



La imagen representa cómo los avances tecnológicos están transformando la percepción que los humanos tienen de sí mismos y de su identidad. Por ello, el autor del artículo examina cómo los principios de guerra justa tendrán que cambiar de forma radical conforme los avances en los sistemas de armas autónomos. (Foto: Pixabay)

# Robots de combate y precauciones a considerar

## Las limitaciones cognitivas de los combatientes humanos y autónomos

Mayor Jules Hurst, Reserva del Ejército de EUA

**D**urante la Segunda Guerra Mundial, los pilotos dependían de calculadoras analógicas y la heurística para que sus municiones dieran en el blanco. En los 80, las computadoras de selección de blancos notificaban con precisión a los pilotos cuando tenían que lanzar su munición, y si el piloto no lanzaba la munición en el momento exacto, podía errar el blanco por una gran distancia, dado que su precisión dependía de la altura, la velocidad y la orientación de la aeronave. En la actualidad, las amplias regiones aceptables de lanzamiento de las municiones guiadas de precisión no hacen necesario tener mucha destreza para acertar en el blanco. Si una munición guiada de precisión falla un blanco que ha sido identificado positivamente, el error probablemente se debe a la ojiva y no a la técnica de lanzamiento del piloto.

Esta situación será cada vez más común mientras sigan proliferando los sistemas semiautónomos en el campo de batalla. A medida que las computadoras adquieran más responsabilidades de combate, las fallas de equipo supondrán un aumento mayor en el porcentaje de errores del empleo del arma. Los soldados pueden lidiar fácilmente con las tasas de errores y las probabilidades de error circular de las armas de precisión, pero los errores que los robots de combate semiautónomos podrían cometer son más difíciles de predecir. Un análisis preciso de estas fallas tal vez requiera conocimiento de los algoritmos y el código que llevaron a las decisiones del robot. En la actualidad, un caza F-35 del programa Joint Strike Fighter requiere 8 millones de líneas de código para funcionar y en el futuro, los sistemas semiautónomos necesitarán más<sup>1</sup>.

Dado lo complejo que será el proceso de toma de decisiones de los robots de combate en el futuro, los combatientes probablemente experimentarán cada vez más dificultades para seguir los principios de guerra justa, o *ius in bello*. Aparte de contar con mayores capacidades, los robots de combate tendrán más libertad de acción en el campo de batalla debido a que sus parámetros para la toma de decisiones por lo general serán configurados antes de salir de fábrica y probablemente no podrán ser modificados por los soldados que los operarán. Para la sociedad, la responsabilidad ética es proporcional a la capacidad del individuo o de la organización para controlar las acciones en cuestión. Cuando los sistemas autónomos letales proliferen en el campo de batalla, los combatientes tal vez no sean los principales

encargados de controlar sus actividades. En su lugar, los funcionarios de adquisiciones y los ingenieros que obtienen y construyen estos sistemas probablemente serán los que tendrán mayor influencia en el comportamiento de los robots de combate en el campo de batalla. Por lo tanto, como las personas no combatientes en la cadena de adquisiciones de sistemas de armas son cada vez más responsables de la conducta ética de las máquinas en el campo de batalla, los conceptos de guerra justa deben ser actualizados.

## ¿Cuáles sistemas autónomos?

No todos los sistemas de armas autónomos letales (LAWS) dificultarán los esfuerzos de los combatientes para respetar los principios de guerra justa. Si los avances en las redes neuronales y la informática hacen posible que los LAWS tomen decisiones y perciban como los seres humanos, es lógico deducir que los sistemas autónomos tendrán las mismas preocupaciones con respecto a la guerra justa que los combatientes humanos. Pero sin duda alguna, habrá un período de transición entre la consecución de un nivel de inteligencia artificial similar al de los seres humanos y el empleo de robots de combate en el campo de batalla.

Los pronósticos de los expertos sobre cuándo surgirá una inteligencia artificial similar a la de los seres humanos varían considerablemente. En 2012, analistas del Instituto de Investigación de la Inteligencia de las Máquinas (MIRI) analizaron 257 predicciones de expertos y no expertos respecto a cuándo las máquinas alcanzarían niveles cognitivos comparables a los de los seres humanos. Estas predicciones iban desde el año 1980 hasta más allá del 2100, con la mayoría situándose entre 2020 y 2060<sup>2</sup>. No existe un consenso sobre cuándo surgirá una inteligencia artificial similar a la de los seres humanos, ni tampoco una definición precisa del término. Además, la aparición de máquinas con inteligencia humana no necesariamente significará que estas tendrán la capacidad para percibir sus entornos de igual manera que los humanos<sup>3</sup>. La inteligencia y la percepción sensorial son habilidades independientes. Si no lo cree,

**El mayor Jules «Jay» Hurst, Reserva del Ejército de EUA,** ha servido como oficial de operaciones en la Célula Analítica de Tecnología e Innovación del Comando de Operaciones Especiales y como analista de inteligencia civil superior en el 1<sup>er</sup> Batallón del 75<sup>o</sup> Regimiento Ranger.

atrévase a hacerle esa pregunta a alguien con discapacidad auditiva o visual y verá lo que le responde. Una inteligencia artificial similar a la inteligencia humana tal vez demore mucho tiempo en desarrollarse, e incluso una vez que se obtenga, es posible que se necesite emplear una computadora del tamaño de una habitación o procesadores cuánticos en temperaturas bajo cero<sup>4</sup>. Probablemente pasará mucho tiempo antes de que la inteligencia artificial similar a la inteligencia humana sea lo suficientemente rentable para incorporarla en sistemas de armas tácticos, como lo ocurrido con las computadoras personales, las cuales tomaron décadas para ser viables financiera y tecnológicamente<sup>5</sup>.

### **El período de transición entre una inteligencia artificial similar a la inteligencia humana y robots de combate competentes**

Independientemente de cuándo los robots lleguen a alcanzar capacidades intelectuales y sensoriales similares a las de los seres humanos, seguramente habrá un período de transición antes de que esto ocurra en el que las naciones emplearán los LAWS en el campo de batalla debido a las ventajas que estos proporcionan cuando ejecutan tareas claramente definidas<sup>6</sup>.

Con pocas excepciones, las máquinas pueden superar a los humanos en cualquier función que esté definida claramente. El empleo de sistemas de armas automatizados en tareas rutinarias específicas ya ofrece a Estados Unidos y a otras fuerzas sofisticadas enormes ventajas tácticas. Estas ventajas aumentarán paralelamente con los avances tecnológicos y mientras más se empleen estos sistemas, mejor será su integración en la cultura militar. En la actualidad, el empleo de máquinas con niveles de inteligencia artificial subhumana en tareas estructuradas (inteligencia artificial débil) repercute profundamente en el poderío militar de una nación y a largo plazo, esto será cada vez más importante<sup>7</sup>. Las naciones se verán fuertemente tentadas a permitir que los robots tengan autonomía letal antes de alcanzar niveles cognitivos y sensoriales humanos.

La política actual de Estados Unidos solo autoriza el desarrollo de vehículos robóticos que «permitan a los comandantes y a los operadores ejercer niveles apropiados de juicio humano sobre el empleo de la fuerza», pero es posible que las fuerzas militares estadounidenses ignoren esta directriz durante un

conflicto por razones prácticas<sup>8</sup>. A menos que la comunidad internacional imponga con éxito un régimen de control de armas que prohíba los robots asesinos o limite su empleo mediante el derecho internacional o normas establecidas, las naciones beligerantes podrían fácilmente justificar el empleo de los LAWS como una necesidad en conflictos prolongados y hasta limitados<sup>9</sup>. Incluso si la comunidad internacional logra establecer un régimen de control de armas o una legislación que gobierne el empleo de sistemas autónomos, la historia demuestra que los Estados rápidamente violan estos acuerdos cuando consideran necesario o cuando lo estiman conveniente. En la década de los 30, tanto el Imperio japonés como el Tercer Reich se retiraron o violaron las disposiciones relacionadas con las armas del Tratado Naval de Washington y el Tratado de Versalles<sup>10</sup>. Los regímenes más liberales también tienden a ignorar los acuerdos y las disposiciones internacionales según sus intereses. Apenas hace 15 años, el presidente George W. Bush retiró a Estados Unidos del Tratado de Misiles Antibalísticos que se firmó con la Unión Soviética en 1972 para empezar la construcción de un sistema nacional de defensa contra misiles balísticos<sup>11</sup>.

Cuando una parte importante de un tratado de control de armas se retira para obtener una ventaja militar, los signatarios y potenciales adversarios pierden los incentivos para cumplir con las provisiones; una manzana podrida echa a perder toda la caja. Si emplear sistemas letales con inteligencia artificial débil supone una gran ventaja militar, los países rápidamente podrían modificar sus restricciones internas o ignorar los tratados y las leyes. Estos cambios de política incluso podrían ser producto de actividades militares bien intencionadas. Occidente continúa desplegando decenas de miles de soldados para combatir las amenazas terroristas en África, Asia y el Oriente Medio. El despliegue de LAWS en vez de soldados humanos tal vez sea una alternativa atractiva para regímenes democráticos que necesitan conciliar las demandas de seguridad nacional del electorado con su renuencia a aceptar bajas de combate<sup>12</sup>.

Aunque la comunidad internacional lograra prohibir el empleo de robots asesinos, los sistemas autónomos todavía podrían tener la oportunidad de tomar decisiones letales en el campo de batalla. Suponiendo que las naciones empleen sistemas robóticos ofensivos,



Un misil estándar 3 (SM-3) es lanzado desde el USS Decatur, un destructor de la clase Arleigh Burke equipado con un sistema de combate Aegis, el 22 de junio de 2007, durante una prueba de misiles balísticos de la Agencia de Defensa contra Misiles. Minutos después, el SM-3 interceptó un misil balístico lanzado desde el Campo de Pruebas de Misiles del Pacífico en Barking Sands, Hawái. Fue la primera vez que este tipo de pruebas se realizó desde un destructor estadounidense equipado con un sistema de defensa contra misiles balísticos. El sistema de combate Aegis tiene un modo automático que permite a las computadoras de control de tiro identificar e interceptar blancos sin la intervención humana. (Foto: Armada de EUA)

incluso con un humano controlando sus acciones, los robots de combate que se encuentren en medio de un ataque electrónico o cibernético de fuerzas enemigas (o propias) podrían quedarse aislados de sus controladores humanos<sup>13</sup>. Las Fuerzas Armadas en todo el mundo han expandido sus capacidades de guerra electrónica en todos los escalones, motivadas principalmente por las comunicaciones inalámbricas y las amenazas que presentan los dispositivos explosivos y los drones. China y Rusia, en particular, han estado buscando maneras de cómo bloquear las comunicaciones tácticas de una fuerza enemiga, lo cual sería un problema potencial para unas fuerzas estadounidenses acostumbradas a un mando y control sin interrupciones<sup>14</sup>. Los robots de combate en el campo de batalla podrían quedar aislados fácilmente del control humano mediante un ataque electrónico o podrían ser modificados a través de medios cibernéticos.

Por otro lado, cuando los robots sufren daños, también podrían ser incapaces de recibir o procesar las entradas de datos de control de los humanos. Los robots de combate o los sistemas autónomos con defectos y separados de sus controladores podrían ser los nuevos elefantes de guerra del siglo XXI, sistemas de armas capaces de infligir enormes daños a fuerzas amigas o enemigas si no están fuertemente controlados. Aunque estén averiados o bajo un ataque electrónico, con la instalación de mecanismos de seguridad, los sistemas autónomos podrían ser forzados a cesar todas las actividades letales si pierden la comunicación con los controladores humanos, pero tal vez otras naciones no sigan estas mismas reglas de enfrentamiento. Si los robots de combate se vuelven esenciales para el éxito militar, negarles la autonomía para llevar a cabo acciones letales podría ser la diferencia entre ganar o perder la guerra.

Tener control total (*in the loop*) o parcial (*on the loop*) de los sistemas autónomos tal vez no reduzca drásticamente las violaciones contra los principios de guerra justa cometidas por los robots de combate porque los humanos tienden a depender del juicio de las máquinas cuando la información es limitada o cuando se encuentran en situaciones estresantes. Los controladores humanos de robots de combate, como los operadores de drones actuales, podrían tomar decisiones letales según la información que obtengan de sensores remotos e incluso de lecturas no visuales en un área con un ancho de banda limitado. Los robots de combate futuros tal vez hasta tengan la capacidad para explicarle de forma concisa a un operador cómo llegaron a una decisión, pero esto no eliminará necesariamente el riesgo de que los operadores dependan de las máquinas si no tienen experiencia de combate o con los sistemas<sup>15</sup>. Como el doctor John Hawley señala en el informe *Patriot Wars: Automation and the Patriot Air and Missile Defense System*, «un sistema automatizado en manos de una tripulación con un adiestramiento inadecuado es un sistema completamente automatizado *de facto*»<sup>16</sup>. Otros investigadores han notado esta tendencia de confiar en las máquinas durante emergencias. En un estudio de 2015 en el Instituto de Tecnología de Georgia, los participantes siguieron un guía robot a una sala de conferencias y observaron que el robot cometió varios errores de navegación durante la trayectoria. Después de llegar a la sala de conferencias, los investigadores activaron la alarma contra incendios del edificio. A pesar de ver que el robot cometió varios errores de navegación apenas unos minutos antes, todos los participantes en el estudio decidieron seguirlo para salir del supuesto edificio en llamas<sup>17</sup>.

Si bien Estados Unidos tiene interrogantes respecto al empleo de máquinas autónomas letales futuras, los sistemas de defensa antiaérea ya han entrado en un período de transición en el que las armas «inteligentes» tienen la capacidad y la autoridad para tomar decisiones letales. El sistema de defensa antiaérea Patriot, el sistema de combate Aegis y los sistemas de defensa de punto (CIWS) cuentan con modos automáticos que permiten que las computadoras de control de tiro accedan a sus sistemas de clasificación e interceptación de blancos<sup>18</sup>. Los tres sistemas ya han estado en combate y han clasificado e interceptado incorrectamente un blanco neutral o amigo, causando daños a aliados y a combatientes<sup>19</sup>. En muchas

de estas instancias, el fratricidio o los daños colaterales fueron producto de una mala comprensión humana de los algoritmos de los sistemas de defensa antiaérea o los operadores confiaron demasiado en el sistema de clasificación e interceptación de blancos de la computadora de control de tiro. Estos errores no se limitan solamente a los sistemas estadounidenses. El derribo en 2015 del vuelo 17 de Malaysia Airlines por un sistema de misiles Buk controlado por separatistas ucranianos es un ejemplo de los errores que pueden surgir cuando humanos y máquinas interactúan<sup>20</sup>.

Los sistemas de armas autónomos letales del futuro enfrentarán situaciones mucho más complejas que las que enfrentan los sistemas de defensa antiaérea automatizados actuales porque ellos se desplazarán por sí mismos y operarán en más ambientes (aéreo, terrestre, marítimo y espacial). Los sistemas de armas autónomos letales del futuro también tendrán que lidiar con un mayor número de amenazas y más importante aún, tendrán que interactuar con humanos en un mayor número de escenarios<sup>21</sup>. También es posible que los LAWS operen a grandes distancias de sus controladores y supervisores humanos. Esta movilidad permitirá a los robots de combate experimentar circunstancias imprevistas, lo cual aumentará la probabilidad de que enfrenten situaciones que no comprenderán del todo y en las que sus algoritmos internos no serán suficientes para procesar su entorno.

## Dificultades de los robots de combate para seguir los principios de guerra justa

En el campo de batalla, reconocer blancos potenciales y emplear fuerza letal contra ellos indiscriminadamente requiere muchos menos cálculos cognitivos que emplear la fuerza de forma justa. No importa si el combatiente es humano o robot. Los soldados humanos que siguen los principios de guerra justa de discriminación, proporcionalidad y necesidad militar tienen un amplio arsenal de experiencias y conocimientos culturales que los ayudan a evitar infligir sufrimiento innecesario en los no combatientes. Ellos inconscientemente también llegan a conclusiones heurísticas y pueden integrar de forma rápida información aparentemente sin relación en su toma de decisiones. Las máquinas, por lo general, tienen dificultades para crear sus propias reglas o relacionar datos dispares tan bien como los humanos y es posible

que nunca puedan llegar a tener esta capacidad<sup>22</sup>. Las siguientes secciones explican los tres principales principios de guerra justa y la razón por la cual los robots tal vez tengan dificultad en seguirlos.

**Discriminación.** Una vez iniciada la guerra, los soldados y combatientes no heridos pueden ser atacados en cualquier momento<sup>23</sup>. Cuando un individuo se alza en armas, él o ella pierde la impunidad civil que impide que sea atacado con fuerza letal. A medida que estos combatientes pierden sus derechos civiles, ganan otros a cambio, como la libertad de atacar a otros combatientes. Los combatientes solo recuperan su inmunidad contra ataques cuando no son capaces físicamente de hacer daño a otros (por estar heridos de gravedad) o cuando renuncian de forma permanente a su papel de combatiente. Si el período de servicio de un soldado finaliza y este decide regresar a su hogar, esa persona para de ser un blanco militar legítimo.

Si tienen que realizar tareas de identificación de blancos, los LAWS probablemente tendrán dificultades para diferenciar entre combatientes y no combatientes fuera de un ambiente de guerra convencional, como los humanos. En la multitud de conflictos que existen por todo el mundo (p. ej., Siria, Ucrania, Irak y Afganistán), gran parte de la población combatiente no forma parte de las estructuras militares convencionales. Uniformes o sistemas de armas similares no son suficientes para identificarlos como combatientes. En la actualidad, los programas de reconocimiento de objetos permiten que las máquinas detecten y clasifiquen los objetos comparándolos con fotos guardadas en una base de datos interna—los robots de combate pueden clasificar un fusil AK-47 o un tanque T-72 y concluir que el individuo asociado con ellos es una amenaza.

Sin embargo, el programa no puede llegar a las mismas conclusiones si los combatientes están sujetando un arma atípica o un objeto de doble propósito—no hay mucha diferencia entre un insurgente que está usando un teléfono celular para detonar un artefacto explosivo improvisado y una persona que está enviando un mensaje de texto. Para percibir la diferencia es necesario la capacidad de procesamiento de la mente humana. Un soldado joven tendría que evaluar las intenciones del insurgente basándose en indicadores visuales y en un análisis de las actividades habituales de los civiles y los insurgentes en el área, y rápidamente recordar información de entrenamientos y experiencias anteriores. Todo esto ocurriría en cuestión de segundos. Los LAWS no harán esta tarea u otras que requieran los modelos heurísticos

complicados de las mentes biológicas. Los vehículos autónomos de Google a veces tienen dificultades para atravesar las intersecciones con señales de pare porque no pueden interpretar las intenciones de los conductores humanos a su alrededor<sup>24</sup>.

Estos microanálisis son importantes. Si un LAWS encuentra a un combatiente herido, ¿cómo determinará si este está incapacitado? ¿Cómo puede una máquina procesar la rendición de soldados que portan armas sin infringir en sus derechos como prisioneros de guerra? Se pueden escribir algoritmos que proporcionen cierta capacidad para hacer esto, pero probablemente no igualarán la capacidad humana en el futuro próximo. Es posible que los robots de combate puedan identificar a los combatientes convencionales con facilidad, pero cuando se trate de no combatientes y soldados irregulares, tendrán problemas<sup>25</sup>.

**Proporcionalidad.** Además de emplear la fuerza discriminadamente, en la teoría de la guerra justa se señala que «la destrucción necesaria para lograr un objetivo militar debe ser proporcional a los beneficios que resulten de su consecución»<sup>26</sup>. En pocas palabras, el nivel de fuerza empleado debe ser proporcional al objetivo; no es necesario lanzar una bomba de 500 libras contra un insurgente con un AK-47 cuando un cartucho de 5.56 mm es suficiente. Si no se sigue este principio, es probable que los daños colaterales y las bajas de combatientes aumenten.

Los robots de combate tendrán dificultades a la hora de tomar decisiones proporcionales sin la intervención de un humano<sup>27</sup>. Los desarrolladores de programas pueden crear un LAWS que siga reglas de enfrentamiento preprogramadas que, por ejemplo, limiten el tipo de armas que pueden emplearse en zonas urbanas, zonas de alto el fuego, etc., y los comandantes pueden establecer límites geográficos e implementar medidas de control de apoyo de fuego para retardar el uso excesivo de la fuerza y el fratricidio por parte de los robots. Pero incluso con estas restricciones, los LAWS tendrán problemas si los controladores humanos no son capaces de predecir cuáles medidas de control serán necesarias. En combates impredecibles, esto será un gran desafío para los humanos.

Estimar los daños colaterales, un elemento esencial de la proporcionalidad, requiere múltiples predicciones y suposiciones. La mente humana hace que este proceso parezca fácil. Si un soldado de infantería es atacado desde un edificio en un ambiente urbano, este automáticamente analiza el propósito de esa estructura y

la probabilidad de que personas no combatientes estén presentes antes de contratar. Los humanos inconscientemente examinan las características del edificio, entre otros indicadores, antes de compararlas con la información sobre estructuras arquitectónicas que han aprendido a lo largo de su vida. La mente humana tiene una increíble capacidad para almacenar y recordar información que le permite recuperar datos de información aparentemente inconexos para aplicarlos a nuevos problemas, las computadoras no<sup>28</sup>.

Los juicios de proporcionalidad son más que identificación estructural. Los humanos pueden analizar fácilmente el efecto que la hora del día, el entorno cultural y los días de la semana tienen en los patrones de vida de la población civil. Es poco probable que el bombardeo de una iglesia un miércoles a la medianoche resulte en muchas muertes civiles, pero si el bombardeo ocurriera en la víspera de Navidad, el número de muertes podría aumentar exponencialmente. La capacidad de los LAWS para llegar a este tipo de conclusiones en el medio del combate es, en el mejor de los casos, una interrogante. Los programadores pueden escribir algoritmos que se aproximen a estos ciclos de decisión humanos, pero escribir un algoritmo para cada escenario posible es una tarea casi imposible. Las diferencias entre las capacidades sensoriales de los humanos y las de los robots también presentarán dificultades<sup>29</sup>. Además,

es posible que los robots de combate tácticos no posean la memoria suficiente para acomodar las bases de datos necesarias o la potencia para contrastarlas, o tenerlas no resulte rentable y haya que conectarlos a una nube.

**Necesidad militar.** En la guerra, el empleo de la violencia contra los blancos debe ser legítima y facilitar la derrota de una fuerza enemiga. Sin embargo, incluso matar a combatientes por motivos relacionados con un objetivo militar puede ser innecesario<sup>30</sup>. Este principio dificulta la labor de los combatientes tanto táctica como estratégicamente. El curso de la guerra es impredecible y a menudo es difícil evaluar si un ataque es necesario o no.

Los LAWS también tendrán problemas a la hora de decidir. Imagine que un LAWS localiza un terrorista que pretende estrellar un coche bomba contra una embajada estadounidense en los próximos días. El terrorista tiene al lado a su hijo de 8 años. ¿Es militarmente necesario atacarlo ahora? ¿O se debería esperar hasta que el terrorista esté solo, incluso si ello significa que el ataque con el coche bomba podría llevarse a cabo? Aunque la respuesta pueda parecer obvia, decisiones como estas no son fáciles. Un humano podría calcular la probabilidad de éxito del ataque del coche bomba frente a la probabilidad de que las fuerzas terrestres capturen al terrorista si el robot no ataca.



Un investigador de materiales examina datos experimentales en el planificador de inteligencia artificial del Sistema de Investigación Autónomo (ARES). El sistema, desarrollado por el Laboratorio de Investigación de la Fuerza Aérea, emplea la inteligencia artificial para diseñar, ejecutar y analizar experimentos a un ritmo más rápido que los métodos de investigación científica tradicionales. (Foto: Departamento de Defensa)

Un robot táctico probablemente no tendrá acceso a información externa o un conocimiento de la situación más amplio para formular un estimado.

En la actualidad, los soldados tienen dificultades para seguir al pie de la letra los principios de guerra justa. Los robots de combate con inteligencia subhumana probablemente tendrán más dificultades. Si esperan impedir que los LAWS violen las reglas de discriminación, proporcionalidad y necesidad militar, los controladores y los supervisores humanos necesitarán estar extremadamente familiarizados con los parámetros de selección e interceptación de blancos. A los soldados

se les requerirá hacerse responsables de las decisiones que los sistemas de armas automatizados tomen en el campo de batalla. Esto ya ocurre con los sistemas de defensa antiaérea y las municiones «dispara y olvida», pero dado el aumento de autonomía de los robots de combate, mayor será la responsabilidad.

Aunque no son el ejemplo perfecto, los perros de guerra son el ejemplo más claro para ilustrar esta situación. Cuando el entrenador suelta al perro en el campo de batalla, no está disparando un arma, está empleando un sistema de armas que toma sus propias decisiones con respecto a cuándo atacar. Para garantizar su efectividad en el combate y para impedir violaciones éticas y el fratricidio, el perro de combate pasa por un entrenamiento intenso de varios meses junto con su entrenador. Mediante el entrenamiento, el entrenador puede enseñarle al perro a realizar tareas simples y aprender sus puntos fuertes y débiles<sup>31</sup>. Los encargados de los robots de combate tal vez tengan que pasar por un entrenamiento similar en entornos físicos y virtuales, pero es posible que estas experiencias no sean suficientes para entender mejor el comportamiento de las máquinas. A pesar de las diferencias entre los humanos y los perros, los sentidos (p. ej., la vista, el oído y el olfato), los patrones de pensamiento y las emociones (p. ej., el miedo y el entusiasmo) tienen un punto de partida evolutivo común que permite a ambos entenderse. Un humano puede ver a un perro interactuar con su entorno y entender parcialmente su intención, sus motivaciones y sus pensamientos. Las marcadas diferencias entre la cognición y la percepción de los humanos y las máquinas dificultarán los esfuerzos humanos para establecer un nivel similar de comprensión.

## El proceso de toma de decisiones de los robots de combate

Los algoritmos, procesos definidos por programadores para que las computadoras resuelvan problemas, son la base de cualquier programa. Los desarrolladores de programas pueden combinar cientos o miles de algoritmos para crear un programa que tome decisiones sin que el usuario tenga conocimiento de ello. Los usuarios manipulan una interfaz gráfica de usuario que ejecuta *scripts*, archivos de órdenes de un lenguaje de programación de alto nivel, que luego son convertidos en datos binarios para el procesador. Estos niveles de abstracción no muestran el proceso de toma de decisiones de la máquina al usuario y los numerosos

cálculos heurísticos, las suposiciones y las fallas que los programadores incorporan voluntaria o involuntariamente en sus programas. Dado que los expertos en inteligencia artificial emplean cada vez más las técnicas de aprendizaje automático y aprendizaje profundo para permitir que las máquinas creen sus propios algoritmos, es posible que los sesgos se vuelvan más imperceptibles. Las técnicas de aprendizaje automático permiten que los algoritmos establezcan generalizaciones y analicen patrones mediante la evaluación de datos de entrenamiento; efectivamente, los algoritmos aprenden por sí mismos a través del ensayo y error. Como las máquinas dependen de los datos de entrenamiento, los científicos de computación inconscientemente pueden incorporar sesgos considerables dependiendo de los datos que utilicen<sup>32</sup>. Por ejemplo, un algoritmo de reconocimiento facial que use anuarios de secundarias de la región central de Estados Unidos como datos de entrenamiento tendrá problemas para identificar minorías étnicas porque estas no son muy comunes en la región. En un contexto comercial, este tipo de sesgo es vergonzoso, pero con los LAWS, puede ser mortal.

Imagine que está desarrollando un algoritmo que identifica varones armados en edad militar. Los programadores podrían lograr esto de muchas maneras. El programa podría emplear sensores para medir la altura del potencial blanco para confirmar que es mayor de edad, buscar indicios de vello facial o calcular la proporción entre la cintura y los hombros para determinar el género y estimar su masa corporal para compararla con los promedios de adultos y niños. Si la precisión de los sensores lo permite, el programa incluso podría medir los objetos que el blanco sujeta para determinar si estos coinciden con las especificaciones de armas que los robots de combate tienen guardadas en sus memorias.

Cada uno de estos cálculos requerirá la ejecución de algoritmos que comparen y apliquen los datos capturados por los sensores del robot de combate con bases de datos internas y algoritmos preestablecidos. El algoritmo para identificar varones armados en edad militar podría fallar de varias maneras: (1) El robot tal vez no puede llegar a una conclusión porque no es capaz de medir con precisión; sus sensores visuales tal vez no pueden identificar los objetos o medir las alturas después de cierta distancia y el algoritmo tal vez obliga a la máquina a tomar una



El soldado de primera clase Colten Carey, técnico de municiones guiadas de precisión del 23° Escuadrón de Mantenimiento, examina bombas con los kits de navegación JDAM el 11 de enero de 2018 en la base aérea Moody, en Georgia. El kit de navegación JDAM convierte las bombas estándares en municiones «inteligentes», capaces de interceptar de forma autónoma blancos designados. (Foto: Soldado Eugene Oliver, Fuerza Aérea de EUA)

decisión de cualquier manera. (2) La base de datos del robot no tiene información suficiente; por ejemplo, el blanco tal vez porta un modelo de arma que no se encuentra entre los datos de entrenamiento del algoritmo o tiene modificaciones que la máquina no puede reconocer. (3) Los cálculos heurísticos o las suposiciones del programador que dan lugar a la decisión tal vez tienen errores o son influenciados por suposiciones falsas. Si el algoritmo intentara diferenciar los hombres de las mujeres basándose en la altura y en la masa corporal, podría incurrir en errores si no tiene en consideración la nacionalidad y/o la etnia del blanco—la mujer escandinava promedio es más alta que el hombre chino promedio y probablemente pesa lo mismo<sup>33</sup>. Por otro lado, el algoritmo para identificar armas podría clasificar un fusil de caza como una herramienta y un AK-47 como un arma debido al criterio de un ingeniero de programas, aunque ambas son letales.

Casi todos los algoritmos presentan los sesgos y los criterios inherentes de las personas que los crean y de los datos de entrenamiento<sup>34</sup>. Los soldados encargados de monitorear y controlar los sistemas autónomos letales necesitarán estar familiarizados con estos sesgos y criterios del programa de los sistemas de armas para evitar accidentes y violaciones del derecho internacional de los conflictos armados.

Entender estos criterios y reconocer cuándo afectarán el rendimiento operacional puede ser una tarea bastante difícil. Los algoritmos que hacen funcionar a los LAWS serán extremadamente complejos. Los robots de combate tal vez dependan de millones de líneas de código para poder operar y cada línea incorporará los sesgos y las suposiciones de los ingenieros de programas. El volumen de código dificultará entender plenamente la complejidad de la toma de decisiones de los algoritmos y sus efectos en diversas situaciones de combate. El empleo de técnicas de aprendizaje automático empeorará potencialmente esta situación.

Es posible que los integrantes de las Fuerzas Armadas experimenten mayor dificultad para entender estos algoritmos por no tener los conocimientos técnicos que los ingenieros de programas que crearon los LAWS tienen. Los operadores de robots de combate tal vez ni siquiera tengan acceso a los algoritmos por motivos de seguridad; los Gobiernos podrían protegerlos como secretos nacionales o de propiedad intelectual<sup>35</sup>. Esta falta de claridad podría ser problemática. Los principios de guerra justa establecen que los combatientes y los comandantes son responsables del empleo ético de los sistemas autónomos letales, incluso cuando las decisiones de los funcionarios de adquisiciones y los ingenieros de programas tienen una repercusión igual o mayor en el comportamiento de los robots.

## Ampliar la responsabilidad de guerra justa

Dado que la mayoría de los soldados no entenderán plenamente el proceso de toma de decisiones letales de sus robots de combate, los ingenieros de programas y los funcionarios de adquisiciones responsables de colocar estos equipos en sus manos deben ser responsables parcialmente de las violaciones de guerra justa que los robots cometan en el campo de batalla. Para los sistemas jurídicos occidentales, la responsabilidad penal no puede ser atribuida a los acusados si estos son incapaces de entender su crimen y por eso, la sociedad no debería considerar a los combatientes responsables de las decisiones que los LAWS tomen en el campo de batalla si ellos no son capaces de entenderlas<sup>36</sup>. Para la sociedad, la responsabilidad del individuo es proporcional a su capacidad para influir en los acontecimientos. Aunque los combatientes siempre serán responsables de las violaciones de los principios de guerra justa que resulten de errores operativos, los diseñadores de sistemas y los funcionarios de adquisiciones también deberían ser responsables dependiendo de su papel con respecto a la violación<sup>37</sup>.

Los oficiales de adquisiciones tienen la responsabilidad ética de escribir los requisitos para los sistemas autónomos letales que reduzcan la probabilidad de que estos violen el derecho de los conflictos armados. Específicamente, los representantes de Gobierno necesitarán establecer estándares rigurosos para entrenar los algoritmos y evaluar los sistemas robóticos en

ambientes simulados y reales para descartar los errores más probables. Por su parte, los contratistas de defensa tienen la responsabilidad de proporcionar robots de combate que cumplan con estos requisitos al más alto nivel tecnológico posible e informar a los Gobiernos de vulnerabilidades identificadas en la toma de decisiones de los algoritmos y en la percepción de los sensores. Las compañías que fabrican vehículos autónomos son responsables de los accidentes causados por defectos de diseño<sup>38</sup>. Por lo tanto, las compañías que producen máquinas de combate con defectos negligentes deben estar sujetas acertadamente a sanciones civiles y penales. Por último, tanto líderes militares como civiles tienen la responsabilidad de decretar políticas y reglas de enfrentamiento que minimicen las oportunidades de los combatientes de emplear sistemas autónomos letales en situaciones que no comprenden.

Independientemente de cuán difícil sea, los líderes militares tienen que esforzarse para ayudar a los soldados a entender los riesgos de emplear sistemas autónomos letales en situaciones de combate predecibles<sup>39</sup>. Es posible que para los ejércitos del futuro, las ventajas económicas que representan los LAWS en términos de automatización disminuyan a medida que aumentan los costos de entrenamiento de los operadores y de las pruebas necesarias para mejorar los algoritmos de inteligencia artificial. La automatización de tareas realizadas por humanos en el campo de batalla probablemente resultará en la participación de menos combatientes con más conocimientos técnicos.

Los ejércitos deberían considerar crear clases especiales de soldados similares a la de los controladores aéreos tácticos conjuntos, quienes reciben un entrenamiento avanzado sobre la complejidad de los sistemas autónomos. Los comandantes podrían potencialmente limitar el control de los robots de combate a estos soldados, a menos que sea una situación de emergencia, hasta que la inteligencia artificial mejore aún más<sup>40</sup>. Ya existen varias metodologías y personal especializado para asesorar a los comandantes durante los ataques aéreos y los fuegos de largo alcance, como los especialistas en selección de blancos, los controladores aéreos tácticos conjuntos y los observadores avanzados. Los ejércitos que empleen los LAWS deberían considerar crear posiciones similares. Los expertos en robótica podrían acompañar a los comandantes en el terreno y ayudarlos a tomar decisiones difíciles con

respecto al empleo de robots de combate en el campo de batalla de la misma manera que los especialistas en selección de blancos ayudan a los comandantes a determinar el riesgo de daños colaterales durante los ataques aéreos. Es posible que los ejércitos también necesiten implementar pruebas de algoritmos y células de evaluación para cultivar datos de entrenamiento y crear simulaciones que permitan la calibración de los LAWS para teatros de combate y reglas de enfrentamiento específicos. Metodologías para estimar el riesgo de los LAWS en el campo de batalla también

podrían ayudar a los comandantes a establecer normas operativas sobre cuándo emplearlos.

En las próximas décadas, los combatientes entrarán en una era en la que sus sistemas de armas serán cada vez más responsables de la aplicación apropiada de fuerza contra blancos militares legítimos. Políticas y conceptos éticos necesitan avanzar paralelamente, distribuyendo las responsabilidades de guerra justa en proporción a los actores humanos involucrados para garantizar que los cambios tecnológicos no resulten en vacíos éticos. ■

## Notas

1. «A Digital Jet for the Modern Battlespace», Lockheed Martin (sitio web), consultado el 31 de enero de 2018, <https://www.f35.com/about/life-cycle/software>.

2. Stuart Armstrong y Kaj Sotala, «How We're Predicting AI—or Failing To», en *Beyond AI: Artificial Dreams*, eds. Jan Romportl y otros (Pilsen, República Checa: Universidad de Bohemia Occidental, 2012), págs. 52–75.

3. Noel E. Sharkey, «The Evitability of Autonomous Robot Warfare», *International Review of the Red Cross* 94, nro. 886 (2012): págs. 788–89.

4. En el Centro Nacional de Supercomputación de China en Wuxi se puede encontrar una computadora del tamaño de una habitación; para más información, consulte: <http://www.nsscwx.cn/wxcyw/>. IBM desarrolló una computadora cuántica que los investigadores pueden utilizar para avanzar aún más el potencial de tan vasta capacidad computacional. Cade Metz, «IBM Is Now Letting Anyone Play with its Quantum Computer», *Wired* (sitio web), 4 de mayo de 2016, consultado el 31 de enero de 2018, <https://www.wired.com/2016/05/ibm-letting-anyone-play-quantum-computer/>.

5. Personal de History.com, «Invention of the PC», History.com, 2011, consultado el 31 de enero de 2018, <http://www.history.com/topics/inventions/invention-of-the-pc>.

6. Patrick Lin, George Bekey y Keith Abney, «Autonomous Military Robotics: Risk, Ethics, and Design», Informe preparado para la Oficina de Investigación Naval, 20 de diciembre de 2008, consultado el 12 de febrero de 2018, [http://ethics.calpoly.edu/onr\\_report.pdf](http://ethics.calpoly.edu/onr_report.pdf).

7. Greg Allen y Taniel Chan, «Artificial Intelligence and National Security» (Cambridge, Massachusetts: Centro Belfer de Ciencias y Asuntos Internacionales, Escuela de Gobierno John F. Kennedy, julio de 2017), págs. 17–18, consultado el 31 de enero de 2018, <https://www.belfercenter.org/publication/artificial-intelligence-and-national-security>.

8. Directriz 3009.09 del Departamento de Defensa, *Autonomy in Weapon Systems* (Washington, DC: Oficina de Publicaciones del Gobierno (GPO), 21 de noviembre de 2012); Joseph Breecher, Heath Niemi y Andrew Hill, «My Droneski Just Ate Your Ethics», War on the Rocks (sitio web), 10 de agosto de 2016, consultado el 31 de enero de 2018, <https://warontherocks.com/2016/08/my-droneski-just-ate-your-ethics/>.

[my-droneski-just-ate-your-ethics/](https://warontherocks.com/2016/08/my-droneski-just-ate-your-ethics/).

9. Grupos como Campaign to Stop Killer Robots (Campaña para detener los robots asesinos) ya han comenzado a presionar para prohibir el empleo de sistemas de armas letales. Para más información sobre este grupo, véase: <https://www.stopkillerrobots.org/>.

10. C. Peter Chen, «Japan's Refusal of Washington Treaty, 19 Dec 1934», WWII Database (sitio web), consultado el 31 de enero de 2018, [https://ww2db.com/battle\\_spec.php?battle\\_id=45](https://ww2db.com/battle_spec.php?battle_id=45).

11. Wade Boese, «U.S. Withdraws from ABM Treaty; Global Response Muted», *Arms Control Today* (sitio web), julio/agosto de 2002, consultado el 31 de enero de 2018, [https://www.armscontrol.org/act/2002\\_07-08/abmjul\\_aug02](https://www.armscontrol.org/act/2002_07-08/abmjul_aug02).

12. Jules Hurst, «Intervention and the Looming Choices of Autonomous Warfighting», War on the Rocks (sitio web), 25 de agosto de 2016, consultado el 31 de enero de 2018, <https://warontherocks.com/2016/08/intervention-and-the-looming-choices-of-autonomous-warfighting/>.

13. *In the loop* se refiere a situaciones en las que los humanos tienen que dar la orden de aplicar fuerza letal a un robot. Los controladores humanos forman parte de la cadena de la muerte (*kill chain*). *On the loop* se refiere a situaciones en las que los humanos supervisan a los robots cuando estos aplican fuerza letal.

14. Thomas Gibbons-Neff, «“We Don't Have the Gear”: How the Pentagon is Struggling with Electronic Warfare», *Washington Post* (sitio web), 9 de febrero de 2016, consultado el 31 de enero de 2018, <https://www.washingtonpost.com/news/checkpoint/wp/2016/02/09/we-dont-have-the-gear-how-the-pentagon-is-struggling-with-electronic-warfare/>.

15. Dave Gershgorn, «We Don't Understand How AI Make Most Decisions, So Now Algorithms are Explaining Themselves», Quartz (sitio web), 20 de diciembre de 2016, consultado el 12 de febrero de 2018, <https://qz.com/865357/we-dont-understand-how-ai-make-most-decisions-so-now-algorithms-are-explaining-themselves/>.

16. John K. Hawley, *Patriot Wars: Automation and the Patriot Air and Missile Defense System*, serie Voices from the Field (Washington, DC: Center for a New American Security, enero de 2017), consultado el 31 de enero de 2018, <https://s3.amazonaws.com/>

[files.cnas.org/documents/CNAS-Report-EthicalAutonomy5-PatriotWars-FINAL.pdf](http://files.cnas.org/documents/CNAS-Report-EthicalAutonomy5-PatriotWars-FINAL.pdf).

17. Paul Robinette y otros, «Overtrust of Robots in Emergency Evacuation Scenarios» (conferencia, The 11th ACM/IEEE International Conference on Human Robot Interaction, Christchurch, Nueva Zelanda, 7 de marzo de 2016), pág. 101, consultado el 12 de febrero de 2018, <http://ieeexplore.ieee.org/document/7451740/>.

18. John Canning, «A Concept of Operations for Armed Autonomous Systems» (presentación en PowerPoint, 3ª Conferencia Anual de Tecnología Disruptiva, Washington, DC, 6–7 septiembre de 2010), consultado el 12 de febrero de 2018, [http://www.dtic.mil/ndia/2006/disruptive\\_tech/canning.pdf](http://www.dtic.mil/ndia/2006/disruptive_tech/canning.pdf).

19. Oficina del subsecretario de Defensa para Adquisiciones, Tecnología y Logística (OUSD[AT&L]), «Report of the Defense Science Board Task Force on Patriot System Performance Report Summary» (informe, Washington, DC: Departamento de Defensa, enero de 2005); George C. Wilson, «Navy Missile Downs Iranian Jetliner», *Washington Post* (sitio web), 4 de julio de 1988, consultado el 12 de febrero de 2018, <http://www.washingtonpost.com/wp-srv/inatl/longterm/flight801/stories/july88crash.htm>; A. J. Plunkett, «Iwo Jima Officer Killed In Firing Exercise», *Daily Press* (sitio web), 12 de octubre de 1989, consultado el 12 de febrero de 2018, [http://articles.dailypress.com/1989-10-12/news/8910120238\\_1\\_iwo-jima-ship-close-in-weapons-system](http://articles.dailypress.com/1989-10-12/news/8910120238_1_iwo-jima-ship-close-in-weapons-system).

20. Simon Tomlinson, «Russian Missile Killed Pilots and Cut Jet in Half but Passengers Could Have Been Conscious for up to a Minute as Plane Plunged, Reveals Official Report into MH17 Downed over Ukraine» *Daily Mail* (sitio web) 13 de octubre de 2015, consultado el 12 de febrero de 2018, <http://www.dailymail.co.uk/news/article-3270355/Doomed-flight-MH17-shot-Russian-BUK-missile-fired-rebel-held-territory-eastern-Ukraine-Dutch-investigators-set-rule.html>.

21. Andrew Ilachinski, «AI, Robots, and Swarms Issues, Questions, and Recommended Studies» (Arlington, Virginia: CNA, enero de 2017), págs. vi–vii, consultado el 12 de febrero de 2018, [https://www.cna.org/CNA\\_files/PDF/DRM-2017-U-014796-Final.pdf](https://www.cna.org/CNA_files/PDF/DRM-2017-U-014796-Final.pdf).

22. Sharkey, «The Evitability of Autonomous Robot Warfare», pág. 789.

23. Michael Walzer, *Just and Unjust Wars: A Moral Argument with Historical Illustrations*, 4ª ed. (1977; reimpr., Nueva York: Basic Books, 2006), pág. 138.

24. Matt Richtel y Conor Dougherty, «Google's Driverless Cars Run into Problem: Cars with Drivers», *New York Times* (sitio web), 1 de septiembre de 2015, 12 de febrero de 2018, <https://www.nytimes.com/2015/09/02/technology/personaltech/google-says-its-not-the-driverless-cars-fault-its-other-drivers.html>.

25. Sharkey, «The Evitability of Autonomous Robot Warfare», págs. 788–89.

26. Brian Orend, *The Morality of War* (Peterborough, Ontario: Broadview Press, 2006), pág. 119.

27. Sharkey, «The Evitability of Autonomous Robot Warfare», pág. 789.

28. Jeneen Interlandi, «New Estimate Boosts the Human Brain's Memory Capacity 10-Fold», *Scientific American* (sitio web), 5 de febrero de 2016, consultado el 12 de febrero de 2018, <https://www.scientificamerican.com/article/new-estimate-boosts-the-human-brain-s-memory-capacity-10-fold/>.

29. Ruth A. David y Paul Nielsen, «Report of the Defense Science Board Summer Study on Autonomy» (Washington, DC: OUSD[AT&L], junio de 2016), pág. 14, consultado el 12 de febrero de 2018, <http://www.dtic.mil/dtic/tr/fulltext/u2/1017790.pdf>.

30. Orend, *The Morality of War*, pág. 119.

31. Department of the Army Pamphlet 190-12, *Military Working Dog Program* (Washington, DC: U.S. GPO, 30 de septiembre de 1993), pág. 24.

32. Aylin Caliskan, Joanna J. Bryson y Arvind Narayanan, «Semantics Derived Automatically from Language Corpora Contain Human-like Biases», *Science* 356, no. 6334 (2017): págs. 183-86.

33. Ian Langtree, «Height Chart of Men and Women in Different Countries», Disabled World (sitio web), última actualización el 19 de diciembre de 2017, consultado el 12 de febrero de 2018, <https://www.disabled-world.com/artman/publish/height-chart.shtml>.

34. Felicitas Kramer, Kees van Overveld y Martin Peterson, «Is There an Ethics of Algorithms?», *Ethics and Information Technology* 13, nro. 3 (2010): pág. 251.

35. Tom Simonite, «For Superpowers, Artificial Intelligence Fuels New Global Arms Race», *Wired* (sitio web), 8 de septiembre de 2017, consultado el 12 de febrero de 2018, <https://www.wired.com/story/for-superpowers-artificial-intelligence-fuels-new-global-arms-race/>.

36. «The Infancy Defense: Criminal Law Basics», The 'Lectric Law Library, consultado el 12 de febrero de 2018, <http://www.lectlaw.com/mjl/cl032.htm>.

37. Aaron M. Johnson y Sidney Axinn, «The Morality of Autonomous Robots», *Journal of Military Ethics* 12, nro. 2 (2013): pág. 131.

38. Anjali Singhvi y Karl Russell, «Inside the Self-Driving Tesla Fatal Accident», *New York Times* (sitio web), 12 de julio de 2016, consultado el 12 de febrero de 2018, <https://www.nytimes.com/interactive/2016/07/01/business/inside-tesla-accident.html>.

39. Esto es más fácil decirlo que hacerlo. Para poder responsabilizar a los comandantes por la falta de preparación de sus soldados para un escenario de combate específico, la preparación tendría que ser parte de un currículo de entrenamiento estandarizado.

40. David y Nielsen, «Report of the Defense Science Board», pág. 38.