

La guerra de Rusia en Ucrania demuestra por qué el mundo debe prohibir ya las armas de Inteligencia Artificial

Por: STUART RUSSELL. 03/03/2023

Los robots que matan sin supervisión humana ya no son ciencia ficción, escribe el informático Stuart Russell. Rusia ya utiliza misiles de crucero semiautónomos contra Ucrania. Y en 2020, el gobierno provisional libio utilizó drones cargados de explosivos para dar caza de forma autónoma a las fuerzas en retirada. Y lo peor podría estar por llegar: una persona podría lanzar enjambres casi ilimitados de microdrones baratos – “slaughterbots”- para acabar con oponentes políticos o poblaciones enteras. Los gobiernos y las organizaciones deben dejar de lado excusas endebles y prohibir ya las armas de inteligencia artificial para evitar un “futuro sombrío de amenazas robóticas”, afirma.

Un año después de la invasión rusa, una carrera armamentística de inteligencia artificial (IA) se desarrolla en suelo ucraniano. El público occidental aplaude cuando las valientes fuerzas ucranianas utilizan cuadricópteros comerciales modificados para lanzar granadas sobre los soldados rusos. Abuchean cuando las brutales fuerzas rusas envían enjambres de misiles de crucero iraníes baratos para destruir hospitales, centrales eléctricas y bloques de apartamentos. Pero esta simple narrativa de “nosotros contra ellos” oculta una tendencia inquietante: las armas son cada vez más inteligentes.

Pronto, los sistemas de armas letales totalmente autónomos podrían convertirse en algo habitual en los conflictos. Algunos ya están en el mercado. Afortunadamente, pocos se han utilizado realmente en la guerra, y ninguno se ha utilizado en Ucrania en el momento de escribir estas líneas. Sin embargo, la evolución de los acontecimientos es motivo de preocupación.

La lógica inevitable de utilizar contramedidas electrónicas contra las armas operadas por control remoto está llevando a ambas partes a aumentar el nivel de autonomía de esas armas. Esto nos acerca cada vez más a un mundo peligroso en el que los sistemas de armas autónomas letales son herramientas baratas y ampliamente disponibles para infligir bajas masivas: armas de destrucción masiva que se encuentran en todos los supermercados de armas, a la venta de cualquier dictador,

señor de la guerra o terrorista.

Aunque es difícil debatir la prohibición de armas que puedan ayudar a la causa ucraniana, ahora es urgente que los gobiernos del mundo lo hagan y limiten el uso de la IA en la guerra. Nadie quiere este futuro sombrío de amenazas robóticas.

Para empezar, los gobiernos deben iniciar negociaciones serias sobre un tratado que prohíba las armas autónomas antipersona, como mínimo. Las sociedades profesionales de la inteligencia artificial y la robótica deberían desarrollar y aplicar códigos de conducta que prohíban trabajar en armas autónomas letales. Y la gente de todo el mundo debería entender que permitir que los algoritmos decidan matar a seres humanos es una idea terrible.

Presiones que conducen a la plena autonomía

¿Qué son exactamente los “sistemas de armas autónomas letales”? Según las Naciones Unidas, son “armas que localizan, seleccionan y atacan objetivos humanos sin supervisión humana”. La palabra “atacar” en esta definición es un eufemismo de “matar”. No me refiero a armas manejadas a distancia por humanos, como el avión no tripulado estadounidense Predator o los lanzagranadas caseros ucranianos, porque no son autónomos. Tampoco me refiero a los sistemas de defensa antimisiles, ni a los drones totalmente autónomos que tanto rusos como ucranianos utilizan para reconocimiento, que no son letales. Y no estoy hablando de los robots de ciencia ficción que aparecen en las películas de “Terminator” - controlados por la espeluznante conciencia emergente del sistema de software Skynet e impulsados por el odio a la humanidad- que los medios de comunicación suelen evocar cuando hablan de armas autónomas. En este caso no se trata de máquinas sin escrúpulos que se apoderan del mundo, sino de armas desplegadas por humanos que reducirán drásticamente nuestra seguridad física.

Los sistemas de IA actuales presentan todas las capacidades necesarias: planificar misiones, navegar, cartografiar en 3D, reconocer objetivos, volar por ciudades y edificios y coordinar ataques. Hay muchas plataformas disponibles. Entre ellas: cuadricópteros de centímetros a metros; aeronaves de ala fija (desde aviones de reparto de paquetes del tamaño de un hobby y drones portamisiles de tamaño completo hasta cazas supersónicos “listos para la autonomía”, como el Taranis de BAE Systems); camiones y tanques autodirigidos; lanchas rápidas, destructores y submarinos autónomos; e incluso robots humanoides esqueléticos.

El camino hacia la plena autonomía en el conflicto Rusia-Ucrania comienza con varios tipos de armas semiautónomas ya en uso. Por ejemplo, Rusia está desplegando misiles de crucero “inteligentes” de gran efecto, que alcanzan objetivos predefinidos como edificios administrativos e instalaciones energéticas. Estas armas incluyen misiles Shahed iraníes, apodados “ciclomotores” por los ucranianos debido a su sonido, que pueden volar bajo a lo largo de los ríos para evitar ser detectados y rodear una zona mientras esperan instrucciones. La clave de estos ataques es el uso de enjambres de misiles para desbordar los sistemas de defensa antiaérea, junto con enlaces de radio mínimos para evitar ser detectados. He oído informes de que los nuevos Shaheds están siendo equipados con detectores de infrarrojos que les permiten localizar fuentes de calor cercanas sin necesidad de actualizaciones de objetivos comunicadas por radio desde los controladores; de ser cierto, esto supondría un paso importante hacia la plena autonomía.

Los ucranianos han desplegado armas turcas teleoperadas Bayraktar contra tanques y otros objetivos desde los primeros días de la guerra. La mejora de las defensas aéreas rusas y las interferencias han hecho que estas armas sean más vulnerables y menos eficaces con el paso del tiempo; además, cuestan alrededor de 5 millones de dólares cada una (250 veces más caras que los Shaheds). Los cuadricópteros comerciales teledirigidos adaptados para lanzar granadas han demostrado su eficacia en operaciones tácticas a pequeña escala, y se han utilizado lanchas pilotadas a distancia para atacar objetivos navales. Pero, a medida que los sistemas de interferencia se convierten en la norma, la teleoperación se hace más difícil y las armas autónomas cada vez más atractivas.

En otros lugares, las armas autónomas letales llevan varios años a la venta. Por ejemplo, desde 2017, un fabricante de propiedad estatal en Turquía (STM) vende el dron Kargu, que tiene el tamaño de un plato de comida y transporta 1 kilogramo de

explosivo. Según el sitio web de la empresa en 2019 (desde entonces matizado), el dron es capaz de realizar impactos “autónomos y precisos” contra vehículos y personas, con “objetivos seleccionados en imágenes” y mediante el “seguimiento de objetivos en movimiento” (ver go.nature.com/3ktq6bb). Según ha informado la ONU, los drones Kargu fueron utilizados en 2020 por el Gobierno de Acuerdo Nacional libio -a pesar de un estricto embargo de armas- para “cazar” de forma autónoma a las fuerzas en retirada 1/.

Otras formas de misiles “merodeadores”, como el Shahed, también muestran una forma de autonomía. El dron israelí Harpy puede sobrevolar una región durante varias horas en busca de objetivos que coincidan con una firma visual o de radar y destruirlos después con su carga explosiva de 23 kilogramos. (El misil ruso Lancet, muy utilizado en Ucrania, tiene características similares). Mientras que el Kargu y el Harpy son armas “kamikaze”, el Ziyun Blowfish A3 chino es un helicóptero autónomo equipado con una ametralladora y varias bombas de gravedad no guiadas. Todos estos sistemas se describen tanto en modo autónomo como teledirigido, lo que hace difícil saber si un ataque determinado ha sido llevado a cabo por un operador humano.

Ventajas y problemas

¿Por qué los ejércitos buscan máquinas que puedan decidir por sí mismas si matan a seres humanos? Al igual que las armas teledirigidas, los aviones, tanques y submarinos autónomos pueden llevar a cabo misiones que serían suicidas para las personas. Son más baratos, más rápidos, más maniobrables y tienen mayor alcance que sus homólogos tripulados; pueden soportar fuerzas g más elevadas en vuelo y funcionar bajo el agua sin sistemas de soporte vital. Pero, a diferencia de las armas teledirigidas, las armas autónomas pueden funcionar incluso cuando la comunicación electrónica es imposible debido a interferencias, y pueden reaccionar incluso más rápido que cualquier arma teledirigida por un ser humano. El experto en IA Kai-Fu Lee, entre otros, ha descrito las armas autónomas como la “tercera revolución de la guerra” después de la pólvora y las armas nucleares 2/.

Un argumento habitual a favor es que hacer la guerra con armas autónomas protegerá vidas militares, como se dice que hacen las armas manejadas a distancia y los misiles de crucero. Pero esto es una falacia. El otro bando también dispondría de esas armas y, como hemos visto en Ucrania, el número de muertos entre soldados y civiles es asombroso.

Otro argumento que se suele esgrimir es que, en comparación con otros modos de guerra, la capacidad de las armas autónomas letales para distinguir a los civiles de los combatientes podría reducir los daños colaterales. Estados Unidos, junto con Rusia, ha estado citando este supuesto beneficio con el efecto de bloquear las negociaciones multilaterales en la Convención sobre Ciertas Armas Convencionales (CCAC) en Ginebra, Suiza – conversaciones que se han producido esporádicamente desde 2014.

El caso se basa en dos afirmaciones. En primer lugar, que los sistemas de IA son menos propensos a cometer errores que los humanos, una proposición dudosa ahora, aunque podría llegar a ser cierta. Y segundo, que las armas autónomas se utilizarán esencialmente en los mismos escenarios que las armas controladas por humanos, como rifles, tanques y drones Predator. Esto parece inequívocamente falso. Si las armas autónomas se utilizan más a menudo, por diferentes partes con objetivos diversos y en escenarios menos claros, como insurrecciones, represión, guerras civiles y terrorismo, entonces cualquier supuesta ventaja en distinguir a civiles de soldados es irrelevante. Por este motivo, creo que el énfasis en la supuesta superioridad de las armas a la hora de distinguir a civiles de combatientes, que tiene su origen en un informe de la ONU de 2013 ^{3/} en el que se señalan los riesgos de una identificación errónea, ha sido erróneo.

Hay muchas más razones por las que desarrollar armas autónomas letales es una mala idea. La mayor, como escribí en *Nature* en 2015 ^{4/}, es que “cabe esperar plataformas desplegadas por millones, cuya agilidad y letalidad dejarán a los humanos totalmente indefensos”. El razonamiento se ilustra en un vídeo de YouTube de 2017 que aboga por el control de armas, que publiqué con el *Future of Life Institute* (véase go.nature.com/4ju4zj2). Muestra “Slaughterbots”, enjambres de microdrones baratos que utilizan IA y reconocimiento facial para asesinar a opositores políticos. Al no necesitar supervisión humana, una sola persona puede lanzar un número casi ilimitado de armas para acabar con poblaciones enteras. Los expertos en armamento coinciden en que los enjambres antipersona deberían

clasificarse como armas de destrucción masiva (véase go.nature.com/3yqjx9h). Por este motivo, la comunidad de la IA se opone casi unánimemente a las armas autónomas.

Además, los sistemas de IA podrían ser pirateados, o los accidentes podrían intensificar los conflictos o reducir el umbral de las guerras. Y la vida humana se devaluaría si los robots tomaran decisiones de vida o muerte, planteando preocupaciones morales y de justicia. En marzo de 2019, el secretario general de la ONU, António Guterres, resumió este caso ante los negociadores de armas autónomas en Ginebra: “Las máquinas con el poder y la discreción de quitar vidas sin participación humana son políticamente inaceptables, moralmente repugnantes y deberían estar prohibidas por el derecho internacional” (ver go.nature.com/3yn6pqt). Sin embargo, todavía no existen normas, más allá del derecho internacional humanitario, contra la fabricación y venta de armas autónomas letales de destrucción masiva.

La acción política, estancada

Por desgracia, la política no ha seguido el ritmo de los avances tecnológicos. Decenas de organizaciones de derechos humanos y de control de armas se han unido a la [Campaña para Detener a los Robots Asesinos](#), que pide la prohibición de las armas autónomas letales. Los políticos y los gobiernos no han actuado, a pesar de que las encuestas sugieren un amplio apoyo público a dicha prohibición (más del 60% de los adultos; véase, por ejemplo, go.nature.com/416myef). Miles de investigadores y líderes en IA, entre los que me incluyo, se han unido a esos llamamientos (véase go.nature.com/4gqmf5), pero, hasta ahora, ninguna sociedad académica ha desarrollado una política sobre armas autónomas debido a la preocupación por debatir asuntos que no son puramente científicos.

Una de las razones por las que las negociaciones en el marco de la CCAC han avanzado poco es la confusión, real o fingida, sobre cuestiones técnicas. Los países siguen discutiendo sin cesar sobre el significado de la palabra “autónomo”. De forma absurda, por ejemplo, Alemania declaró que un arma es autónoma sólo si tiene “la capacidad de aprender y desarrollar conciencia de sí misma”. China, que apoya ostensiblemente la prohibición de las armas autónomas, afirma que en cuanto las armas sean capaces de distinguir de forma autónoma entre civiles y soldados, ya no contarán como autónomas y, por tanto, no se prohibirían. El Reino Unido se ha comprometido a no desarrollar ni utilizar nunca armas autónomas letales, pero no

deja de redefinirlas para que su compromiso carezca de sentido. Por ejemplo, en 2011, el Ministerio de Defensa del Reino Unido escribió que “un grado de funcionamiento autónomo es probablemente alcanzable ahora”, pero en 2017 declaró que “un sistema autónomo es capaz de comprender la intención de alto nivel”. Michael Fallon, entonces secretario de Estado de Defensa, escribió en 2016 que “los sistemas totalmente autónomos aún no existen y no es probable que lo hagan durante muchos años, si es que existen”, y concluyó que “es demasiado pronto para prohibir algo que simplemente no podemos definir” (véase go.nature.com/3xrzt6).

Es poco probable que se produzcan nuevos avances en Ginebra. Estados Unidos y Rusia se niegan a permitir negociaciones sobre un acuerdo jurídicamente vinculante. A Estados Unidos le preocupa que un tratado no sea verificable, lo que llevaría a otras partes a eludir una prohibición y crearía un riesgo de sorpresa estratégica. Rusia objeta ahora que se la está discriminando, debido a su invasión de Ucrania.

Un camino pragmático

En lugar de bloquear las negociaciones, sería mejor que Estados Unidos y otros países se centraran en idear medidas prácticas para fomentar la confianza en la adhesión. Éstas podrían incluir acuerdos de inspección, limitaciones de diseño que disuadan de la conversión a la plena autonomía, normas que obliguen a los proveedores industriales a comprobar la buena fe de los clientes, etcétera. Tendría sentido debatir las competencias de una versión para la IA de la Organización para la Prohibición de las Armas Químicas, que ha ideado medidas técnicas similares para aplicar la [Convención sobre Armas Químicas](#). Estas medidas no han sobrecargado a la industria química ni han restringido la investigación química. Del mismo modo, el nuevo tratado START entre Estados Unidos y Rusia permite 18 inspecciones in situ de instalaciones de armas nucleares al año. Y el Tratado de Prohibición Completa de los Ensayos Nucleares podría no haber existido nunca si los científicos de todas las partes no hubieran trabajado juntos para desarrollar el Sistema Internacional de Vigilancia que detecta las violaciones clandestinas.

A pesar del punto muerto en Ginebra, hay destellos de esperanza. De los países que se han pronunciado, la gran mayoría está a favor de la prohibición. Las negociaciones podrían avanzar en la Asamblea General de la ONU en Nueva York, donde ningún país tiene derecho a veto, y en reuniones a nivel ministerial. La semana pasada, el gobierno de los Países Bajos organizó una reunión en La Haya

sobre “la IA responsable en el ámbito militar”, en la que se planteó la cuestión de si es ético introducir este tipo de armas. Durante la reunión, Estados Unidos anunció una “declaración política” de principios y mejores prácticas para el uso militar de la IA e instó a otras naciones a suscribirlos (véase go.nature.com/3xsj779). Quizá lo más importante sea la afirmación de que “Los Estados deben mantener el control y la participación humanos en todas las acciones críticas para informar y ejecutar decisiones soberanas relativas al empleo de armas nucleares”. Ya se han adherido a la declaración más de 60 países, incluida China. Por desgracia, no es vinculante y no descarta ninguna categoría de arma autónoma.

Los días 23 y 24 de febrero, Costa Rica acogerá una reunión de naciones latinoamericanas y caribeñas sobre el “impacto social y humanitario de las armas autónomas”, que incluye las amenazas de actores no estatales que podrían utilizarlas indiscriminadamente. Estas mismas naciones organizaron la primera zona libre de armas nucleares, lo que hace albergar esperanzas de que también puedan iniciar un tratado que declare una zona libre de armas autónomas.

Próximos pasos

En mi opinión -y sospecho que en la de la mayoría de los habitantes de la Tierra- la mejor solución es simplemente prohibir las armas autónomas letales, quizá a través de un proceso iniciado por la Asamblea General de la ONU. Otra posibilidad, sugerida como medida de compromiso por un grupo de expertos (véase go.nature.com/3jugzxy) y propuesta formalmente a la comunidad internacional por el Comité Internacional de la Cruz Roja (véase go.nature.com/3k3tpan), sería prohibir las armas autónomas antipersona. Al igual que la Declaración de San Petersburgo de 1868, que prohibía la explosión de artefactos de menos de 400 gramos, un tratado de este tipo podría poner límites más bajos al tamaño y la carga útil de las armas, imposibilitando el despliegue de enormes enjambres de pequeños artefactos que funcionen como armas de destrucción masiva.

Ética de la inteligencia artificial

En lugar de bloquear los avances en Ginebra, los países deberían colaborar con la comunidad científica para desarrollar las medidas técnicas y jurídicas que podrían hacer verificable y aplicable una prohibición de las armas autónomas. Entre las cuestiones técnicas figuran las siguientes ¿Qué parámetros físicos deberían utilizarse para definir el límite inferior de las armas permitidas? ¿Qué son las

plataformas “precursoras” (que pueden ampliarse hasta alcanzar la plena autonomía) y cómo debe gestionarse su producción y venta? ¿Deberían utilizarse restricciones de diseño, como exigir una señal de “retirada”? ¿Pueden separarse físicamente los circuitos de disparo de la informática de a bordo, para evitar que las armas pilotadas por humanos se conviertan fácilmente en armas autónomas? ¿Pueden diseñarse protocolos verificables para evitar la escalada accidental de hostilidades entre sistemas autónomos?

En el ámbito civil, las sociedades profesionales de inteligencia artificial y robótica (como la Association for the Advancement of Artificial Intelligence, la Association for Computing Machinery y el Institute of Electrical and Electronics Engineers) deberían elaborar y aplicar códigos de conducta que prohíban trabajar en armas autónomas letales. Hay muchos precedentes: por ejemplo, la American Chemical Society tiene una estricta política sobre armas químicas (véase go.nature.com/3yn8ajt) y la American Physical Society pide a Estados Unidos que ratifique el Tratado de Prohibición Completa de los Ensayos Nucleares (véase go.nature.com/3jrajvr), se opone al uso de armas nucleares contra Estados no nucleares (véase go.nature.com/3k4akq8) y aboga por sólidos programas de investigación en ciencia y tecnología de verificación en beneficio de la paz y la seguridad (véase go.nature.com/3hzjkkv).

Mientras se desarrolla la guerra de Rusia en Ucrania y la tecnología de las armas autónomas avanza a toda velocidad (junto con el deseo de utilizarla), el mundo no puede permitirse otra década de posturas diplomáticas y confusión. Los gobiernos tienen que cumplir lo que parece una simple petición: dar a sus ciudadanos algún tipo de protección contra la caza y la muerte por robots.

Notas

- 1/ United Nations. Final Report of the Panel of Experts on Libya Established Pursuant to Security Council Resolution 1973 (2011) (UN Security Council, 2021).
- 2/ Lee, K.-F. ‘The third revolution in warfare’ The Atlantic (11 September 2021).
- 3/ UN General Assembly. Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns (United Nations, 2013).

4/ Russell, S. Nature 521, 415–418 (2015).

21/2/2023

Traducción: **viento sur**

[LEER EL ARTÍCULO ORIGINAL PULSANDO AQUÍ](#)

Fotografía: Viento sur. *Misil Shahed 131 recuperado en Ucrania. Foto: FA de Ucrania/ Wikimedia*

Fecha de creación

2023/03/03