

# Inteligencia artificial y armas autónomas: una combinación letal

TICA FONT

Iniciemos este artículo estableciendo la diferencia entre sistema de armas automáticas y sistemas de armas autónomas letales. Disponemos de muchos tipos de armas automáticas o automatizadas como los “sistemas antimisiles”, o sistemas de defensa activa o los usados en navíos de combate que, utilizando sensores, radares, y ordenadores detectan misiles y cuando están situados a menos de una cierta distancia, de manera automática disparan para destruir el misil, cohete de artillería o embarcaciones. Entre estos modelos estaría el escudo antimisiles ABM, diseñado para interceptar misiles enemigos antes de que lleguen a su destino.

Estos sistemas de armas utilizan algoritmos de toma de decisión precisos, establecidos mediante criterios del operador humano. En este grupo de armas tenemos las armas centinela, utilizadas por Corea del Sur o Israel, armas estacionadas en la frontera o en el muro Israel-Palestina. Suelen ser torretas dotadas de diversas clases de sensores y radares, conectadas a ordenadores y dotadas de diversos tipos de proyectiles que identifican vehículos o personas que se aproximan a zonas o franjas terrestres, en caso de detectar un intruso, le avisan y, si no retrocede, disparan. Estos sistemas de armas automatizados tienen capacidad de respuesta sin que un humano dé la orden. Ello se justifica por la necesidad de respuestas rápidas –no hay tiempo para hacer consultas–. Se trata de armas de carácter defensivo y sus algoritmos de toma de decisiones están muy definidos y preestablecidos. Estos algoritmos siguen ecuaciones matemáticas o con base a la física, no tienen parámetros que haya que ajustar y los errores potenciales provienen de un mal uso o de errores humanos. Si hay un fallo siempre podremos saber qué parámetro ha fallado y podremos corregirlo. En el caso de los algoritmos de las armas centinelas, los algoritmos, a través de los sensores, tienen que diferenciar entre un humano y un animal cruzando la frontera y suelen hacerlo con reglas preestablecidas basándose en altura, la forma corporal o la velocidad.

En una segunda categoría se encuentran los llamados sistemas de armas autónomos. La diferencia entre un sistema de armas automático o automatizado y un sistema de armas autónomo, se encuentra en la estructura del sistema de toma de decisiones. En las armas automatizadas el sistema de toma de decisiones se construye con algoritmos que siguen reglas ya programadas, definidas, bien definidas. En los sistemas de armas autónomas, el sistema de toma de decisiones se basa en un modelo de realidad obtenido con redes neuronales de aprendizaje profundo a partir de datos.

Los algoritmos de aprendizaje automático son complejos, se basan en una gran cantidad de parámetros, centenares de millones, que requieren conjuntos de datos

**En los sistemas de armas autónomas, el sistema de toma de decisiones se basa en un modelo de realidad obtenido con redes neuronales de aprendizaje profundo**

enormes para entrenarse, muchos más datos que parámetros haya que ajustar en el entrenamiento o durante el aprendizaje. Estos algoritmos son opacos y llevan asociado un grado de error que es intrínseco al sistema y, por tanto, su fiabilidad es limitada. Si un sistema basado en algoritmos de aprendizaje profundo falla, no podemos explicar

qué neurona o grupo de neuronas ha fallado, por eso se les suele llamar “cajas negras”.

Hay aplicaciones que utilizan sistemas de aprendizaje automático como son los traductores, son útiles y todos somos conscientes que cualquier traducción requiere una supervisión posterior por parte del usuario en la fase final, y son las personas las que finalmente asumen la responsabilidad de la traducción. Esto no pasa con los sistemas (sean militares o civiles) de aprendizaje automático dotados de autonomía de uso, ya que por definición no contemplan ningún tipo de postsupervisión humana y el resultado pueden ser víctimas civiles.

En el caso de un vehículo autónomo, somos conscientes que un error del sistema puede provocar accidentalmente la muerte de alguna persona. En el caso de aplicar estos algoritmos a armas, o utilizarlos para uso militar, la cuestión será si el error del sistema es aceptable o no. Todavía más: si aceptamos que sea un sistema el que tome la decisión, sin supervisión humana, de llevar a cabo una actuación que provocará víctimas humanas. En definitiva, vamos a centrar este debate en si es aceptable el uso de sistemas militares dotados de autonomía.

Los sistemas militares con autonomía de uso o armas autónomas llevan a cabo cinco tareas clave: 1) La tarea de búsqueda (*find*), permite obtener una lista de potenciales objetivos, lista que puede ser elaborada por humanos o por el mismo sistema. Los drones merodeadores, los de reconocimiento y los enjambres de drones están especialmente preparados para llevar a cabo esta tarea; 2) la tarea de definir un objetivo (*fix*), seleccionar un objetivo al que atacar; 3) la tarea de seguimiento del objetivo (*tracking*), mantener localizado el objetivo, sea móvil o fijo, y seguirlo en caso de que sea móvil; 4) la tarea de tomar la decisión de ataque (*target*); y 5) la tarea de atacar (*engage* o *destroy*). De todas estas tareas hay dos especialmente críticas y que en todos los sistemas militares clásicos quedan a cargo de los militares; son las de *fix* y *target*. Pero los sistemas militares con autonomía de uso se caracterizan por no tener ninguna supervisión humana es estas dos tareas, ni en ninguna de las cinco.

En los sistemas de armas automatizados el operador militar tiene la opción de parar la acción que se lleva a cabo durante un tiempo limitado; en el caso de sistemas de armas autónomos el operador no tiene esta opción.

## Reto jurídico de los sistemas de armas autónomas letales

Como se ha mencionado, los sistemas de armas autónomas letales (*Lethal autonomous weapons* o LAWs) una vez activados, pueden seleccionar y atacar objetivos sin intervención adicional de un operador humano. A diferencia de los sistemas de armas defensivos automatizados o altamente automatizados, las armas autónomas podrían operar fuera de límites espaciales y temporales estrechos, podrían aprender a adaptar su funcionamiento en respuesta a circunstancias cambiantes en el entorno en el que se las despliegan.<sup>1</sup>

Si alguien activa un arma autónoma, no sabe exactamente a que personas u objetos atacará, ni tampoco sabe con precisión dónde ni cuando se producirá el ataque. Esto se debe a que estas armas se activan mediante sensores y *software* que comparan aquello que los propios sensores detectan en el entorno con un “perfil de objetivo”. El sistema podría activarse por la forma de un vehículo o por

<sup>1</sup> Joaquín Rodríguez, Xavi Mojal, Tica Font y Pere Brunet, (2019), *Nuevas armas contra la ética y las personas. Drones armados y drones autónomos*, Centre Delàs, Informe núm. 39, disponible en: [https://arxiu.centredelas.org/images/INFORMES\\_i\\_altres\\_PDF/informe39\\_DronesArmados\\_CAST\\_web\\_DEF.pdf](https://arxiu.centredelas.org/images/INFORMES_i_altres_PDF/informe39_DronesArmados_CAST_web_DEF.pdf)

identificar a una persona con un cierto perfil. En este caso el vehículo o la persona serían los activadores de un sistema que acabaría en un ataque.

En la actualidad se han diseñado algunas clases de armas autónomas, aunque solamente se utilizan en circunstancias muy limitadas, como son las municiones

**Los sistemas de armas autónomas letales, una vez activados, pueden seleccionar y atacar objetivos sin intervención adicional de un operador humano**

tipo *loitering* o “drones kamikazes”, es una combinación de dron armado o misil. Estos dispositivos son capaces de permanecer en posición sobre una zona de objetivos, realizando un reconocimiento y proporcionando datos o imágenes para la asignación de un objetivo en vuelo. Este tipo de sistema permite que el ataque se lleve a cabo en el momento preciso o abortar la misión en caso de que no se detecte el objetivo o no se den las condiciones óptimas para el

ataque. En esta situación, el operador humano permanece involucrado en todo el proceso y es él el que toma la decisión del ataque.

El desarrollo de este tipo de armamento y su posible uso ha abierto un debate sobre si es necesario regular el uso de estos sistemas o si hay que regular el propio sistema. ¿Dónde hay que situar los límites? La segunda cuestión es si ante los avances producidos por estas tecnologías, tenemos que modificar, adaptar o replantear el derecho internacional humanitario y los derechos humanos a estas tecnologías, o al revés, si son las tecnologías las que tienen que adaptarse a las normas internacionales.

En este trabajo tomaremos la posición de analizar estas armas a la luz de la legislación internacional existente. La primera consideración es que hay muy pocas normas sobre las armas. Como determina el Comité Internacional de la Cruz Roja (CICR), el Derecho Internacional Humanitario dice que «el uso de medios y métodos de guerra cuya naturaleza cause lesiones o sufrimientos innecesarios está prohibido» (norma 70); y que «el uso de armas que por su naturaleza sean indiscriminadas está prohibido» (norma 71). De ahí que cualquier arma que necesariamente cause lesiones o sufrimientos excesivos o no pueda estar dirigida hacia un objetivo militar específico debería estar prohibida y no ser utilizada nunca.<sup>2</sup>

---

<sup>2</sup> Jaume Saura, «Algunas implicaciones del empleo de drones en perspectiva jurídica internacional», en *El arma de moda: impacto del uso de los drones en las relaciones Internacionales y el derecho internacional contemporáneo*, ICIP Research, núm. 4, ICIP, Barcelona, 2014, pp. 133.

Las armas que han sido prohibidas mediante tratados específicos y que actualmente forman parte del derecho internacional vigente son: las armas químicas, armas biológicas, las minas antipersona, las bombas de dispersión y las armas nucleares.

Desde esta perspectiva del derecho internacional debe asegurarse que los sistemas de armas autónomos puedan cumplir con los principios de responsabilidad, distinción y proporcionalidad.

## Principio de responsabilidad

Cuando se cometa una violación de derecho internacional humanitario es fundamental establecer el responsable de los actos para poder hacer justicia y disuadir de futuras violaciones.

En general, las investigaciones por ataques militares se centran en la persona que disparó el arma y el jefe militar que dio la orden de ataque. Pero cuando se utilizan armas autónomas, ¿quién dará explicaciones de por qué se atacó un mercado o una escuela? En caso de error o un acto de crimen de guerra, ¿quién será el responsable?, ¿el programador?, ¿el fabricante?, ¿el militar a cargo del sistema?, ¿el político? ¿y si hubiera sido usado por un agente privado? Definir con claridad en quién recae la responsabilidad de los actos llevados a cabo por un sistema de arma autónomo es fundamental para garantizar, por un lado, la redición de cuentas y, por otra parte, para asumir las consecuencias penales que puedan derivarse de la actuación.

De entrada, sería el militar el que asumiría la responsabilidad del sistema de arma autónoma, pero si él no lo ha activado o desconocía las intenciones del blanco (en el caso de ser una persona), ¿cómo puede ser el responsable? La experiencia nos dice que salvo en acciones comprobables en las que el militar actúe de manera imprudente, él no sería el responsable. Ello conduce a la impunidad.

Cualquier arma es susceptible de fallar o sufrir accidentes. En este caso, un fallo podría provocar que el sistema cometa un error y que seleccione y ataque inapropiadamente un cierto número de objetivos. Las víctimas de este tipo de armas pierden el derecho a saber la verdad, pierden el derecho de justicia y de

reparación. Porque es evidente que un sistema de armas autónomo no tiene personalidad jurídica, no puedes llevarlo a los tribunales y no puede ser castigado por ello.

## Principio de discriminación o distinción

Este principio obliga a las partes beligerantes en un conflicto armado a distinguir en todo momento sus ataques entre combatientes y objetivos militares, por un

**Definir con claridad en quién recae la responsabilidad de los actos llevados a cabo por un sistema de arma autónomo es fundamental**

lado, y a personas y bienes civiles por el otro. Por ejemplo, los algoritmos de la IA tendrían que poder distinguir entre un pastor pastún armado con un kalashnikov y un miembro armado talibán, o entre un cazador y un combatiente. También tendría que poder distinguir entre combatientes, sean comba-

tientes activos o combatientes que han quedado fuera de combate (se han rendido, están heridos o porque son prisioneros); entre civiles que participan activa o directamente en las hostilidades y civiles armados, como pudiera ser el personal de seguridad.

Por otra parte, los algoritmos de la IA tendrían que comprender el contexto en el que actúan. Si, por ejemplo, pueden identificar combatientes heridos, o que están en disposición de rendirse o combatientes con miedo. Estamos hablando de que los algoritmos de la IA puedan comprender intenciones. Muchas de las actuaciones previas al uso de la fuerza armada implican la toma de decisiones subjetivas, intuitivas, que requieren del juicio humano como por ejemplo distinguir entre un civil con miedo o un enemigo amenazante. Percibir, entender una situación, implica entender las intenciones que hay detrás de una expresión humana.

La dificultad o imposibilidad de que un sistema armado interprete o evalúe intenciones o emociones es un factor clave para decidir si es correcto o si es legal el uso de robots armados autónomos. Por tanto, si un arma o sistema de armas autónomo que inherentemente no pueda discriminar debe quedar prohibido.<sup>3</sup>

---

<sup>3</sup> Posición de la Comisión de la Cruz Roja Internacional sobre estas armas, CICR, 12 de mayo de 2021 [en línea], disponible en: <https://www.icrc.org/es/document/posicion-del-cicr-sobre-los-sistemas-de-armas-autonomos>

## Principio de proporcionalidad

La norma de proporcionalidad exige que, antes de atacar, debe evaluarse el daño que pueda causarse a la población civil en cuanto a la ventaja militar obtenida con la acción. Este principio se aplica a actuaciones concretas con el fin de valorar si los daños causados eran proporcionales en relación con la ventaja militar que se obtuvo, o si, por el contrario, los daños fueron excesivos.

La proporcionalidad es propia del discernimiento humano y se basa en conceptos como el sentido común, actuar de buena fe o una orden razonable. Las interpretaciones jurídicas imperantes se basan explícitamente en estos conceptos de sentido común o de buena fe. Queda por determinar en qué medida estos conceptos pueden traducirse en algoritmos y programas tanto en la actualidad como en futuro.

Analizar el principio de proporcionalidad en cualquier actuación militar es muy complejo y es un ejercicio puramente cualitativo y difícil de cuantificar. Es imposible obtener un consenso sobre cuantos civiles es “proporcional” matar o, en una actuación determinada cuantos civiles o daños colaterales son “excesivos”. Según Markus 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 y no es posible llegar a consensos.<sup>4</sup>

Por tanto, volvemos a hacernos la misma pregunta: ¿es posible que los algoritmos de la IA actual o en el futuro, puedan realizar este tipo de valoración o evaluación tan dependiente de valores y de contexto?

## La ética y los sistemas de armas autónomos

Los defensores de desarrollar armas autónomas piden no poner límites al desarrollo de estas armas. Alegan que los humanos, bajo ciertas condiciones como calor, rabia, miedo, ira, rencor o deseo de venganza, actuamos de la peor manera posible. Los robots, es cierto, no pueden actuar bajo estos estados de ánimo, no

<sup>4</sup> Markus Wagner M., (2014), «The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems», *Vanderbilt Journal of Transnational Law*, 47, 1371-1424, disponible en: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2541628](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2541628)

**Según los desarrolladores, los robots no pueden actuar bajo estados de rabia o rencor, por lo que podrían evitar muertes innecesarias**

pueden tener sentimientos como estos y, por tanto, podrían evitar muertes innecesarias. Los defensores de no limitar la creación de armas autónomas letales argumentan que las armas completamente autónomas pueden ser capaces de cumplir con los tres principios del derecho internacional humanitario y que pueden cumplirlo mejor que los humanos,

ya que pueden procesar más información, más rápidamente que los humanos y no estar inhibidos por preservarse a sí mismos, ni estar influenciados por las emociones.

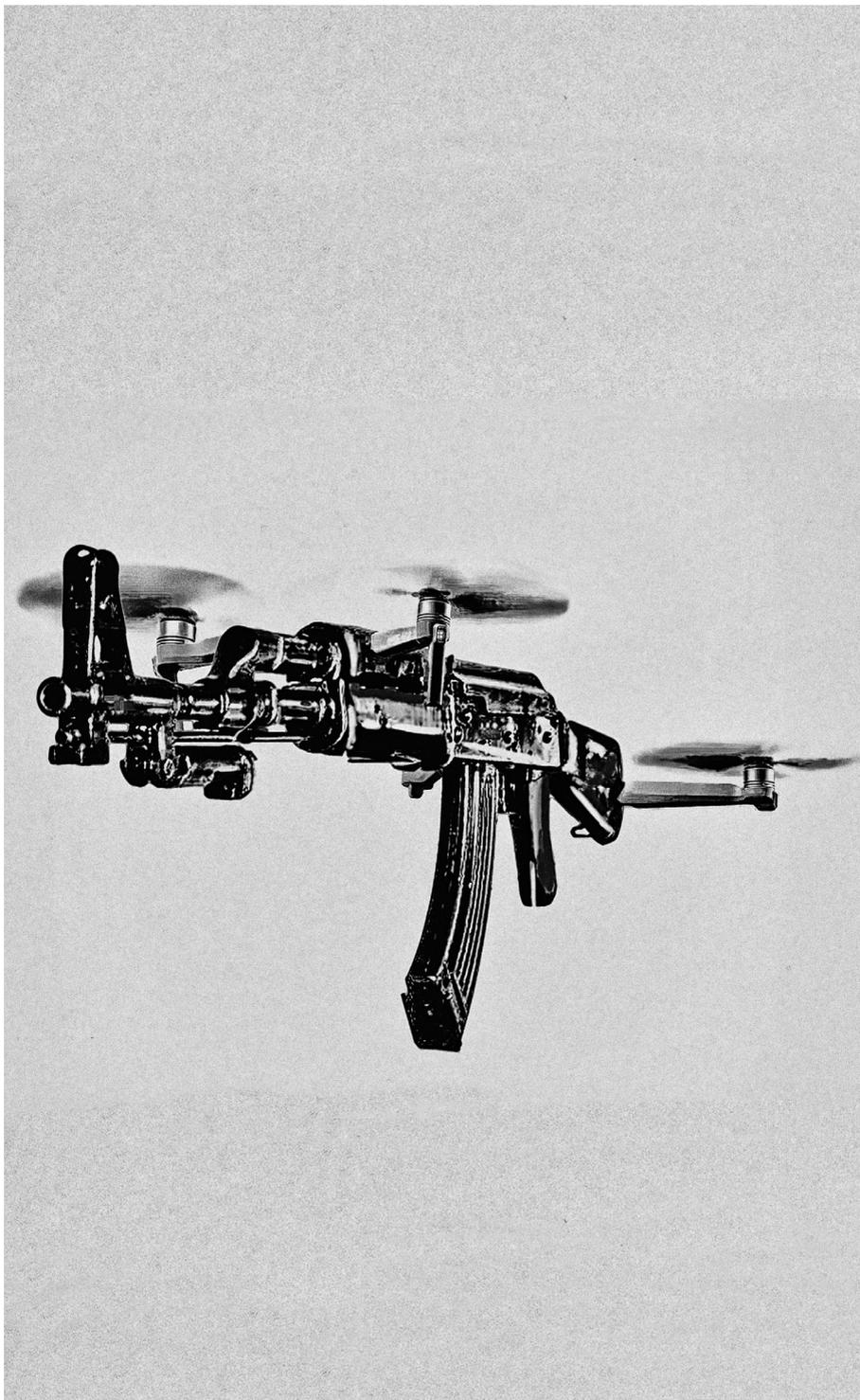
Los defensores de no limitar el desarrollo de estas armas totalmente autónomas nos recuerdan que en situaciones de conflicto armado los humanos cometen muchas vilezas como violar a mujeres y niñas o torturar.

Es cierto que los robots y las armas totalmente autónomas, a menos que estén programados específicamente para ello, no violarían a mujeres, no torturarían y no causarían daños intencionadamente a la población civil.

Es cierto que estas armas no tienen sentido de riesgo, salvo que se las programe específicamente para preservarse, no tienen miedo y no toman decisiones influidos por emociones. Pero hay que reconocer que hay emociones o sentimientos que representan una salvaguarda. Sin emociones se puede matar más fácilmente. Los sistemas de armas totalmente autónomas no tendrán el sentido común de los humanos, no pueden sentir compasión, lástima, empatía o intuición. Todos estos sentimientos son una salvaguarda frente a la atrocidad.

Es cierto que los humanos somos falibles y estas máquinas no –se las programa para ejecutar misiones militares de ataque–, pero el ser falibles es la condición que nos hace diferentes del resto de especies, representa la esencia de la humanidad. Estas armas no sienten y no pueden sentir culpabilidad por sus actos; en cambio un humano sí.

Estas armas pueden ser programadas para abordar y evaluar situaciones de manera cuantitativa, pero su capacidad para hacerlo de manera cualitativa es y será muy limitada. Estas habilidades son muy necesarias cuando se trata de decidir sobre la vida o la muerte de seres humanos. Lo mismo podemos decir sobre



la capacidad de los sistemas de armas totalmente autónomos para distinguir y evaluar entre órdenes lícitas o ilícitas, o su capacidad para interpretar un contexto, evaluarlo en cálculos basados en valores. Por ejemplo, un arma totalmente autónoma podría decidir disparar sobre un niño que lleva un arma en sus manos, su decisión podría estar conforme a la ley, el niño iba armado, pero esta actuación puede no ser considerada ética. Un militar ante esta misma situación puede recordar o evocar a su hijo, puede pensar que él puede no matar al niño y buscar una solución alternativa como puede ser capturarlo.

Hasta aquí hemos abordado la utilización de la IA aplicada a una clase específica de sistemas de armas, las autónomas, que pueden ser utilizados en el campo de batalla. Estamos considerando sistemas de armas que sustituyen a los humanos en las tareas del combate. Pero también podemos pensar en que estos sistemas pueden ser utilizados en otras tareas éticamente cuestionables como interrogar o torturar a sospechosos. En este contexto cabe tener presente que un robot, a diferencia de un médico, no ha hecho un juramento hipocrático de no hacer daño a un humano; a diferencia de un humano, ¿cómo llevará a cabo el seguimiento de los signos vitales o del dolor de un humano que está siendo interrogado o torturado?<sup>5</sup>

## Impacto político de los sistemas de armas totalmente autónomos

Hay un consenso general en que la aplicación de la IA en todos los aspectos militares producirá cambios de gran calado en las guerras, y no solamente en términos de capacidades militares, sino que cambiará los enfoques, las estrategias, las tácticas y las teorías de cómo hacer la guerra.

Los nuevos sistemas de armas autónomas están supervisados por militares que pueden estar situados a cientos o a miles de kilómetros. El militar está muy lejos del llamado campo de batalla, y no se requiere su presencia física en este espacio geográfico. De esta manera, los militares pueden hacer la guerra a turnos, acabado su turno, el soldado vuelve a casa a cenar y revisar los deberes escolares con los hijos. Según Singer,<sup>6</sup> este tipo de armas interpone una gran distancia física

---

<sup>5</sup> Patrick Lin, «Drone-Ethics Briefing: What a Leading Robot Expert Told the CIA», *The Atlantic*, 15 de diciembre de 2011, disponible en: <https://www.theatlantic.com/technology/archive/2011/12/drone-ethics-briefing-what-a-leading-robot-expert-told-the-cia/250060/>

<sup>6</sup> P. W. Singer, *Wired for War. The Robotics Revolution and Conflicts in the 21st Century*, Penguin Press, Nueva York, 2009.

entre el ser humano, la propia arma y las consecuencias de su utilización. Tal distanciamiento físico puede provocar un distanciamiento psicológico y moral o una disminución del sentido y de la conciencia de la responsabilidad de la actuación. Con los sistemas de armas autónomas, las guerras serán más fáciles de empezar, las barreras morales, éticas y psicológicas serán más débiles, los políticos pueden poner menor resistencia a involucrarse en el desarrollo de las guerras –ya que no estarían tan presionados por el rechazo de los ciudadanos a las pérdidas de las vidas humanas– y los políticos pueden reducir esfuerzos en buscar soluciones no violentas a un conflicto. En definitiva, la utilización de toda esta generación de nuevas armas en combate nos abre el debate sobre la banalización del hecho bélico a escala política y social.

**Hay un consenso general en que la aplicación de la IA en todos los aspectos militares producirá cambios de gran calado en las guerras**

Estos nuevos sistemas de armas suponen una tecnología diseñada para disminuir las bajas de soldados en combate, que no quiere decir que mueran menos civiles en las guerras y que no mueran más personas por ataques de robots o armas autónomas que por armas convencionales. El uso de robots en la guerra eliminará el obstáculo político de hacer la guerra; provocará que los políticos eviten asumir ante la sociedad lo que representa el retorno de los cadáveres de soldados muertos. Cada vez más los políticos piden a los militares que diseñen intervenciones bajo la doctrina de “cero muertos”, es por eso por lo que impulsan el desarrollo de estas nuevas armas, les es más fácil enviar a los robots a hacer la guerra, que no enviar a nuestros hijos y tener que hacer cartas de pésame a las familias de los soldados muertos. Ante la opinión pública los políticos evitarán la responsabilidad, ya que con la utilización de los robots deshumanizamos la guerra y eliminamos el factor humano. La utilización de robots en el combate puede conseguir hacer más aceptable la guerra para la sociedad y además puede evitar la “fatiga del combate” que provoca una guerra de larga duración.<sup>7</sup>

Armas autónomas, drones y otras municiones presentan ventajas: los militares no pueden ser abatidos y además suponen menor coste económico. El valor de un dron es mucho menor que un avión de combate. En definitiva, nos lleva a la guerra *low cost*, no solamente en términos económicos sino también en términos éticos

<sup>7</sup> Pere Vilanova, «Drones y política exterior: un instrumento de múltiples lecturas», en: *El arma de moda: impacto del uso de los drones en las relaciones internacionales y el derecho internacional contemporáneo*, ICIP Research núm. 4, 2014, pp. 73.

y políticos. Estas nuevas armas ofrecen mayor proyección de fuerza con menor riesgo para la vida de los militares; permite más acciones militares con menos soldados. Ello nos lleva a preguntarnos si la vida de los militares es más valiosa que la de los civiles, si hay vidas que tienen más valor que otras, si hay vidas de las que se puede prescindir. También nos interpela como sociedad. Nos dice que tenemos una doble moral en relación con los sacrificios que estamos dispuestos a asumir; una doble moral ante una sociedad que demanda intervenir militarmente sin sacrificios humanos propios y sin riesgo para los gobernantes políticos.

## Conclusiones

La aplicación de la IA nos abre la puerta a plantear las guerras sin riesgo para la vida de los militares, con lo que comporta prevalecer las soluciones militares por encima de las soluciones políticas. Hace más fácil pensar en guerras a distancia y las hace más abstractas.

Estas armas nos sitúan en un escenario de deshumanización. La narrativa de las bondades de la IA tiende a ignorar aquellos aspectos que afectan al sentido de la humanidad y que afectan o cuestionan los derechos civiles y derechos humanos.

Como hemos recogido al inicio del artículo, los sistemas militares dotados de IA con autonomía constructiva, tienen comportamientos no explicables, con una probabilidad garantizada de error que es significativa, de aquí que si estos sistemas de nuevas armas no pueden cumplir con los principios jurídicos del derecho internacional humanitario –el principio de responsabilidad, de distinción y de proporcionalidad– hay que considerar que estas armas son indiscriminadas y, por tanto, tendrían que ser ilegales. Finalmente, señalar que deberíamos dotarnos de un instrumento jurídicamente vinculante que prohibiera su desarrollo y uso.

**Tica Font Gregori** es experta en economía de la defensa y comercio de armas y fundadora e investigadora del Centre d'Estudis per la Pau J.M. Delàs.

