todo caso, es importante subrayar que las definiciones de inteligencia artificial hablan de máquinas que actúan de manera inteligente pudiendo llegar a hacerlo de manera indistinguible en relación a los humanos. La inteligencia artificial es comportamiento, actuación. Las definiciones científicas hablan de actuar, no de pensar. Es importante distinguir entre estos dos aspectos.

Uno de los primeros logros de la inteligencia artificial nos llegó de la mano de los traductores automáticos. El investigador Franz Josef Och fue pionero en diseñar los primeros algoritmos en 2003,[3] y luego Google los incorporó entre los años 2005 y 2007. Estos nuevos traductores funcionaban tras aprender de ingentes cantidades de datos. Según Och, para poder traducir bien entre dos idiomas se necesita un corpus de texto bilingüe de más de 150 millones de palabras y dos corpus monolingües de más de mil millones de palabras.

En este contexto, ¿por qué este tema llenó miles de páginas entre 1950 y 2005, cuando era una quimera, y por qué continúa llenándolas? ¿por qué nos llegan mensajes apocalípticos que nos hablan de máquinas pensantes que podrán dominarnos? ¿por qué tendemos a considerarlos veraces? ¿por qué tenemos esta

actitud ambivalente que nos lleva a la vez a desearlos y temerlos? ¿por qué nos atrae, la inteligencia artificial?

Creo que la fascinación se fomenta, como mínimo, desde tres ámbitos. Desde los intereses ocultos de aquellos que, para cultivar sus negocios particulares, nos hablan de grandes virtudes mientras ocultan los defectos y problemas. Desde los tecnólogos que, con su euforia creativa, presentan muchos de sus nuevos sistemas con los sesgos propios de aquellos padres que no aceptan públicamente los defectos de sus hijos. Y desde aquellos que ven la oportunidad de utilizar estas nuevas técnicas para el control pre-totalitario (o directamente totalitario) de la población.

Y la fascinación se amplifica porque llueve sobre nuestra innata tendencia a generar mitos y a disfrutar de ellos. Creamos máquinas y soñamos pensando que nos dominarán. Pero nuestro deber es separar los mitos de la realidad, y en este punto, la ciencia nos puede ayudar. Podemos inventar grandes historias sobre lo que nos puede deparar la inteligencia artificial, pero debemos dejarlas en el rincón de los mitos y en cambio escuchar a las personas científicas para saber cuál va a ser la realidad. En este sentido, Michael Shermer habla de la imposibilidad de que lleguemos a ver máquinas que piensen, que sean auto-conscientes y que tengan emociones. Este apocalipsis, esta singularidad, dice irónicamente, lo más probable es que nos llegue en algún momento entre los años 2525 y 9595.[4] Luego, citando a Michael Chorost, dice que "en el momento en que un sistema de IA desee algo, pasará a vivir en un universo con recompensas y castigos, incluidos los castigos a los que le someteremos".[5] Como el de desenchufar la máquina, si ello llegase a ser necesario.

La fascinación no cesará. Y será perfectamente aceptable si sabemos mantenerla en el ámbito de los mitos mientras, al mismo tiempo, nos esforzamos por entender la realidad y los hechos objetivos. Porque quienes nos querrán controlar serán personas concretas, no máquinas. Hay quien dice que las máquinas pensantes acabarán dominándonos dentro de pocas décadas. Pero lo cierto es que, con muy alta probabilidad, dentro de varios siglos habrá quienes continúen diciendo exactamente lo mismo.

Notas:

1. Datos de diciembre de 2020
2. La prueba o test de Turing requiere dos personas y un ordenador (o dos). Una de las dos personas, la interrogadora, teclea preguntas en el ordenador. Tras cada pregunta y al cabo de poco, aparece una respuesta en pantalla. Algunas veces, aleatoriamente, esta respuesta ha sido tecleada por la segunda persona, que se encuentra en una habitación separada; otras, ha sido generada de manera automática por el ordenador, sin intervención humana. Tras ver la respuesta, la persona interrogadora debe responder si cree que la respuesta es humana o proviene de la máquina. Según Turing, puede considerarse que la máquina (ordenador) es inteligente si actúa de manera inteligente y logra engañar al interrogador, dando respuestas que éste considera que provienen de una persona humana. Ver por ejemplo: http://www.cse.chalmers.se/~aikmitr/papers/Turing.pdf#page=442
3. Ver: Och, Franz Josef (2003), "Statistical Machine Translation: From Single-Word Models to Alignment Templates", Technical Report, RWTH Aachen, Department of Computer Science: http://www-i6.informatik.rwth-aachen.de/publications/download/520/OchF.J.--StatisticalMachineTranslationFromSingle-WordModelstoAlignmentTemplates--2002.pdf - Ver también su presentación de 2005, ya como empleado de Google: "Machine Translation", Summit 2005, Phuket: http://www.mt-archive.info/MTS-2005-Och.pdf
4. Michael Shermer (2012), "In the year 9595". Ver: https://michaelshermer.com/sciam-columns/in-the-year-9595/
5. Michael Shermer (2017), "Apocalypse AI". Ver: https://michaelshermer.com/sciam-columns/apocalypse-ai/

[02] ¿EN QUÉ CONSISTE LA INTELIGENCIA ARTIFICIAL?

La inteligencia artificial es un concepto muy amplio que incluye una gran variedad de técnicas y algoritmos. Una definición bastante clarificadora es la que dice que es la inteligencia que pueden llegar a tener las máquinas, realizando tareas que típicamente requieren el uso de capacidades humanas inteligentes.[1]

La inteligencia artificial es por tanto "inteligencia" de máquinas, y se basa en la posibilidad de actuar, en el marco de determinadas tareas, de manera parecida los humanos. Se trata de una "habilidad" para realizar y resolver tareas, captando la realidad con sensores y luego actuando. En este sentido, no incluye la posibilidad de razonar ni de pensar.

La informática impregna nuestras vidas. En muchos casos, nos encontramos con utensilios que usan algoritmos fiables (los sistemas GPS, los sistemas de lectura de códigos de barras en comercios, los cajeros bancarios) que funcionan casi sin errores; en otros, usamos sistemas automáticos cuyo comportamiento es predecible (lavaplatos, lavadoras, etc.). Los nuevos sistemas de inteligencia artificial, en cambio, acaban "tomando decisiones" antes de actuar, y éstas pueden ser imprevisibles.

Tras el auge inicial de los algoritmos basados en el conocimiento,[2] durante las últimas décadas la inteligencia artificial (IA) se ha ido materializando básicamente en nuevos algoritmos denominados de aprendizaje automático profundo.[3] Estos sistemas de IA con aprendizaje profundo (*Deep Learning*, DL, en inglés) primero deben aprender de un número ingente de datos antes de empezar a actuar. Para aprender, necesitan grandes cantidades de información, el llamado *Big Data*. No puede haber sistemas de inteligencia artificial basados en aprendizaje profundo[4] sin *Big Data*.

Los sistemas con DL se basan en una red neuronal profunda, que no es más que una gigantesca red de conexiones entre una inmensidad de neuronas digitales,[5] organizadas en múltiples capas de la red.

En cada caso concreto, los expertos en datos deben primero analizar el problema para decidir la estructura de red más adecuada al problema. A continuación viene el primer paso, de entrenamiento o aprendizaje, que implica procesar grandes cantidades de datos para ajustar el modelo. Durante este proceso, los algoritmos estadísticos de aprendizaje van optimizando los parámetros asociados a todas las conexiones entre las neuronas de la red.[6] Al final de este proceso de aprendizaje o entrenamiento, que se hace en grandes ordenadores, la red neuronal acaba teniendo sus millones de parámetros ajustados en base a los datos de entrenamiento[7] y puede ser ya instalada en el sistema o artilugio que la utilizará en la práctica. Este paso podríamos decir que construye la red neuronal, porque "aprende" de los datos y "personaliza" los parámetros de todas sus conexiones.

En el segundo paso, el uso posterior de estas redes en aplicaciones reales es muy eficiente y requiere poca potencia de cálculo. Supongamos el caso de un sistema de reconocimiento de caras. Los datos de entrada de la red neuronal en este caso serían todos los píxeles de una imagen de la persona que se desea analizar (el sensor de entrada del sistema de IA en este caso es una cámara). Estos píxeles alimentan la capa de entrada de neuronas y generan señales que se van transmitiendo capa a capa hasta llegar a la capa final de salida, que genera el resultado del sistema de IA[8] y que en este caso podría ser una determinada clasificación de la persona según la foto de su cara.[9] El proceso de reconocimiento en cada caso concreto es muy rápido y puede integrarse en micro-ordenadores y en diversos dispositivos, dado que las operaciones matemáticas a nivel de neuronas son extremadamente sencillas.

Notas:

1. Stuart Russell y Peter Norvig (1004), "Artificial Intelligence: A Modern Approach", Prentice Hall: http://aima.cs.berkeley.edu
2. Steels & Lopez de Mántaras (2018), "The Barcelona declaration for the proper development and usage of AI": https://content.iospress.com/articles/ai-communications/aic180607
3. Los algoritmos de aprendizaje automático aprenden de datos reales. Pueden dividirse en cinco categorías principales: los algoritmos evolutivos genéticos, los algoritmos basados en analogía, los sistemas de aprendizaje simbólicos, las máquinas de aprendizaje bayesianas y los algoritmos de aprendizaje profundo. Estos últimos han experimentado una rápida evolución durante los últimos años. Ver: Domingos, Pedro (2018): "Artificial Intelligence Will Serve Humans, Not Enslave Them", Scientific American, September 2018: https://www.scientificamerican.com/article/artificial-intelligence-will-serve- humans-not-enslave-them/
4. En lo que sigue, a los sistemas de inteligencia artificial basados en aprendizaje profundo se les denominará, por simplicidad, sistemas de inteligencia artificial (IA).
5. Los nodos (neuronas digitales, inmateriales, consistentes en pequeños trozos inmateriales de software) de estas redes neuronales profundas tienen entradas y salidas y se organizan en capas, como las neuronas de nuestro cerebro. Se dice que son "profundas" cuando incluyen múltiples capas intermedias entre la entrada y la salida, con un gran número de neuronas y multitud de conexiones. Cada una de estas neuronas recibe información de otras muchas neuronas de la capa anterior, las pondera con parámetros que el sistema ajusta en base a los datos de aprendizaje, y genera una señal que se transmite a las neuronas de la capa siguiente. La estructura global es simple: neuronas conectadas a otras neuronas. Toda neurona incluye tantos parámetros como conexiones de entrada tiene con neuronas de la capa anterior (si en una conexión de entrada a una cierta neurona N el parámetro es del 35%, esto significa que toda señal que entre a esta neurona a través de esta conexión se amortiguará en un 35%). El problema es la inmensa cantidad de parámetros que hay que ajustar durante el aprendizaje: tantos como conexiones. En una red neuronal de tipo DL podemos estar hablando de cientos de millones de parámetros. La estructura de estas redes puede ser diversa: tenemos las redes neuronales recurrentes, las redes neuronales basadas en convolución (CNN) o las redes generativas adversarias (GAN) entre otras. Ver, por ejemplo: Samira Pouyanfar et al. (2018): "A Survey on Deep Learning: Algorithms, Techniques, and Applications", ACM Computing Surveys, Volume 51 Issue 5, November 2018: https://dl.acm.org/citation.cfm?id=3234150
6. Es una optimización que nunca puede llegar a ser perfecta, porque llegar al óptimo requeriría un tiempo de cálculo gigantesco.
7. Este entrenamiento inicial de las redes puede consumir mucho tiempo y también requiere una gran cantidad de datos (el volumen de datos de entrenamiento debe ser como mínimo del mismo orden de magnitud que el número de conexiones neuronales y parámetros).
8. Cada neurona de la DL calcula la media ponderada de las señales procedentes de las neuronas de la capa anterior en base al valor del parámetro asociado a cada conexión, aplica ciertas operaciones no lineales (funciones de umbral y de activación) y envía su salida a las neuronas de la siguiente capa. El uso de un conjunto de funciones no lineales de activación es fundamental para garantizar que cada neurona intervenga de manera diferenciada en el resultado final (de lo contrario, toda la red DL se convertiría en un enorme sistema lineal que se podría simplificar en una única multiplicación de matrices).
9. Esta salida algunas veces se utiliza para mejorar el ajuste de los parámetros de la red, en los esquemas que incluyen aprendizaje dinámico con retroalimentación.

[03] ¿LOS ALGORITMOS DE INTELIGENCIA ARTIFICIAL, SON SEGUROS Y FIABLES?

La respuesta, según Virginia Eubanks y otros muchos autores, es categórica: no lo son.[1]

Podríamos argumentar con razón que pocas cosas, en este mundo, son realmente seguras y fiables. Nosotros mismos somos vulnerables e imprevisibles, y todo lo que vemos, creamos y construimos está sujeto a multitud de imponderables que pueden acabar descontrolando su comportamiento.

Pero, en relación con los algoritmos de IA, constatamos dos hechos: estos sistemas son mucho menos fiables de lo que nos explican, y tanto sus desarrolladores como sus panegiristas y ensalzadores suelen tender a esconder fallos y errores, dando una visión sesgada de su grado de fiabilidad. Porque el nivel de fallos y errores de los algoritmos y sistemas de inteligencia artificial es muy superior al de los algoritmos clásicos,[2] deterministas y no basados en Big Data, aquellos que nos ayudan en nuestros desplazamientos (como el sistema GPS), los que calculan la cuenta de nuestra compra en el supermercado, o los que nos permiten interaccionar durante las video llamadas, por citar solo algunos ejemplos.

No existe ningún sistema de IA capaz de contextualizar y de hacer el tipo de inferencias básicas que incluso un niño realiza sin esfuerzo.[3] Y de hecho, los sistemas de reconocimiento de imágenes basados en IA se han demostrado altamente inestables. Ciertos cambios que son imperceptibles para los ojos humanos pueden hacer que un sistema de IA deje de reconocer una imagen como la de un león, pasándola a clasificar como la de una biblioteca, por ejemplo.[4] Y es por ello que un buen número de activistas se marca la cara con algunas líneas de pintura.[5] Con ello, consiguen que muchos sistemas de reconocimiento facial con IA se descontrolen y no puedan reconocerles.

La fiabilidad de los algoritmos de IA depende de la aplicación y de la complejidad de la salida deseada (no es lo mismo diseñar un sistema para clasificar imágenes de productos envasados en dos categorías correcto/defectuoso que construir un sistema que deba inferir caras de personas a partir de información sobre las mismas, por ejemplo), del diseño estructural de la red neuronal, de si el aprendizaje continúa o no durante el uso de la red, del tamaño de la propia red... y de los datos utilizados para el entrenamiento. En este caso, la fiabilidad de los resultados es función del número de datos y los posibles sesgos inherentes a los mismos, sesgos que terminan siempre perjudicando a las minorías raciales y de género y a las personas más vulnerables.[6] Los datos que se suministran a los sistemas de IA para su aprendizaje están habitualmente sesgados, heredando los prejuicios de aquellas personas que han intervenido en los procesos y en el negocio de los datos. En consecuencia, los sistemas de IA acaban reproduciendo estos sesgos y reduciendo su grado de fiabilidad. Pero además, debido a su estructura masivamente heurística y a un

proceso de aprendizaje que es necesariamente subóptimo, adolecen de una fiabilidad que es intrínsicamente limitada, algo que es inherente a su estructura y que implica una probabilidad de error no despreciable.

En ciertas aplicaciones críticas como las de diagnóstico en medicina, y ante errores del orden del 12%,[7] los expertos entienden que la intervención de los expertos en la toma de decisiones es imprescindible y que hay que incorporar evaluaciones clínicas con pacientes durante los necesarios procesos de validación previa.[8] Porque, gracias a ONGs como Big Brother,[9] sabemos de la baja fiabilidad de los sistemas de reconocimiento facial que se utilizaron durante los carnavales de Candem: sólo un 5% de las identificaciones de criminales hechas a través del sistema de IA fueron correctas, con un error promedio del 95%.[10]

Todos los nuevos sistemas que utilizamos deben someterse a pruebas y procesos de certificación para determinar su fiabilidad y seguridad. Deberíamos pedir lo mismo con los sistemas de IA y sobretodo con los basados en datos y deberíamos disponer de información cuantitativa sobre su grado de fiabilidad. Pero, además de disponer de metodologías de verificación y validación adecuadas, debemos exigir la creación de agencias independientes de certificación que validen todas las nuevas aplicaciones de IA antes de que se utilicen de forma generalizada.[11]

Notas:

1. Virginia Eubanks (2018), "Automating Inequality", St. Martin's Press: https://virginia-eubanks.com/
2. La definición de algoritmo que proporciona Richard Dawkins en su libro "Escalando el monte improbable" es especialmente interesante: es una manera muy adecuada de resumir el conocimiento que tenemos sobre cualquier conjunto de reglas". Los algoritmos de los sistemas GPS o de las cajas de cobro en los supermercados siguen paso a paso conjuntos de reglas precisas y no ambiguas que garantizan resultados correctos.
3. Ramón López de Mántaras (2020), "El traje nuevo de la inteligencia artificial", Investigación y ciencia, Julio de 2020: https://www.investigacionyciencia.es/revistas/investigacion-y-ciencia/una-nueva-era-para-el-alzhimer-803/el-traje-nuevo-de-la-inteligencia-artificial-18746 - Ramon López de Mántaras es fundador y exdirector del Instituto de Investigación en Inteligencia Artificial del CSIC, en Barcelona.
4. Ver: Nguyen, Anh (2015), "Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images", IEEE CVPR 2015: https://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/ - Ver también; Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, Pascal Frossard (2016), "DeepFool: a simple and accurate method to fool deep neural networks", Proceedings of the IEEE CVPR, https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Moosavi-Dezfooli_DeepFool_A_Simple_CVPR_2016_paper.pdf
5. Ver por ejemplo: https://i-d.vice.com/en_uk/article/jge5jg/dazzle-club-surveillance-activists-makeup-marches-london-interview
6. Virginia Eubanks: Una respuesta al DHS del condado de Allegheny, sobre la herramienta de evaluación familiar de Allegheny (AFST): "Creo que el sistema es injusto y discriminatorio. Además, la declaración del condado del 31 de enero sugiere que para el 55% de las familias, la mayoría, la recepción de servicios públicos de hecho aumenta su puntuación AFST, dejándolos desproporcionadamente vulnerables en las investigaciones sobre bienestar infantil": https://virginia-eubanks.com/2018/02/16/a-response-to-allegheny-county-dhs/
7. Xinyuan Zhang, Shiqi Wang, Jie Liu & Cui Tao (2018), "Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge", BMC Medical Informatics and Decision Making: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-018-0631-9?optIn=true
8. Emma Beede (2020), "Healthcare AI systems that put people at the center": https://www.blog.google/technology/health/healthcare-ai-systems-put-people-center/
9. Ver: https://bigbrotherwatch.org.uk/campaigns/stop-facial-recognition/#facial-recognition-uk
10. Roser Martínez Quirante y Joaquín Rodríguez (2020), "El costat fosc de la intel·ligència artificial - El cas dels sistemes d'armament letal autònom o els Killer Robots", Revista Idees: https://revistaidees.cat/el-costat-fosc-de-la-intelligencia-artificial/
11. Luca Steels y Ramón Lólez de Mántaras (2018), "The Barcelona declaration for the proper development and usage of artificial intelligence in Europe": https://content.iospress.com/articles/ai-communications/aic180607

[04] ¿EN QUÉ CASOS TIENE SENTIDO USAR TÉCNICAS BASADAS EN INTELIGENCIA ARTIFICIAL?

La Inteligencia Artificial puede ser una buena herramienta para el progreso social, pero existe un peligro real de uso inapropiado, prematuro o malintencionado. Dado que no siempre es recomendable, es importante analizar su idoneidad en distintos ámbitos.

En todo caso, habría que distinguir entre la IA basada en el conocimiento y la IA basada en datos.[1] La IA basada en el conocimiento, que se empezó a usar a finales de los setenta, intenta modelar el conocimiento humano en base al diseño de reglas y algoritmos. En ella, los expertos diseñan las soluciones a partir del análisis de los problemas. La IA basada en datos, en cambio, también conocida comúnmente como aprendizaje automático, trabaja de forma ascendente a partir de grandes cantidades de datos y ha aparecido como una gran novedad ya entrado el siglo XXI. En este caso, podríamos decir que los expertos van experimentando con distintos esquemas para ver si con alguno de ellos, una ingente cantidad de datos les resuelve el problema.

En lo que sigue, nos centraremos en esta última, la Inteligencia Artificial basada en datos y aprendizaje automático, por ser la que genera más expectativas y más interés social. En este caso, se sabe y se ha demostrado que las aplicaciones del aprendizaje profundo a dominios que involucran comportamientos regidos por reglas y cuestiones humanas, como la toma de decisiones financieras, la gestión de recursos humanos o la aplicación de la ley, son problemáticas desde un punto de vista humanista[1].

Y es que estos sistemas, basados habitualmente en redes neuronales profundas,[2] son masivamente heurísticos,[3] con un número ingente de parámetros que hay que ajustar durante su etapa de aprendizaje. El hecho de que este proceso de aprendizaje sea subóptimo y que utilice grandes conjuntos de datos que inevitablemente son sesgados, lleva a que estos sistemas sean menos fiables de lo que cabría esperar,[4] con errores en sus resultados y decisiones.

En este contexto, y una vez aceptado el inherente margen de error de estos sistemas, podemos distinguir distintos tipos de usos:

- Aplicaciones no críticas para casos concretos, con post-supervisión: se trata de sistemas para aplicaciones específicas, que requieren que la persona interesada revise los resultados y corrija posibles errores. Un ejemplo típico serían los sistemas de traducción automática, ciertamente útiles cuando conllevan una revisión sistemática del texto traducido.
- Aplicaciones críticas para casos concretos, con post-supervisión: Como ejemplo tenemos los sistemas de ayuda al diagnóstico médico y de interpretación

de imagen médica. Existe un muy elevado consenso entre los expertos en la necesidad de post-supervisión, de manera que el diagnóstico final sea responsabilidad del experto.[5]

- Aplicaciones no críticas que funcionan en promedio y sin post-supervisión: En este caso, podemos citar el ejemplo de los sistemas publicitarios que nos envían mensajes con recomendaciones personalizadas de todo tipo, basadas en nuestro comportamiento anterior. Son sistemas éticamente discutibles, invasivos para las personas, pero que pueden ser beneficiosos para las empresas anunciantes: aunque no incluyan post-supervisión, los errores individuales no impiden que, en promedio, estos sistemas acaben generando un aumento global de ventas.
- Aplicaciones críticas que funcionan en promedio y sin post-supervisión: Como ejemplo, podríamos mencionar los sistemas personalizados de propaganda electoral fragmentada que intentan enviar mensajes electorales a medida, en concordancia con los intereses concretos de cada persona y con lo que desearía que le prometiesen.[6] Como en el caso anterior, a pesar de sus errores pueden ser efectivos porque consiguen más votos de los que hacen perder. Pero son ética y democráticamente inaceptables. La investigación de la empresa Cambridge Analytica obligó a su cierre.

En resumen, los sistemas de Inteligencia Artificial con post-supervisión pueden ser útiles en muchos casos, como ayuda a la toma de decisiones de los expertos y en nuestra vida diaria (sistemas de reconocimiento de voz y música, sistemas aconsejan determinados productos, etc.). Pero su post-supervisión por parte de humanos es esencial. Por lo que respecta a las aplicaciones que funcionan en promedio y sin post-supervisión (críticas o no) deberían ser objeto de una regulación que garantice los derechos de todas las personas.

Notas:

1. Luca Steels y Ramón López de Mántaras (2018), "The Barcelona declaration for the proper development and usage of artificial intelligence in Europe", IOS Press: https://content.iospress.com/articles/ai-communications/aic180607
2. Véase la respuesta a la pregunta 2
3. Las técnicas heurísticas son métodos prácticos para la resolución de problemas que no garantizan un resultado óptimo ni perfecto, pero que alcanzan soluciones aproximadas aceptables y suficientes. Habitualmente, los sistemas heurísticos incluyen un conjunto de parámetros que hay que ajustar experimentalmente para lograr que su comportamiento sea adecuado.
A este proceso de ajuste se le denomina "afinado" ("tuning" en inglés) por su similitud con el método de puesta a punto de automóviles, durante el cual se ajustan manualmente multitud de variables o parámetros. Los sistemas masivamente heurísticos se caracterizan por depender del ajuste de millones o miles de millones de parámetros. Sobre el ajuste de parámetros en sistemas heurísticos, véase por ejemplo: Steven P. Coy et al. (2000), "Using Experimental Design to Find Effective Parameter Settings for Heuristics", Journal of Heuristics, Volume 7: http://yalma.fime.uanl.mx/~roger/work/teaching/mecbs5122/Papers/DOE/joh-2001-coy-heuristic%20doe.pdf
4. Véase la respuesta a la pregunta 3
5. Emma Beede (2020), "Healthcare AI systems that put people at the center": https://www.blog.google/technology/health/healthcare-ai-systems-put-people-center/ - Véase también: Will Douglas (2020), "Google's medical AI was super accurate in a lab. Real life was a different story": https://www.technologyreview.com/2020/04/27/1000658/google-medical-ai-accurate-lab-real-life-clinic-covid-diabetes-retina-disease/
6. Véase, por ejemplo, el caso de la empresa Cambridge Analytica: Pere Brunet (2018): https://virvigblogs.cs.upc.edu/2018/05/11/segmentar-el-missatge/

[05] ¿EL COMPORTAMIENTO Y LA MANERA DE ACTUAR DE LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL, ES FÁCILMENTE EXPLICABLE?

Cuando los sistemas de Inteligencia Artificial basados en aprendizaje profundo (IA) aciertan y dan el resultado esperado, lamentablemente no podemos saber por qué han funcionado bien. Pero tampoco sabemos por qué fallan cuando se equivocan. De hecho, es algo que no saben los usuarios pero que tampoco pueden saber los diseñadores de estos sistemas de IA.[1] Es el denominado "problema de la caja negra", que hace que sea prácticamente imposible explicar las decisiones que toman estos sistemas. Ésta es una de las diferencias esenciales con otros artilugios que usamos a diario. Si una lámpara no se enciende, sabemos que puede ser un problema de la bombilla, del cable o del interruptor. Y si un coche deja de funcionar, el mecánico sabrá encontrar fácilmente la causa, sabiendo explicar el motivo y pudiendo proceder a su reparación. Poder explicar el por qué de un problema es el primer paso hacia su resolución.

Pero cuando los sistemas de IA tienen un comportamiento erróneo e inesperado, no hay forma de entender qué ha pasado. Por ello, decimos que los sistemas actuales de Inteligencia Artificial no son explicables. En los casos en que no funcionan, nadie puede explicar por qué han fallado. Es uno de sus graves inconvenientes. Son cajas negras.

Este comportamiento no explicable es consecuencia de la extraordinaria complejidad de las redes neuronales que conforman estos sistemas, de su carácter heurístico,[2] y de la ingente cantidad de parámetros que los gobiernan. Pero también es consecuencia de la multiplicidad de sesgos que con toda probabilidad contenían los datos de aprendizaje, y de su carácter inestable. Como resultado de estos sesgos y de la enormidad de implicaciones que acaban teniendo en el proceso de ajuste de los parámetros de la red, los sistemas de IA es probable que tomen "decisiones" incorrectas[3] a la vez que extrañas y por ello inexplicables. Por otra parte, su bien demostrada inestabilidad acaba agravando este carácter no explicable, porque cambios muy sutiles en la información que el sistema debe clasificar pueden llevar, inexplicablemente, a decisiones extraordinariamente distantes. Hay muchos ejemplos que han mostrado esta inestabilidad, por ejemplo, en los sistemas de clasificación y detección a partir de imágenes. Y de hecho, la modificación de unos pocos píxeles en la imagen de entrada, un cambio que es totalmente imperceptible para el ojo humano, es suficiente para inestabilizar el sistema, haciendo que salte de una clasificación a otra.[4] Un típico ejemplo muestra una fotografía de ImageNet con un autobús escolar; cuando se distorsiona de manera inapreciable, el sistema de IA sorprendentemente la clasificó como un avestruz. Son comportamientos inesperados y por ello inexplicables. En este contexto, parte de los actuales investigadores en

el campo de la IA están estudiando maneras que permitan que el propio sistema, además de dar su respuesta al problema planteado, nos dé información que explique el por qué de dicha respuesta. Todo ello tendrá, no obstante, nuevas limitaciones.[5]

Esta falta de explicabilidad está relacionada com la opacidad de las "cajas negras" y con no poder detectar cuales han sido los fallos internos que han llevado a determinados resultados erróneos. Todo ello impide repararlos e imposibilita que en el futuro podamos evitar errores similares.

Notas:

1. Ramón López de Mántaras (2020), "El traje nuevo de la inteligencia artificial", Investigación y ciencia: https://www.investigacionyciencia.es/revistas/investigacion-y-ciencia/una-nueva-era-para-el-alzhimer-803/el-traje-nuevo-de-la-inteligencia-artificial-18746 - Ramon López de Mántaras es fundador y exdirector del Instituto de Investigación en Inteligencia Artificial del CSIC, en Barcelona. Como explica López de Mántaras, "las personas tampoco podemos explicar siempre nuestras decisiones. Sin embargo, hay una diferencia fundamental: los humanos tendemos a confiar unos en otros porque creemos que los mecanismos de pensamiento de los demás son similares a los nuestros. Es lo que los psicólogos llaman tener una «teoría de la mente» sobre los demás. No obstante, ninguno de nosotros tiene una teoría de la mente sobre ninguna máquina, ni desde luego ninguna máquina la tiene sobre nosotros. Por ello, resulta perfectamente razonable exigir más explicaciones a una máquina que a una persona".
2. Véanse las respuestas a las preguntas 2 y 4.
3. Pere Brunet (2020), "El negocio de las armas que van contra la ética y las personas", El Salto Diario, enero de 2020: https://www.elsaltodiario.com/industria-armamentistica/negocio-armas-contra-etica-personas
4. Jiawei Su, Danilo Vasconcellos Vargas, Kouichi Sakurai (2019), "One Pixel Attack for Fooling Deep Neural Networks", IEEE Transactions on Evolutionary Computation, Vol. 23: https://arxiv.org/pdf/1710.08864.pdf
5. Dado que la fiabilidad de los sistemas de IA es siempre limitada, sus "explicaciones" tendrán una cierta probabilidad de ser erróneas. Véase la respuesta a la pregunta 3.

[06] ¿CUÁLES SON LOS PRINCIPALES PROBLEMAS ÉTICOS RELACIONADOS CON LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL?

A lo largo de la ultima década se ha ido construyendo una falsa narrativa sobre las bondades de la inteligencia artificial que tiene tendencia a ignorar todos aquellos aspectos que los expertos y académicos están expresando. De hecho, los sistemas de IA presentan comportamientos no explicables, con una probabilidad garantizada de error que es significativa y no pequeña. Esto los hace esencialmente discutibles en determinadas situaciones críticas en las que los errores podrán poner en riesgo vidas humanas (diagnóstico médico, vehículos autónomos o el caso de las armas autónomas, en las que los errores se convertirán directamente en vidas humanas) y en las que puede ser difícil la rendición de cuentas.[1] Los principales problemas éticos relacionados con estos sistemas derivan de su **carácter oscuro y no explicable,**[2] de su **fiabilidad limitada**[3] **y probabilidad de error,** del hecho de **no poder resolver las ambigüedades** que aparecen en las situaciones reales, de sus **inevitables sesgos,** y de la **necesidad de post-supervisión** en el caso de aplicaciones críticas. Los sistemas de IA aprenden, captan y actúan. Pero **pueden ser imprevisibles** y no siempre lo hacen de la manera esperada.

Uno de los campos de aplicación que se nos presentan como más prometedores en el campo de los sistemas de IA es el de los sistemas autónomos. La **diferencia constructiva** entre un sistema automático y uno autónomo es que en el primer caso (puertas automáticas, lavadoras y lavaplatos), su comportamiento es previsible, mientras que los sistemas autónomos, que han sido construidos para poder actuar en base a “sus” decisiones (basadas en su aprendizaje y en la información que captan), pueden ser imprevisibles, además de no explicables y en casos, erróneos.

Pero los problemas éticos relacionados con los sistemas de Inteligencia Artificial provienen no tanto de sus potencialidades constructivas sino de **su uso, que es el que puede ser ético o no.** En este sentido podemos distinguir tres casos:

- Los sistemas basados en IA y autónomos desde un punto de vista constructivo que se usan bajo control humano. Sería el caso de los vehículos autónomos que deberían ser constantemente controlados por la persona que los conduce. En este caso, los problemas éticos asociados a los sesgos, baja fiabilidad y no explicabilidad, quedan resueltos por la existencia de una persona responsable que puede y debe asumir una posible rendición de cuentas.
- Los sistemas basados en IA que se usan con post-supervisión. Sería el caso de los sistemas de ayuda al diagnóstico médico en los que los expertos acaban tomando la decisión final. Como el caso anterior, su uso suscita pocos problemas éticos, aunque es importante recordar el sesgo de automatización,[4] que habría que tener siempre en cuenta.

- Los sistemas basados en IA que funcionan de manera autónoma, sin control ni intervención humana. En este caso, los problemas éticos asociados a los sesgos así como a la baja fiabilidad, comportamiento imprevisible y no explicabilidad, son muy relevantes.

En este contexto, algunos autores como Alan Winfield y Marina Jirotka proponen que los robots y los sistemas autónomos deban ir equipados con una "caja negra ética" que seria el equivalente de los grabadores de datos de vuelo de los aviones, y que registraría continuamente los datos y el estado interno de los sistemas de inteligencia artificial en aplicaciones críticas.[5] Esta caja negra ética seria esencial para poder entender lo que ha ocurrido en caso de víctimas y facilitaría el establecimiento de responsabilidades.

En todo caso, el uso de sistemas de IA en equipos de armamento autónomos (las denominadas "LAWS" en inglés) nos lleva al máximo grado de problemas éticos. La escalada hacia los sistemas armados autónomos es ética y jurídicamente inaceptable, porque delegar en una máquina las decisiones de matar va en contra de la dignidad humana y de los derechos de las personas. Hay que situar el concepto de dignidad humana como límite insalvable,[6] marcando una "línea roja" más allá de la cual la autonomía en los sistemas de armas ya no puede ser aceptable.[7] Los drones letales autónomos no podrán tomar decisiones éticas complejas en campos de batalla dinámicos, ni podrán distinguir adecuadamente entre soldados y civiles, y tampoco podrán evaluar el grado de proporcionalidad de un ataque. Sin hablar de su comportamiento imprevisible, la posible pérdida de control, los accidentes "habituales" y los potenciales malos usos.

Notas:

1. Joaquín Rodríguez, Xavi Mojal, Tica Font y Pere Brunet (2019), "Nuevas armas contra la ética y las personas. Drones armados y drones autónomos", Informe 39, Centro Delàs de Estudios para la Paz: http://centredelas.org/wp-content/uploads/2019/11/informe39_DronesArmados_RE_CAST_web_DEF-1.pdf
2. Véase la respuesta a la pregunta 5.
3. Véase la respuesta a la pregunta 3.
4. El sesgo de automatización es la tendencia humana a dar por bueno aquello que nos proponen las máquinas. Noel Sharkey explica que hay que tener en cuenta este sesgo, que rebaja fuertemente el posible apoyo ético a estos sistemas. Este sesgo hace, según Sharkey, que "los operadores estén predispuestos a aceptar las recomendaciones informáticas sin buscar otras informaciones que permitan su confirmación. La presión temporal añadida hace que los operadores caigan en todas las trampas del razonamiento automático: en lugar de pensar, pasan a creer y aceptar aquello que la máquina los propone; ignoran la ambigüedad, suprimen la duda, inventan causas e intenciones, se centran en las pruebas existentes e ignoran las pruebas ausentes que ellos tendrían que buscar". Véase: Sharkey, Noel (2014): "Towards a principle for the human supervisory control of robot weapons". UNOG: https://www.unog.ch/80256EDD006B8954/(httpAssets)/2002471923EBF52AC1257CCC0047C791/$file/Article_Sharkey_PrincipleforHumanSupervisory.pdf
5. Alan Winfield and Marina Jirotka (2017): "The Case for an Ethical Black Box", Proc. of the Annual Conference Towards Autonomous Robotic Systems TAROS 2017, pp. 262-273: https://link.springer.com/chapter/10.1007/978-3-319-64107-2_21
6. Palmerini E., Azzarri F., et al.. (2016) "Robolaw: Guidelines on Regulating Robotics": https://www.researchgate.net/publication/322041670_Guidelines_on_Regulating_Robotics
7. Daan Kayser and Alice Beck (2018): "Crunch Time". PAX Report, Noviembre 2018: https://www.paxvoorvrede.nl/media/files/pax-rapport-crunch-time.pdf

[07] ¿QUÉ RELACIÓN HAY ENTRE LOS DRONES, LOS ROBOTS DE COMBATE, LAS ARMAS DIRIGIDAS A DISTANCIA Y LAS ARMAS AUTÓNOMAS?

Los Drones, los robots de combate, las armas dirigidas a distancia, así como los sistemas de armas autónomos, forman parte de una nueva generación de armamento que aprovecha los avances tecnológicos, producidos mayoritariamente en ámbitos civiles a fin de substituir el despliegue de tropas (humanas) por sistemas robóticos (Rodríguez et al., 2011). Todos estas tipologías de armamento, comparten características y propiedades comunes:

En primer lugar tienen la capacidad de reducir el número de tropas combatientes desplegadas en el terreno por parte de los ejércitos poseedores de las mismas, reduciendo así los costes de intervención (fundamentalmente los logísticos asociados al despliegue) así como las bajas propias. (Scharre & Norton, 2018)

En segundo lugar, tienen una capacidad intrínseca de invisibilización del conflicto a ojos de la opinión pública (Rodríguez-Álvarez & Martinez-Quirante, 2019), ya que para los periodistas y expertos en la materia es prácticamente imposible tener constancia de las intervenciones realizadas con este tipo de armamento. Ya que, al no estar precedidos por despliegues de tropas, o incluso declaraciones formales de guerra pueden pasar desapercibidos dando lugar a nuevas dinámicas de conflicto de baja intensidad, dotando, así de una mayor opacidad a los operativos militares.

En tercer lugar, todos estos tipos de armamento, profundizan, aún más, las dinámicas de la Guerra Asimétrica que se dan entre ejércitos o fuerzas combatientes que tienen un profundo “gap” tecnológico entre ellas (Thornton & Miron, 2020). Como por ejemplo en el caso de Israel y Palestina, donde la superioridad tecnológica de los primeros, determina un tipo de conflicto, no solo completamente desigual, sino además carente de incentivos reales para su resolución. Profundizando de esta forma la dinámicas de opresión preexistentes en el sistema.

En cuarto lugar y tal como se ha venido denunciando a lo largo de la última década, todas estas tipologías de armamento tienen el potencial de erosionar el Derecho Humanitario Internacional al aumentar el número de ejecuciones extrajudiciales, así como otro tipo de operaciones de tipo “quirúrgico”. Como ejemplo de este tipo de intervenciones podemos destacar la ejecución extrajudicial del General Iraní Qasem Soleimani, abatido el 3 de Enero de 2020 por un dron Estadounidense en el aeropuerto Internacional de Bagdad (BBC, 2020).

En quinto lugar, todos estos tipos de armamento, por motivos diferentes, representan problemas de tipo ético, relativos a la responsabilidad de los actos (tanto en

situaciones que involucran operadores humanos, como en las que no) además de añadir factores de incerteza e imprevisibilidad asociados a cualquier sistema tecnológico complejo, pudiendo generar errores que desemboquen en víctimas civiles. A modo de ejemplo se podría recordar la bomba MK 82 guiada por láser de 500 libras (227 kilos) fabricada por Lockheed Martin que impactó contra un autobús de Naciones Unidas en Yemen niños el 9 de agosto de 2018, matando a 40 niños o el ataque de marzo del mismo año, también en Yemen, esta vez contra un mercado, –en esa ocasión con una bomba MK 84 guiada a precisión, que según se reportó entonces– dejó 97 muertos civiles. (Elgabir et al., 2018)

Si bien, además de estas características y propiedades comunes, existen diferencias importantes entre las mismas a tenor de la participación o control humano sobre las mismas. (Musco, 2020)

En este sentido debemos tener en cuenta que lo que distingue a un arma autónoma del resto, es que pese a que el armamento autónomo puede tomar forma de drone, robots de combate, misil... etc. La clave reside en que este puede operar sin controladores humanos involucrados en el ciclo de acción del arma, a través de una Inteligencia Artificial. Es decir, lo que diferencia a un drone "normal" de un drone "autónomo" es que el primero necesita que un piloto "humano" lo maneje de forma remota, mientras que el drone autónomo puede llegar a tener capacidad operativa completa sin intervención humana.

Bibliografía:

Rodríguez, J., Mojal, X., Font, T., & Brunet, P. (2011). Nuevas Armas contra la ética de las personas: Drones armados y drones autónomos. www.defenceimagery.mod.uk/OGL;

Rodríguez-Álvarez, J., & Martinez-Quirante, R. (2019). Towards a new AI race. The challenge of lethal autonomous weapons systems (Laws) for the United Nations. Aranzadi - Tomson Reuters.

Scharre, P., & Norton, W. W. (2018). Army of None: Autonomous Weapons and the Future of War Army of None: Autonomous Weapons and the Future of War Arms Control Today. https://www.armscontrol.org

Thornton, R., & Miron, M. (2020). Towards the 'Third Revolution in Military Affairs': The Russian Military's Use of AI-Enabled Cyber Warfare. RUSI Journal, 165③, 12-21. https://doi.org/10.1080/03071847.2020.1765514

BBC. (2020, January 3). Qasem Soleimani: EE.UU. mata en Irak al poderoso general, líder de la fuerza élite Quds de Irán - BBC News Mundo. Bbc.Com. https://www.bbc.com/mundo/noticias-internacional-50979843

Elgabir, N., Abdelaziz, S., Brone, R., Arvanitidis, B., & Smith-Spark, L. (2018, September 18). La bomba que mató a 40 niños en Yemen fue proporcionada por Estados Unidos. CNN. https://cnnespanol.cnn.com/2018/08/18/la-bomba-que-mato-a-40-ninos-en-yemen-fue-proporcionada-por-estados-unidos/

Musco, A. (2020). Meaningful Human Control of Autonomous Weapon Systems Definitions and Key Elements in the Light of International Humanitarian Law and International Human Rights Law.

[08] ¿QUÉ SON LOS SISTEMAS DE ARMAMENTO AUTÓNOMOS Y LETALES? (CONOCIDOS POR LAS SIGLAS “LAWS” EN INGLÉS)

Los sistemas de armamento autónomo, son sistemas de armamento que debido al uso extensivo de algoritmos de Inteligencia Artificial pueden operar sin la necesidad de involucrar decisores humanos.(Rodríguez-Álvarez & Martinez-Quirante, 2019)

Este tipo de armamento, que han sido calificados por algunos autores, como la tercera revolución de la guerra (Thornton & Miron, 2020), tras la pólvora y el armamento nuclear, englobaría, a Grosso modo, cualquier sistema de armamento que posea tres características básicas:

- Pueden moverse independientemente a través de su entorno a lugares que ellos escogen de manera arbitraria. Sus capacidades son: movilidad, persistencia y orientación y navegación.
- Pueden seleccionar y disparar contra objetivos en su entorno. Sus capacidades son: identificación propia de objetivos, discriminación para categorizar objetivos, priorización de objetivos y selección del tipo de arma apropiada al objetivo.
- Pueden crear y/o modificar sus objetivos incorporando la observación de su entorno y la comunicación con otros agentes. Sus capacidades son: autodeterminación, autocompromiso, comunicación autónoma con otros sistemas, automodificación de objetivos basada en información adquirida de fuentes autónomas, planificación de objetivos y aprendizaje y adaptación constantes.

(Martinez-Quirante & Rodríguez-Alvarez, 2018 p.36)
Es decir, estos sistemas de armamento, suponen una transformación sustancial en la propia comprensión e historia del conflicto ya que implican la delegación de capacidades letales a entes no humanos (Martinez- Quirante & Rodríguez-Álvarez, 2020), con la problemática que ello supone, no solo a nivel legal en lo relativo a la asunción de responsabilidades, en casos por ejemplo donde se produzcan víctimas civiles(Rodríguez et al., 2011). Sino a nivel ético, ya que implica delegar decisiones de vida o muerte sobre entes incapaces de reconocer el propio valor de una vida humana que queda reducida a simples bites de información. Hecho que implicaría no solo una violación a la noción de dignidad humana, sino una amenaza sin precedentes para la especie. Que por primera vez en su historia coexistiría con entes no humanos armados.

Los riesgos asociados a este armamento, además de legales y éticos, también implican los operativos y tácticos, fundamentalmente en situaciones en que ambos combatientes dispongan de este tipo de sistemas (Martinez-Quirante & Rodríguez-Alvarez, 2018). Unos sistemas que tienen capacidades operativas mucho más veloces

que las capacidades humanas, lo que supone un riesgo de escalabilidad del conflicto sin precedentes (Sharkey, 2011), ofreciendo además serias dudas sobre la capacidad de desactivación de dichos sistemas una vez estén situados en un contexto de escalada bélica.

A esto debemos añadir, los riesgos puramente técnicos, ya que en tanto que creación humana, la Inteligencia Artificial es falible, y sus rangos de error garantizado hace imposible el cumplimiento de nociones básicas del Derecho Humanitario Internacional (Asaro, 2012), tal y como podría ser la distinción entre combatientes y no combatientes. Además, en tanto que proceso heurístico cuyo proceso se asemeja a una caja negra donde entran millones de inputs de información, a partir de los cuales se obtiene un output en forma de decisión/acción, es practicamente imposible determinar el porqué de la misma, hecho que de facto descartaría sus posibilidades de uso legal, al ser imposible determinar los porqués relativos a una determinada acción (Sharkey, 2010).

En última instancia, cabe además añadir que la proliferación de este tipo de armamento, puede producir transformaciones radicales no solo en la forma y estructura de los ejércitos, sino también de las fuerzas y cuerpos de seguridad del estado, que comenzarán a experimentar de forma paulatina una penetración exponencial de sets tecnológicos (Sharkey, 2018), que en muchos casos irán orientados a la propia sustitución de humanos en ciertos procesos de toma de decisión. Este hecho puede significar a largo plazo una reconcentración del poder debido a la superación de la necesidad de generar consentimiento entre la sociedad, así como los grupos afines. Erosionando de esta forma el “contrato social”.

Bibliografía:

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

Rodríguez, J., Mojal, X., Font, T., & Brunet, P. (2011). Nuevas Armas contra la ética de las personas: Drones armados y drones autónomos. www.defenceimagery.mod.uk/OGL;

Martinez-Quirante, R., & Rodríguez-Alvarez, J. (2018). Inteligencia artificial y armas letales autónomas: un nuevo reto para Naciones Unidas. Trea.

Rodríguez-Álvarez, J., & Martinez-Quirante, R. (2019). Towards a new AI race. The challenge of lethal autonomous weapons systems (Laws) for the United Nations. Aranzadi - Tomson Reuters.

Martinez- Quirante, R., & Rodríguez-Álvarez, J. (2020). El lado oscuro de la Inteligencia artificial . IDEES, 48. https://revistaidees.cat/es/el-lado-oscuro-de-la-inteligencia-artificial/

Sharkey, N. (2010). Saying ‘No!’ to Lethal Autonomous Targeting. Journal of Military Ethics, 9④, 369-383.

Sharkey, N. (2011). The Automation and Proliferation of Military Drones and the Protection of Civilians. Journal of Law Innovation and Technology, 2③, 229-240.

Sharkey, N. (2018). Mama Mia It’s Sophia: A Show Robot Or Dangerous Platform To Mislead? Forbes.

Thornton, R., & Miron, M. (2020). Towards the ‘Third Revolution in Military Affairs’: The Russian Military’s Use of AI-Enabled Cyber Warfare. RUSI Journal, 165③, 12-21. https://doi.org/10.1080/03071847.2020.1765514

[09] ¿CUÁLES SON LAS RAZONES QUE SE APORTAN PARA JUSTIFICAR SU FABRICACIÓN Y USO?

En la actualidad, la mayor parte de grandes potencias, así como la industria militar privada se han lanzado a una carrera en pos de profundizar en los usos militares de la Inteligencia Artificial. Países como Estados Unidos, Rusia, China, Israel, Corea de Sur o el Reino Unido (Burri, 2016; SKR, 2018) están en la actualidad liderando la investigación y desarrollo de este tipo de armamento. Esta reorientación del complejo industrial militar se debe por una parte a razones de tipo objetivo:

- La primera razón sería que una sustitución de personal militar por máquinas conllevaría una reducción de las bajas propias en determinadas situaciones operativas. (Scharre & Norton, 2018)
- La segunda razón es la reducción de costes logísticos asociados al despliegue de tropas. El armamento autónomo permite intervenciones a un menos coste económico. (*ibídem*)

Si bien, a parte de estas razones, existen otra variedad de argumentos que pueden ser calificados como subjetivos/ especulativos, en tanto y cuanto se sustentan sobre relatos que tienen más de mitológicos que de realidad.

Por una parte, se argumenta que el armamento autónomo reducirá los márgenes de errores, así como la victimología civil debido a que la inteligencia artificial es más eficaz y eficiente que un humano. Hecho que, si bien puede aceptarse en determinadas circunstancias, nunca puede ser aceptado como un postulado general, ya que existen numerosas evidencias de la capacidad de la IA, de reproducir sesgos humanos (O'Neil, 2017) (Statt, 2020). Además de serias dudas sobre la capacidad de la misma de distinguir entre conceptos tan básicos como combatiente y no combatiente. Creer en la infalibilidad de los algoritmos puede llevar a resultados dramáticos como nos mostró el *American Civil Liberties Union* con respecto a los sistemas de reconocimiento facial que tienen una alta tendencia a identificar sujetos no caucásicos como criminales (Snow, 2018).

Por otra parte se suele argumentar, que el uso de este tipo de armamentos evitará la comisión de crímenes de guerra, como las violaciones, pasando por alto que la mayor parte de la comisión de este tipo de crímenes forma parte de la propia estrategia de los combatientes para la desmoralización del adversario (Spade & Willse, 2014). Además, pese al hecho de que una máquina no pueda cometer ciertos tipos de crímenes de guerra, está claro que este tipo de armamento puede propiciar otras violaciones del derecho humanitario internacional tal y como podrías ser la potenciación de las ejecuciones extrajudiciales.

En tercer lugar existe un argumento/mito que insiste en construir una pseudo-humanización de la tecnología a partir de una teórica (y no probada) codificación

ética de los algoritmos. De esta forma se argumentaría que pueden adoptar comportamientos éticos-morales si estos son correctamente codificados. Pero es evidente que una máquina no puede tener ni ética ni moral ni intuición propia. En todo caso podrá tener la ética de quien lo ha codificado. Será una simulación de la ética del programador, una réplica del ingeniero o una combinación de los datos que encuentre en la nube. ¿Sin embargo, podemos preguntarnos si una vez codificada la IA, el sistema evolucionará por sí solo? ¿o si nos condenará a una sociedad de tipo inmovilista donde el bien y el mal queden cristalizados en la base de una construcción subjetivizada en los algoritmos? ¿Y si evoluciona...cuál será su hito? (Martinez- Quirante & Rodríguez-Álvarez, 2020).

Y finalmente, el tercer argumento/mito empleado por los defensores de la militarización de la IA, sería el relativo a la fiabilidad, que afirma que la inteligencia artificial es más fiable que la inteligencia humana, cosa que en análisis muy específicos podría ser aceptado, pero nunca en términos generales. Hay que destacar aquí el trabajo hecho por la ONG británica Big Brother is Watching us que apelando al acto de libertad de información, consiguieron que el gobierno revelara la fiabilidad de los sistemas de reconocimiento facial que se utilizaron durante el Carnaval de Candem. El resultado fue que sólo un 5% de las identificaciones de criminales hechas a través del sistema d'IA eran correctas, dando un error medio del 95%. (BBW, 2017).

Bibliografía:

BBW. (2017). Big Brother Watch: Defending Civil Liberties, Protecting Privacy. https://bigbrotherwatch.org.uk/

Burri, T. (2016). The Politics of Robot Autonomy. European Journal of Risk Regulation, 7(02), 341-360. https://doi.org/10.1017/S1867299X00005766

Martinez- Quirante, R., & Rodríguez-Álvarez, J. (2020). El lado oscuro de la Inteligencia artificial . IDEES, 48. https://revistaidees.cat/es/el-lado-oscuro-de-la-inteligencia-artificial/

O'Neil, C. (2017). Weapons of math destruction : how big data increases inequality and threatens democracy. Broadway Books.

Scharre, P., & Norton, W. W. (2018). Army of None: Autonomous Weapons and the Future of War Army of None: Autonomous Weapons and the Future of War Arms Control Today. https://www.armscontrol.org

SKR. (2018). Campaign to Stop Killer Robots. https://www.stopkillerrobots.org/

Snow, J. (2018, July 26). Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots | American Civil Liberties Union. ACLU. https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28

Spade, D., & Willse, C. (2014). Sex, Gender, and War in an Age of Multicultural Imperialism. http://againstequality.org/files/Spade_Wilse_Manning_2014.pdf

Statt, N. (2020, June 10). Amazon bans police from using its facial recognition technology for the next year - The Verge. The Verge. https://www.theverge.com/2020/6/10/21287101/amazon-rekognition-facial-recognition-police-ban-one-year-ai-racial-bias

[10] ¿QUÉ NIVELES DE CONTROL HUMANO PODEMOS ENCONTRAR EN LAS NUEVAS ARMAS?

Noel Sharkey clasifica el nivel de control humano de las nuevas armas en cinco niveles:[1]

1. Existe una deliberación en relación a los objetivos, por parte de personas responsables, antes de iniciar cualquier ataque.
2. Se dispone de sistemas automáticos que proporcionan una lista de posibles objetivos, que son analizados por personas que acaban escogiendo y decidiendo qué objetivo será atacado.
3. Los sistemas automáticos escogen un objetivo a atacar, objetivo que determinadas personas deberán aprobar antes de que se proceda al ataque.
4. Los sistemas automáticos escogen un objetivo a atacar y proceden a atacarlo si las personas a cargo de estos sistemas no detienen la acción durante un periodo corto y limitado de tiempo.
5. Los sistemas automáticos escogen un objetivo a atacar y proceden a atacarlo sin intervención humana alguna.

Sharkey continúa explicando que, ya en el nivel 1, cualquier deliberación debe cumplir unos mínimos requisitos: "el comandante / operador humano debe tener una conciencia contextual y situacional completa del área objetivo en el momento del ataque específico, siendo capaz de percibir y reaccionar ante cualquier cambio o situación imprevista que pueda haber surgido desde la planificación del ataque. Debe haber una participación cognitiva activa en el ataque, con tiempo suficiente para deliberar sobre la naturaleza del objetivo, su importancia en términos de la necesidad y conveniencia del ataque y los posibles efectos incidentales y accidentales [daños colaterales] del ataque. También deben existir medios para la suspensión rápida o aborto del ataque". Por lo que respecta al nivel 2, y continuando con el razonamiento anterior, Sharkey considera que el proceso puede ser aceptable si se demuestra que cumple con el requisito de control humano deliberativo indicado en el nivel 1. Sin embargo, el nivel 3 es ya inaceptable según Sharkey, porque hay que tomar en consideración el llamado **sesgo de automatización**, que rebaja fuertemente sus fundamentos éticos. El sesgo de automatización es la tendencia humana a dar por bueno aquello que nos proponen las máquinas.[2] Incluso si el sistema dispone de un botón rojo para parar el arma robótica, el sesgo de la automatización puede acabar influyendo algunas veces a los operadores, induciéndoles a seguir las acciones propuestas por estos sistemas, sin reflexionar y sin pulsar el botón. Finalmente, los niveles 4 y 5 son inaceptables.[3]

Por otra parte, podríamos hablar del modelo de circuito ("Loop Model") propuesto por Boulanin y otros.[4] Según este modelo, los sistemas militares armados se pueden agrupar en alguna de las tres categorías siguientes:

1. Sistemas “dentro del circuito” (“In-the-Loop”): Estos sistemas de armamento requieren intervención humana en las tareas relacionadas con la selección de objetivos y las decisiones de ataque. Se corresponden con los niveles 1 y 2 de Noel Sharkey; en este caso, las personas responsables son parte del circuito.
2. Sistemas “sobre el circuito” (“On-the-Loop”): Son sistemas de armamento con capacidad para seleccionar objetivos y decidir ataques pos sí solos, pero que esperan y solo atacan siguiendo órdenes explícitas efectuadas por las personas que operan el sistema. Se corresponden con los niveles 3 y 4 de Noel Sharkey.
3. Sistemas “fuera del circuito” (“Out-of-the-Loop”): Son los sistemas de armamento que pueden seleccionar objetivos, decidir ataques y atacar los objetivos de forma autónoma y sin ninguna intervención ni interacción humana. Es el caso de los sistemas clasificados en nivel 5 por Noel Sharkey, en los que las personas quedan fuera de las decisiones letales.

Sabemos que los ordenadores y las máquinas son superiores a los humanos en la realización de cálculos, en la ordenación de grandes cantidades de información y en la búsqueda dentro de ellas, en la realización simultánea y/o repetitiva de multitud de tareas, en la respuesta rápida en situaciones que hay que controlar, y en la resolución de muchos problemas complejos. Pero las personas superamos a las máquinas en razonamiento cognitivo y en la posibilidad de deliberar, razonar y decidir aplicando nuestro conocimiento, basado en experiencias pasadas totalmente diversas, a nuevas situaciones imprevistas. Y además, las personas somos responsables, de manera que la sociedad puede exigirnos rendición de cuentas. Por todo ello, los caminos que van del nivel 1 al 5 de Sharkey o del nivel 1 al 3 de Boulanin son de una legitimidad ética cada vez más dudosa e inquietante.

Notas:

1. Sharkey, Noel (2014): “Towards a principle for the human supervisory control of robot weapons”: https://www.onug.ch/80256EDD006B8954/(httpAssets)/2002471923EBF52AC1257CCC0047C791/$file/Article_Sharkey_PrincipleforHumanSupervisory.pdf
2. Este sesgo hace, según Sharkey, que “los operadores estén predispuestos a aceptar las recomendaciones informáticas sin buscar otras informaciones que permitan su confirmación. La presión temporal añadida hace que los operadores caigan en todas las trampas del razonamiento automático: en lugar de pensar, pasan a creer y aceptar aquello que la máquina les propone; ignoran la ambigüedad, suprimen la duda, inventan causas e intenciones, se centran en las pruebas existentes e ignoran las pruebas ausentes que tendrían que buscar”. Véase: Rodríguez, J., Mojal, X., Font,. T., Brunet, P., (2019), “Nuevas armas contra la ética y las personas”, Informe del Centro Delàs, págs.. 22-23-24: http://arxiu.centredelas.org/images/INFORMES_i_altres_PDF/informe39_DronesArmados_CAST_web_DEF.pdf
3. Véase la respuesta a la pregunta 18: ¿Las armas autónomas, son legales?
4. Boulanin, Vincent & Verbruggen, Maaike (2017), Mapping the Development of Autonomy in Weapon Systems, Estocolmo, SIPRI: https://www.sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_systems_1117_1.pdf

[11] ¿QUÉ SON LOS DRONES MILITARES QUE RONDAN?

Los drones militares que merodean, conocidos como "*Loitering drones*" (o "*Loitering munition*" en inglés), son vehículos aéreos no tripulados diseñados para atacar objetivos terrestres con una carga explosiva. Están equipados con cámaras ópticas e infrarrojas de alta resolución que permiten localizar, vigilar y guiar el vehículo hacia el objetivo a destruir.[1] Una característica definitoria de estos drones merodeadores es su capacidad de "vagar" por el aire y sobre una determinada zona objetivo durante un período prolongado de tiempo antes de atacar, lo que permite decidir cuándo y qué atacar.

Estos drones pueden ser de reconocimiento o de reconocimiento y ataque. En los dos casos, su interés está en el hecho de que no están dirigidos a un objetivo prefijado, sino a una zona o región objetivo. Van volando por la región asignada, sobrevolando y captando información de todo lo que encuentran. En el caso de los de reconocimiento, simplemente la comunican (en tiempo real o no) a la base. En cambio, cuando son de ataque, su sistema de decisión puede activar (siguiendo o no las órdenes del operador, según su grado de autonomía) sus sistemas de armamento.[2] Los drones que merodean pueden realizar misiones ofensivas y defensivas que serían consideradas peligrosas o arriesgadas para otros tipos de sistemas tripulados. Cuando son de ataque, son de un solo uso porque se autodestruyen en el ataque. Su utilidad operativa radica en el hecho de que no están dirigidas a un objetivo predefinido, sino que (a diferencia de las municiones guiadas) pueden moverse libremente dentro del área objetivo.[3]

Se fabrican en una amplia gama de modelos desde hace décadas y su uso está cada vez más extendido[1]. En los últimos años, por ejemplo, las fuerzas armadas de EEUU, Israel, Turquía e Irán han diseñado y fabricado diversos drones de ataque de este tipo, también denominados drones "suicidas" o "kamikaze", que muy pronto podrían revolucionar la forma en que se libran las guerras. Como ejemplo podemos hablar de los drones SwitchBlade,[4] de los Harop y FireFly, del Kargu-2 o del Qasef-1.[5]

El problema, desde una perspectiva ética, es que los modelos más modernos de estos drones armados que merodean ya tienen la capacidad de actuar de manera totalmente autónoma. Esto significa que pueden buscar y destruir objetivos utilizando algoritmos informáticos en vez de ser guiados por un operador humano. Una vez definida una cierta región objetivo, van volando y vagando por ella, vigilando, captando información, detectando posibles objetivos, decidiendo cuál atacar y finalmente destruyéndolo sin intervención humana alguna. Los drones armados que merodean y que incluyen funcionalidades autónomas se basan únicamente en las indicaciones de sus sistemas de inteligencia artificial.

Como ejemplo, el dron Harop es uno de estos drones "*loitering*" que puede actuar, según el tipo de software que se le active, en modo controlado o en modo autónomo.[6]

Según Boulanin[3], ya en 2017 el Harop era uno de los 49 sistemas desplegados que podían detectar posibles objetivos y atacarlos sin intervención humana.

Los drones armados que merodean son unos de los principales candidatos a incorporar funcionalidades autónomas y a convertirse, por tanto, en robots asesinos[7] voladores, término que "combina escalofriantemente dos de los grandes y temibles conceptos de la ciencia ficción: las armas peculiarmente poderosas y la inteligencia no humana", en palabras de Paul Iddon.

Notas:

1. Informe sobre "Loitering Munitions" del Centro para el Estudio de los Drones de la Universidad de Bard (2017): https://dronecenter.bard.edu/files/2017/02/CSD-Loitering-Munitions.pdf
2. Joaquín Rodríguez, Xavi Mojal, Tica Font, Pere Brunet (2019), "Nuevas armas contra la ética y las personas. Drones armados y drones autónomos", Informe 39, Centro Delàs de Estudios para la Paz, p. 18: http://centredelas.org/wp-content/uploads/2019/11/informe39_DronesArmados_RE_CAST_web_DEF-1.pdf
3. Boulanin, Vincent & Verbruggen, Maaike (2017), "Mapping the Development of Autonomy in Weapon Systems", Informe del Instituto SIPRI, p. 50-53: https://www.sipri.org/sites/default/files/2017- 11/siprireport_mapping_ the_development_of_autonomy_in_weapon_systems_1117_1.pdf
4. Dron SwitchBlade de AeroVironment: Véase: https://www.avinc.com/tms/switchblade
5. Drones Harop y FireFly, Kargu-2 y Qasef-1: Paul Iddon (2020), "Turkey, Israel And Iran Have Built Some Very Lethal Loitering Munitions", Forbes: https://www.forbes.com/sites/pauliddon/2020/07/19/turkey-israel-and-iran-have-built-some-very-lethal-loitering-munitions/amp/?streamIndex=1
6. The Economist Briefing (2019), "Autonomous weapons and the new laws of war": https://amp.economist.com/briefing/2019/01/19/autonomous-weapons-and-the-new-laws-of-war
7. Campaña Stop Killer Robots: https://www.stopkillerrobots.org/

[12] ¿QUÉ SON LOS ENJAMBRES DE DRONES MILITARES?

Los enjambres de drones militares , conocidos por las siglas “Drone swarms” en inglés, son conjuntos de decenas, centenares o miles de mini-drones que actúan coordinadamente gracias a un sistema específico de comunicación que posibilita la interacción entre ellos (Informe,[1] página 19). Pueden ser armados o de reconocimiento. Se inspiran en el comportamiento de los enjambres de pájaros y son extraordinariamente resistentes a los accidentes y adversidades, porque en el caso de problemas, cualquier subconjunto de drones del enjambre puede continuar desarrollando las tareas asignadas. La base rusa de Khmeimim en el oeste de Siria ha recibido diversos ataques con enjambres de drones, alguno con hasta 60 drones.[2]

Los enjambres de drones, junto con los drones que merodean, son uno de los sistemas de ataque que más probablemente incorporarán elementos autónomos de decisión basados en inteligencia artificial.

El proyecto OFFSET de los EEUU quiere poder gestionar enjambres de 250 drones armados,[3] mientras que el proyecto Gremlins se basa en disponer de grandes enjambres de drones recuperables.[4] Por otra parte, el Reino Unido está trabajando intensamente para progresar en el desarrollo de drones autónomos y en particular, de enjambres de drones con autonomía: El proyecto Many Drones Make Light Work, liderado por Blue Bear Systems (con un consorcio que incluye Airbus y otros) se plantea desarrollar enjambres de drones autónomos de bajo coste, que tienen que “llevar a un nuevo paradigma a las operaciones en campos de batalla... esto permitirá que el enjambre pueda realizar simultáneamente misiones complejas contra objetivos simples o múltiples de manera altamente eficaz” (Informe, página 31). Por otra parte, en Europa, el proyecto Roborder investiga el uso de técnicas de inteligencia artificial para el desarrollo de enjambres de drones inteligentes para vigilar las fronteras europeas por tierra, mar y aire. Y uno de los proyectos más significativos de la empresa China Ziyan se basa en el desarrollo de enjambres de drones de ataque inteligente, con la peculiaridad que no tienen que ser todos iguales. Unos de ellos pueden servir para tirar cargas explosivas, mientras que otros pueden incorporar sistemas lanzadores de granadas, o ser directamente drones kamikazes. El enjambre encontrará su ruta hasta los objetivos que debe destruir (Informe, página 30).

Notas:

1. Joaquín Rodríguez, Xavi Mojal, Tica Font, Pere Brunet (2019), “Nuevas armas contra la ética y las personas. Drones armados y drones autónomos”, Informe 39, Centro Delàs de Estudios para la Paz, páginas 19, 30, 31 y 39: http://centredelas.org/wp-content/uploads/2019/11/informe39_DronesArmados_RE_CAST_web_DEF-1.pdf
2. Michael Safi (2019), “Are drone swarms the future of aerial warfare?”, The Guardian: https://www.theguardian.com/news/2019/dec/04/are-drone-swarms-the-future-of-aerial-warfare
3. Proyecto OFFSET de DARPA (OFFensive Swarm-Enabled Tactics): https://www.darpa.mil/work-with-us/offensive-swarm-enabled-tactics
4. Theresa Hitchens (2020), “Gremlins swarming drones”, Breaking Defense, July 23, 2020: https://breakingdefense.com/2020/07/darpa-gremlins-drone-test-back-on-after-covid-delay/

[13] ¿QUÉ SON LOS SISTEMAS ANTI-DRONES?

Los sistemas anti-drones tienen como misión la detección, interceptación o destrucción de drones. En paralelo con el auge que están teniendo estos últimos y con la creciente preocupación en torno a las posibles amenazas a la seguridad que pueden representar (tanto a nivel civil como militar), puede constatarse que el nuevo mercado de tecnología contra drones está emergiendo rápidamente.

Es importante distinguir entre sistemas para la vigilancia y seguimiento de drones y sistemas para destruirlos o inutilizarlos. En el primer caso, el objetivo del sistema es únicamente la detección, clasificación/identificación, seguimiento del dron y alerta, de manera que las actuaciones posteriores son responsabilidad de los operadores y equipos humanos de decisión. Sus sistemas pueden basarse en sensores ópticos, acústicos, de radio o radar. En el segundo caso, en cambio, el objetivo es destruir el dron, neutralizarlo o tomar el control del mismo. Para ello se utilizan, entre otros:[1]

- Bloqueadores de radiofrecuencia que impiden las comunicaciones. Podemos ver, por ejemplo, sistemas electrónicos de este tipo en la Marina de los EUA:[2]
- Redes lanzadas desde cañones específicos desde tierra o desde otros drones, como las del sistema Karnivora (Rusia).[3]
- Sistemas láser de alta energía para la destrucción de drones en vuelo, como el que está desarrollando la empresa Raytheon en los EUA.[4]
- Sistemas de concentración de micro-ondas, que las dirigen con precisión al dron objetivo para desrtruirlo, como el sistema THOR[5] o el nuevo sistema proyectado por el Departamento de Defensa de los EUA contra enjambres de drones.[6]
- Sistemas que actúan generado interferencias en el GPS del dron.[7]
- Sistemas híbridos (que incluyen varios de los sistemas anteriores).
- En una categoría distinta a las anteriores entrarían proyectos como el de la universidad Gen Gurion de Israel, para identificar, localizar y matar directamente a los operadores que dirigen los drones militares a distancia.[8]

Para más información, el informe 2018 sobre sistemas anti-drones del Centro de Estudios sobre los Drones de la Universidad de Bard incluye una tabla con los 235 principales productos comerciales anti-dron existentes en aquél momento.[9]

Pero, mientras los drones militares están evolucionando rápidamente y van adquiriendo capacidades autónomas, la tecnología de los sistemas contra-drones también se está volviendo más inteligente. Los nuevos sistemas se están volviendo más sofisticados, integrando y fusionando diferentes tecnologías y usando enfoques innovadores como el aprendizaje automático, la fusión de sensores y los radares cognitivos y holográficos[10] para conseguir un funcionamiento cada vez más autónomo.

En todo caso, no existen estándares internacionales sobre el diseño y uso adecuado de estos sistemas. Esta ausencia de estándares plantea interrogantes sobre su seguridad, además de abrir la puerta a variaciones significativas en su rendimiento y fiabilidad.

Y en el ámbito militar, los sistemas anti-drones y sobretodo los futuros sistemas autónomos para la destrucción de drones pueden conducir a una escalada incontrolada de los conflictos bélicos. Ya están apareciendo sistemas para el ataque a los sistemas anti-drones (los llamados "Counter Counter-Drone Systems"[11]).
El despliege y uso incontrolado de estos sistemas podría generar altísimos niveles de destrucción (principalmente de equipos y sistemas) con costes económicos indescriptibles, todo ello en intervalos de tiempo muy cortos. El peligro, como dice Frank Sauer, es que la interacción entre sistemas regidos por algoritmos autónomos sin control humano es demasiado rápida y por ello incontrolable.[12]

Notas:

1. Robinradar (2020), "9 Counter-Drone Technologies to Detect and Stop Drones Today", 22-3-2020: https://www.robinradar.com/press/blog/9-counter-drone-technologies-to-detect-and-stop-drones-today
2. Ver por ejemplo: https://www.marinecorpstimes.com/news/2018/09/19/the-corps-just-slapped-a-counter-drone-system-on-an-mrzr-all-terrain-vehicle
3. El dron Karnivora ha sido desarrollado por la empresa rusa *Micran Research and Production*. El sistema está diseñado para interceptar otros drones lanzando una red que captura el dron hostil y lo hace aterrizar con paracaídas: https://tass.com/defense/1042083
4. Ver: https://aviationweek.com/awindefense/raytheon-deploy-counter-uas-laser-overseas-bases
5. El sistema Tactical High Power Microwave Operational Responder (THOR) de los EUA: https://taskandpurpose.com/air-force-thor-microwave-weapon
6. Sistema contra enjambre de drones basado en micro-ondas de alta energia, EUA: https://www.defensenews.com/digital-show-dailies/smd/2019/08/07/the-armys-indirect-fires-protection-system-is-getting-a-high-power-microwave/
7. Por ejemplo, el sistema de CENTUM: https://centum.com/es/centum-presenta-la-primera-solucion-integral-anti-drones/
8. Proyecto para atacar directamente a los operadores de los drones: https://breakingdefense.com/2020/07/israelis-crafting-counter-drone-system-to-track-kill-operators/
9. Arthur Holland Michel (2018), "Counter-Drone Systems", Centro de Estudios sobre los Drones de la Universidad de Bard, Tabla incluída en las págs.. 13-20: https://dronecenter.bard.edu/files/2018/02/CSD-Counter-Drone-Systems-Report.pdf
10. Claudio Palestini (2020), "Countering drones: looking for the silver bullet", NATO, 16-12-2020: https://www.nato.int/docu/review/articles/2020/12/16/countering-drones-looking-for-the-silver-bullet/index.html
11. Apartado 3.6 del "Counter-Drone Market Report 2020": https://droneii.com/product/counter-drone-market-report-2020
12. Frank Sauer, ICRAC (2019), "Why we are working to prohibit killer robots: Proliferation & Definitions", Campaign to Stop Killer Robots, Global Campaign Meeting, Berlin 22-23 March 2019.

[14] ¿QUÉ PROBLEMAS ÉTICOS PLANTEAN LAS NUEVAS TECNOLOGÍAS MILITARES EMERGENTES?

El 27 de septiembre de 2020 y a lo largo de seis semanas las republicas de Armenia y Azerbaiyán se enfrentaron en una guerra por el control de Nagorno Karabaj. No es la primera vez que lo hacen, lo singular en esta ocasión ha sido la utilización de los drones (aviones o sistemas aéreos no tripulados) en este conflicto por parte de Azerbaiyán y la incapacidad de las fuerzas armadas armenias, en concreto de sus sistemas antiaéreos convencionales, de contrarrestar estos ataques.[1]

Se han utilizado drones de fabricación turca e israelí tanto para misiones de reconocimiento como en misiones de ataque y han utilizado drones de tipo munición merodeadora[2] (loitering munition). Hasta ahora los drones, teledirigidos desde tierra y a varios kilómetros de distancia del aparato, han sido utilizados para llevar a cabo asesinatos selectivos o asesinatos extrajudiciales, por parte de Estados Unidos, de presuntos terroristas en Afganistán, Pakistán, Yemen o Somalia.

La existencia de drones armados representa un gran escollo para impedir el uso de la fuerza entre Estados y avanzar en el respeto de los derechos humanos. En el campo de batalla, durante una guerra (aunque ya ningún país declara la guerra, simplemente la hace), es posible que el uso de drones armados cumpla las normas fundamentales del DIH de distinción y proporcionalidad (aunque atribuir responsabilidad penal internacional por el uso ilícito de estas armas puede plantear grandes dificultades). Fuera del campo de batalla, el uso de ataques efectuados con drones representa, en general, una violación de los derechos humanos fundamentales. Es urgente esclarecer el sistema jurídico aplicable y formular las limitaciones necesarias para impedir que siga proliferando la tecnología relativa a los drones.[3]

En términos militares los drones y los drones munición presentan muchas ventajas, en tanto que son vehículos no tripulados, los militares no pueden ser abatidos; por otra parte, representan un menor coste económico, un misil antiaéreo es más caro que un dron, un avión de combate es mucho más caro que un dron y la formación de un piloto de avión de combate es costosa y requiere muchas horas de entrenamiento, para operar un dron se requieren habilidades de “playstation”.

En definitiva, estos nuevos aparatos nos llevan a las guerras “low cost”, no solamente en términos económicos sino también en términos éticos y políticos.

Estos aparatos ofrecen mayor proyección de fuerza con menor riesgo para la vida de los soldados (no para las victimas civiles) y permite llevar a cabo más actuaciones militares con un menos número de soldados. Lo que nos plantea una cuestión ética o nos lleva a cuestionar si evitar el riesgo para la vida de los combatientes esta por

encima de la vida de las victimas civiles o no combatientes, en definitiva, hay vidas que tienen más valor que otras.

Por otra parte, hay que considerar que el distanciamiento físico de los combatientes del mismo escenario de combate puede llevar a un distanciamiento psicológico y moral o a una disminución del sentido y conciencia moral de la responsabilidad de la acción llevada a cabo. Al distanciamiento físico hay que añadir que un militar se aproxima a lo que podemos llamar el campo de batalla a través de una cámara, puede que en infrarrojo y que la visión sea en blanco y negro, en la que se observa en la lejanía el blanco a abatir, en la que puede que los humanos sean una mera sombra en la que no distinguimos ni la ropa que llevan puesta; como en una pantalla de videojuego. Para la sociedad civil ver lo que sucede en la guerra real a través de estas imágenes no es muy diferente de la violencia que observamos en un videojuego, la sociedad nos estamos acostumbrando tanto a esta clase de imágenes que ya no generan en nosotros sensibilidad, empatía con las victimas o indignación con las consecuencias de la guerra, en definitiva, estamos tan sobreexpuestos a la violencia que su efecto es estéril psicológica y políticamente.[4]

Por otra parte, los drones son dirigidos desde una oficina a cientos o miles de kilómetros de distancia, el militar que hace la guerra la hace a turnos laborales, acabado su turno se va a casa a cenar y revisar los deberes escolares con sus hijos, produciendo un distanciamiento psicológico y moral sobre sus actos.

En este contexto los políticos pueden poner menor resistencia a involucrarse en una guerra ya que no estarían tan presionados por el rechazo de los ciudadanos a las perdidas de las vidas humanas, al mismo tiempo que los políticos pueden notar que disminuye la presión hacia ellos para la búsqueda de soluciones diplomáticas a un conflicto. Por ello los políticos cada vez más piden a los militares que diseñen intervenciones militares bajo la doctrina de “cero muertos”, es más fácil comandar a distancia a un dron que no enviar a nuestros hijos y tener que mandar cartas de pésame a las madres de soldados muertos.

En definitiva, utilizar estos drones o drones munición en combate nos abre un debate sobre la banalización del hecho bélico a escala política y social, y plantea una doble moral en relación con los sacrificios que puede aceptar una sociedad que demanda intervenciones militares sin sacrificios humanos propios y sin riesgos para los políticos. La utilización de estas armas en la guerra se nos presenta como una manera de hacer la “guerra limpia” o la “guerra inteligente” y como una manera más aceptable para la sociedad. Sin victimas militares propias, las encuestas de opinión son más favorables a los políticos.

Notas:

1. José Alberto Marín Delgado profundiza en toda la nueva tecnología utilizada en este conflicto del Cáucaso Sur http://www.ieee.es/Galerias/fichero/docs_opinion/2021/DIEEEO21_2021_JOSMAR_DronesCaucaso.pdf
2. Véase la respuesta a la pregunta 12: ¿Qué son los drones militares que merodean?
3. Stuart Casey-Maslen, (2012), ¿La caja de pandora? Ataques con drones: perspectiva desde jus ad belum, el just in bello y el derecho internacional de los derechos humanos; International Revieu of the Red Cross, nº 886
4. Puede ser de interés para el lector visionar el video “Il ny aura plus de nuit”, pueden ver una critica al mismo en: http://criticum.net/il-ny-aura-plus-de-nuit-el-ojo-que-todo-lo-ve o un tráiler en: https://www.youtube.com/watch?v=K7yJGGNqqNg

[15] ¿QUÉ PROBLEMAS ÉTICOS SE PLANTEAN EN EL CASO DE LOS SISTEMAS DE ARMAMENTO AUTÓNOMOS Y LETALES?

En la actualidad se desarrollan proyectos piloto de maquinas y tecnología capaz de tomar decisiones y llevar a cabo operaciones militares sin la intervención de los humanos. Ello abre todo un abanico de cuestiones éticas y morales.

Muchos ámbitos de nuestra vida cotidiana, en el trabajo, el ocio, las compras cada vez tenemos mayor participación de las aplicaciones y sus algoritmos matemáticos que nos ayudan en la toma de decisiones, escoger las vacaciones de verano, la mejor ruta aérea o el mejor precio de un producto. Las personas podemos confiar en ciertas aplicaciones y en los algoritmos de búsqueda, pero algunos se formulan preguntas como ¿hasta que punto queremos vivir en una sociedad en donde las maquinas tomen decisiones por nosotros? Pero lo que aquí nos preguntamos es va más allá de la cotidianidad de nuestras vidas y lo que nos preguntamos es ¿podemos delegar la decisión de quien vive y quien muere en una maquina y en unos algoritmos?

Para evitar la "deshumanización" tenemos abordar la ética de estos sistemas y de los algoritmos de Inteligencia Artificial en que se sustentan las decisiones bajo las cuales se decidirá la vida o la muerte de una persona. Para autores como Wallach y Allen[1] se trata de crear moralidad en las propias maquinas para que tomen decisiones morales por ellas mismas, es decir, diseñar algoritmos morales para construir ética y que las armas autónomas creadas tomen decisiones de acuerdo con los valores morales humanos.

Los tres componentes de la condición humana que estructuran el ser humano y que nos han hecho singulares y únicos con respecto del resto de los animales son: la inteligencia, la voluntad y las emociones. En este corto espacio de tiempo abordaremos este tercer aspecto, las emociones.

Es cierto que los humanos bajo ciertas condiciones como calor, rabia, miedo, ira rencor, odio, deseo de venganza, etc., actuamos de la peor manera posible, es cierto que en contextos de conflicto armado los humanos cometen vilezas como violar a mujeres y niñas, torturar a otros seres humanos, mutilarlos, etc., también es cierto que los robots no pueden actuar bajo estos estados de ánimo, es cierto que no pueden tener estos sentimientos y que por tanto pueden evitar muertes innecesarias.

También es verdad que los robots no tienen sentido del riesgo, no tienen miedo, no toman decisiones influidos por las emociones y que un robot no tendría, a diferencia de un humano, el instinto de supervivencia. Pero hay que reconocer que las emociones pueden ser la mejor salvaguarda, sin emociones se puede matar más fácilmente. Los robots no pueden tener el sentido común de los humanos, no pueden

sentir compasión, lástima o no pueden tener intuición. Es cierto que los humanos somos falibles y muchas veces los militares consideran este "factor humano" como negativo o como el eslabón más débil en contexto de conflicto armado, pero esa es una de las características más relevantes de la identidad del ser humano.[2]

Las decisiones sobre la vida y la muerte de los humanos en un conflicto armado pueden requerir de visión de conjunto, de comprensión de intenciones, de compasión de intuición o de sentido común. En definitiva, estamos cuestionando la capacidad de generar algoritmos que puedan reproducir emociones tan singulares como la empatía el amor, la piedad o la culpa. Lo mismo podemos decir sobre la capacidad para distinguir y evaluar entre ordenes lícitas o ilícitas o su capacidad para interpretar un contexto y evaluarlo en cálculos basados en valores.

Con especial sensibilidad hay que abordar los tres pilares del derecho internacional humanitario. El principio de responsabilidad, en caso de un error o un crimen de guerra ¿quién será el responsable? Una maquina no puede ser responsable, la responsabilidad solo es posible cuando es apropiado atribuir a alguien la culpa, el castigo o la recompensa por lo que ha hecho. Una maquina no puede ser castigada por que ello implicaría que sus algoritmos permiten el sentimiento y por tanto aprender del sufrimiento de castigo para no volver a hacer un acto.[3] En este aspecto cabe tener presentes a las victimas humanas y su necesidad de mirar cara a cara al perpetrador.

El principio de distinción, el robot no solamente tiene que ser capaz de distinguir si el objetivo es un combatiente o no, si es un insurgente armado o un civil inocente, sino que también tiene que hacer un balance de intenciones y cabe la posibilidad de que los insurgentes engañen al robot ocultando sus armas o avanzándose a sus limitaciones sensoriales.

El principio de proporcionalidad, que exige que antes de atacar debe de evaluarse el daño que pueda causarse a la población civil y los beneficios obtenidos con el mismo. La proporcionalidad es muy propia del discernimiento humano, es un concepto complejo y es un ejercicio puramente cualitativo y difícil de cuantificar; en todo caso se requeriría cuantificar cuantos civiles es "proporcional" matar en un ataque (0, 1, 2, 100) o cuantificar cuanto daño colateral es "excesivo" según Wagner el equilibrio, la proporcionalidad o el exceso dependen de los valores del individuo que haga los cálculos, por lo tanto el principio de proporcionalidad es, por naturaleza subjetivo.[4]

Por todo ello es necesario ejercer el control humano total sobre todo el proceso y sus partes desde el principio, no podemos dejar que la Inteligencia Artificial considere a los humanos irrelevantes o prescindibles.

Notas:

1. Wallach W. y Allen C. (2010), *Moral Machines: Teaching Robots Right From Wrong.* New York. Oxford University Press.
2. Rodríguez, J., Mojal, X., Font, T., & Brunet, P. (2011). *Nuevas Armas contra la ética de las personas: Drones armados y drones autónomos* http://centredelas.org/wp-content/uploads/2019/11/informe39_DronesArmados_RE_CAST_web_DEF-1.pdf
3. Sparrow, R. (2007) Killer Robots, *Journal of Applied Philosophy*, 24/1, 62-77 y Sparrow, R. (2016), Robots and respect: Assessing the case against Autonomous Weapon Systems. *Ethics and International Affairs*30①: 93-116
4. Wagner, Markus (2014): *The Dehumanization of International Humanitarian Law: Legal, Ethical and Political Implications of Autonomous Weapon Systems*" Vanderbilt Journal of Transnational Law, Vol. 47, pá g. 1380. [Acceso 20 de agosto de 2019].

[16] ¿LAS ARMAS AUTÓNOMAS, SON LEGALES?

A día de hoy, ningún país ha regulado sobre el diseño, la investigación, la producción o la tenencia de estas armas autónomas. Las organizaciones civiles agrupadas en la campaña *Stop Killer Robots*[1] abogan por la aprobación de un tratado internacional que ilegalice estas armas.

Las regulaciones sobre armas pueden ser de carácter internacional, en este caso un tratado. Los tratados internacionales son normas jurídicas de naturaleza internacional, vinculantes y de obligado cumplimiento para los Estados de lo ratifican, estos tratados son aprobados en una Asamblea General de Naciones Unidas, los Estados en primer lugar lo votan en dicha asamblea, posteriormente los firman y en tercer lugar lo ratifican en sus respectivos parlamentos; aquellos Estados que ratifican un tratado incorporan a su legislación nacional las obligaciones a que se comprometen con la ratificación. Si un Estado no ratifica un tratado no tiene la obligación de respetarlo, pero a menudo la carga moral imperante suele o puede provocar que el tratado sea respetado, aunque no se haya ratificado.

En la actualidad existen diversos tratados de prohibición sobre armamentos, por ejemplo, la Convención sobre Armas Químicas[2] (entró en vigor en 1997), la Convención sobre Armas Biológicas[3] (entró en vigor en 1975), la Convención sobre la Prohibición de las Minas Antipersonal, también llamado Trato de Ottawa[4] (entro en vigor en 1999), Convención sobre Armas de Racimo[5] (entró en vigor en 2010) y el último Tratado sobre la Prohibición de las Armas Nucleares (entró en vigor 22 de enero de 2021).[6] Todos los Estados que han ratificado estos tratados se han comprometido a no poseer, desarrollar, desplegar, probar, usar o amenazar con usar estas armas. Desde la entrada en vigor de esta clase de tratados estas armas han pasado a considerarse ilegales.

Hay tratados que no son de prohibición de un cierto armamento como el Tratado sobre Comercio de Armas,[7] Este tratado se centra en dos aspectos claves ¿qué armas convencionales quedan sometidas al tratado? armas pesadas como carros de combate, vehículos blindados de combate, sistemas de artillería de gran calibre, aviones combate, helicópteros de ataque, buques de guerra y misiles y lanzamisiles; y algunas armas pequeñas como pistolas automáticas, fusiles y carabinas, metralletas, fusiles de asalto, ametralladoras ligeras o armas ligeras como ametralladoras pesadas, cañones antitanque, fusiles sin retroceso, lanzadores portátiles de misiles antitanques y morteros. La segunda cuestión relevante ¿Cuándo queda prohibido exportar o vender armas? El tratado establece unos criterios que los estados parte deben de respetar y bajo ciertas circunstancias denegar o prohibir una exportación, por ejemplo, un estado parte no podrá autorizar una transferencia de armas, de las referenciadas en el tratado, si dicha exportación supone una violación de un embargo decretado por el Consejo de Seguridad de las Naciones Unidas. En definitiva, hay unas pocas categorías de armas reguladas, quedan muchas fuera de este tratado, y solo bajo unos pocos criterios se considera ilegal su exportación. En definitiva, podemos

afirmar que las armas convencionales son legales, no hay regulación sobre su producción y solo hay normativa internacional sobre su exportación. Cada Estado lo que si que ha regulado es quien puede tener acceso a una clase de armas y ha creado normas jurídicas sobre su uso. Las armas convencionales podemos afirmar que son legales, una pistola es legal, puede ser ilegal que un ciudadano tenga esa arma y puede estar penalizado el uso que haya echo de la misma.

En resumen, las armas biológicas, químicas, las minas antipersona, las bombas de racimo y las armas nucleares son ilegales. Las armas convencionales son legales, la legislación internacional regula su comercio y las diversas legislaciones nacionales regulan la tenencia y el uso.

Volviendo a las armas autónomas, estas han puesto de manifiesto diversas cuestiones de índole científica, ética y jurídica. Por una parte, la Inteligencia Artificial, cuyos algoritmos tienen comportamientos no explicables con una probabilidad garantizada de error significativa y no pequeña. Lo cual nos lleva a rechazar la aplicación de esta tecnología en un contexto en donde la decisión de una máquina sea matar personas humanas, incluso por error.[8] En el terreno ético cabe destacar que estas nuevas armas nos sitúan en un escenario de deshumanización, es cierto que los humanos somos falibles, pero esa es la característica de la condición humana y delegar en una máquina la decisión de matar va en contra de la dignidad humana y de los derechos de las personas. El problema ético aparece cuando los sistemas militares dejan de ser operados por personas y ejecutan sus tareas con autonomía de uso, sin intervención humana en el proceso de decisión y ataque.[9] En el terreno jurídico las armas autónomas tendrían que respetar los principios jurídicos de proporcionalidad, que analiza si los daños causados son proporcionales a los beneficios militares obtenidas o si los daños civiles son excesivos. Por otro lado, es necesario que estas armas respeten el principio de distinción, que obliga a distinguir entre combatientes y no combatientes, la cuestión es si estos sistemas de armas pueden comprender el contexto, distinguir entre un civil con miedo o un enemigo amenazante, si pueden entender las intenciones que hay detrás de una expresión humana. Finalmente, estos sistemas de armas tienen que respetar el principio de responsabilidad: si se produce un error o un crimen de guerra, ¿quién es el responsable? El soldado, quien da la orden, el político, el fabricante, el programador, ... ante la dilución de responsabilidades hay que esperar que todos los implicados intenten evadir la responsabilidad y por tanto imperará la impunidad.[10]

Teniendo en cuenta todo ello, en primer lugar, cabe aplicar el principio de precaución, detener el desarrollo de estas armas y en segundo lugar, desarrollar un instrumento jurídicamente vinculante, un tratado internacional, que prohíba el diseño, desarrollo, producción tenencia o uso de este tipo de armamento.

Notas:
1. En la web de la campaña se puede consultar las organizaciones que forman parte de esta, sus actividades y el estado de debates sobre estas armas https://www.stopkillerrobots.org/?lang=es
2. https://www.un.org/disarmament/es/adm/armas-quimicas/
3. https://www.un.org/disarmament/es/adm/armas-biologicas/
4. https://es.wikipedia.org/wiki/Convención_sobre_la_prohibición_de_minas_antipersonales
5. https://www.un.org/disarmament/es/armas-convencionales/municiones-en-racimo/
6. https://news.un.org/es/story/2020/10/1483002 o artículos que aparecen en 1 prensa de estos días https://www.eldiario.es/opinion/tribuna-abierta/entra-vigor-tratado-prohibicion-armas-nucleares_129_6945542.html
7. https://www.un.org/disarmament/es/armas-convencionales/el-tratado-sobre-el-comercio-de-armas/
8. http://centredelas.org/wp-content/uploads/2019/11/informe39_DronesArmados_RE_CAST_web_DEF-1.pdf
9. Ibid
10. Ibid

[17] ¿QUÉ PAÍSES DESARROLLAN ROBOTS AUTÓNOMOS?

Los robots y drones militares están siendo actualmente utilizados por más de 70 países. La mayoría de ellos los usan para tareas de vigilancia, y en este caso son drones no armados y con un alcance limitado. Pero, a pesar de la falta de información disponible sobre el número de drones militares existentes en los diferentes países, puede afirmarse que tan solo la OTAN posee miles de drones, con más de 60 modelos diferentes, y más de 2.200 estaciones de control terrestre.[1]

Muchos estados tienen posturas oficiales aun poco definidas, pero en la práctica prefieren optar por el uso de sistemas militares robóticos y drones militares por razones de competitividad: "Si lo hacen los otros tenemos que hacerlo nosotros, para no quedarnos atrás". Por otra parte, las armas robóticas representan una tecnología que nos abre a una nueva perspectiva de "guerra limpia" sin bajas propias, haciéndola más aceptable a nivel social.[2]

Estados Unidos, Israel, Rusia y China son, junto con algunos países europeos (Reino Unido, Francia, Alemania, Italia, Austria y España), los grandes diseñadores, fabricantes y exportadores de drones militares a nivel mundial. Y además de ellos, la lista de grandes fabricantes continúa con Corea del Sur, Turquía, Irán, Australia, la India, Ucrania y Japón.

Sin embargo, se dispone de mucha menos información en el caso de los robots y drones militares autónomos letales, por el hecho de estar clasificada. Una posible estrategia para aproximarse a una estimación de los países que desarrollan o desarrollarán drones letales autónomos es la de analizar los que están ya produciendo los llamados "loitering drones" o drones que rondan,[3] así como los que desarrollan enjambres de drones,[4] dado que estas dos tipologías pueden fácilmente incorporar mecanismos de decisión autónoma. En este caso, podemos citar a Estados Unidos, Israel, la China, Rusia, Europa, Turquía e Irán como los principales actores a nivel mundial.

Otra aproximación muy interesante es la que plantean Justin Haner y Denise Garcia. En su artículo de 2019, explican que en la actualidad, la tecnología de los robots y drones autónomos letales se concentra en unos pocos países ricos y poderosos que tienen los recursos necesarios para invertir fuertemente en robótica avanzada y en investigación en inteligencia artificial. Sin embargo, dicen, la Ley de Moore y la disminución de los costos de producción, incluida la impresión 3D, pronto permitirán a muchos actores estatales y no estatales fabricar robots asesinos.[5] Por todo ello, las armas autónomas tendrán una rápida proliferación.

Dado que la mayoría de datos relacionados con la investigación militar están clasificados, Justin Haner y Denise Garcia se han basado en un estudio de correlación en base a datos públicos, que les permite estimar el grado de desarrollo

de cada país en relación a los tres componentes críticos de las armas autónomas letales: su voluntad política (en base a las actuaciones y declaraciones oficiales), su nivel de desarrollo en lo que se refiere a investigación en inteligencia artificial, y su capacidad tecnológica para el diseño y fabricación de los sistemas de vuelo, detección y ataque. En base a todo ello, Justin Haner y Denise Garcia concluyen que los cinco mayores actores a nivel mundial en el desarrollo de armas autónomas letales son los Estados Unidos, la China, Rusia, Corea del Sur y Europa (incluyendo la Unión Europea y el Reino Unido), aunque deberían también considerarse, en segundo término, Israel, la India y Japón.

Conectando las diversas perspectivas que hemos analizado, podríamos afirmar que los principales potenciales desarrolladores de robots militares letales y autónomos son los Estados Unidos, Israel, la China, Rusia, Corea del Sur, Reino Unido y la Unión Europea. Aunque no habría que descartar otros países con fuertes intereses como Turquía, la India, Japón e Irán.

Notas:
1. Jordi Calvo (2015): entrada "Drones (aviones no tripulados" en el Diccionario de la Guerra, la Paz y el Desarme. Centro Delàs de Estudios por la Paz. Versión online: http://diccionarioguerrapazdesarme.centredelas.org/es/ - Enlace directo a la entrada: http://diccionarioguerrapazdesarme.centredelas.org/es/drones-aviones-no-tripulados/
2. Rodriguez, J., Mojal, X., Font, T., Brunet, P., (2019): "Nuevas armas contra la ética y las personas. Drones armados y drones autónomos". Informe 39. Centre Delàs d'Estudis per la Pau: http://centredelas.org/publicacions/nuevasarmascontraeticaypersonas/?lang=es
3. Véase la respuesta a la pregunta 11
4. Véase la respuesta a la pregunta 12
5. Justin Haner & Denise Garcia (2019), "The Artificial Intelligence Arms Race: Trends and World Leaders in Autonomous Weapons Development", In Global Policy Vol. 10 ③, September 2019: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1758-5899.12713

[18] ¿QUÉ EMPRESAS ESTÁN DESARROLLANDO ROBOTS AUTÓNOMOS?

Estamos asistiendo a una primera generación de nuevas armas que se empiezan a usar en el desempeño de diversas tareas como información, asesinatos o en el mismo combate. Estos nuevos sistemas marcarán las guerras del futuro, con ellas se cambiarán las doctrinas, las estrategias o las operaciones de combate en el futuro.

En tanto que hablamos de sistemas de armas nuevas, de nuevos prototipos de armas que utilizan una tecnología muy distinta a la tecnología de las armas convencionales, podemos decir que estamos asistiendo una carrera tecnológica a nivel mundial. Todos los países con capacidad tecnológica puntera y sus empresas más relevantes en este sector compiten por desarrollar nuevas tecnologías y aplicarlas al sector militar.

Dado que este sector industrial es emergente y dado que todavía estamos en los albores del diseño y generación de estos nuevos sistemas de armas, no podemos establecer un ranquin de compañías, no podemos listarlas en función de la cuota de mercado o por el éxito que hayan tenido sus prototipos.

Por esta razón ofrecemos un listado de empresas que han diseñado y producido drones armados que ya han sido utilizados o bien en el campo de batalla o que están siendo adquiridos por diversas fuerzas armadas. En el desarrollo de este mercado participan empresas que provienen del sector militar convencional y empresas cuya especialidad es el desarrollo de nuevas tecnologías. Citaremos una pequeña muestra.

General Atomics (EE. UU. www.ga.com). Ha desarrollado modelos de drones armados como el MQ-9 Reaper, MQ-9B SkyGuardian adquirido por muchos ejércitos u otros modelos no tan relevantes como el MQ-1, MQ-1 Gray Eagle o el MQ-9B Protector.

Northrop Grumman (EE. UU. www.northropgrumman.com). Con sus prototipos MQ-4C Triton (avión no tripulado de reconocimiento) y OFFSET Swarm Program, que desarrolla tácticas para enjambres de más de 250 aeronaves no tripuladas, para entornos urbanos o el modelo MQ-5B que puede llevar munición guiada por laser.

Boeing (EE. UU. www.boeing.es). Desarrolla varios modelos de aviones no tripulados de reconocimiento RQ-21 Blackjack, ScanEagle o el Switchblade con capacidad de reconocimiento y ataques

En Estados Unidos hay muchas más compañías que de dedican al diseño y fabricación de vehículos autónomos de uso militar con capacidades diversas reconocimiento o ataque. Varias de ellas muy conocidas en el sector militar como **Lockheed Martin, L3 Technologies, Raytheon o AeroVironment.** Las dos primeras, General Atomics y Northrop Grumman, son las más relevantes ya que copan el 56% de las adquisiciones por parte de las fuerzas armadas norteamericanas.[1]

Israel Aeroespace Industries (www.iai.co.il). Ha desarrollado el modelo Heron (o también Majatz-1), es un dron de combate.

Ebit Systems (Israel https://elbitsystems.com). Ha desarrollado el Hermes 900, un UAV de reconocimiento y vigilancia.

China Aerospace Science and Technology Corporation (CASC) (http://english.spacechina.com/n16421/index.html). Ha desarrollado el dron de combate CH-5 (conocido como Rainbow), muy parecido al norteamericano MQ-9 Reaper,

Aviation Industry Corporation of China (AVIC) (https://enm.avic.com). Ha diseñado el UAV el Wing Loong II, que puede utilizarse para vigilancia y ataque.

Turkish Aerospace Industries (TAI) (www.tusas.com/en). Ha desarrollado el modelo Anka diseñado para misiones de reconocimiento, vigilancia, incluye un sistema de identificación amigo/enemigo (IFF), buscador laser, munición inteligente y lanzamisiles.

STM Defense Technologies Engineering (Turquía www.stm.com.tr/en). Ha diseñado cientos de drones, de los cuales destacar el Kargu-2, conocido como "dron Kamikaze" dotado de un alto grado de autonomía es similar a la munición merodeadora, con capacidad de localizar, rastrear e identificar objetivos sin ayuda humana, dispone de reconocimiento facial por lo que puede rastrear individuos específicos.

Kalashnikov Group (Rusia https://en.kalashnikovgroup.ru). Esta desarrollando un dron suicida KUB-BLA, en esencia es un misil crucero pequeño lento y económico, similar al Harpy israeliano que se lanza desde un camión; diseñado para destruir los sistemas antiaéreos.

Kronshtadt Group (Rusia). Desarrolla el dron Orión, que puede disparar misiles guiados y bombas de deslizamiento.

La lista de empresas que se dedican a diseñar prototipos de drones armados dotados de autonomía humana es bastante larga, todos los países dotados de equipos de investigación en nuevas tecnologías destinan ingentes recursos a la investigación, no quieren quedarse atrás en esta carrera tecnológica. No hemos citado empresas indias o de los países de la Unión Europea, pero en todos estos países se están desarrollando prototipos de UAV para uso militar.

Nota:
1. Según Michael Peck en su articulo *Four companies dominate the military drone market*. Hay cuatro compañías que dominan el mercado de UAV militares en Estados Unidos https://www.c4isrnet.com/unmanned/uas/2016/04/06/four-companies-dominate-the-military-drone-market/

CAMPAIGN TO STOP
KILLER ROBOTS

www.stopkillerrobots.org
@SKR_Espana + @BanKillerRobots
@stopkillerrobots

Material elaborado por el equipo de la campaña SKR en España y por el grupo de investigación sobre armas autónomas del Centre Delàs d'Estudis per la Pau y la UAB (Pere Brunet, Tica Font y Joaquín Rodríguez)

CENTRE DELÀS D'ESTUDIS PER LA PAU

UAB Universitat Autònoma de Barcelona