

Más que robots asesinos. Inteligencia artificial en la toma de decisiones en la guerra

Ángel Gómez de Ágreda¹

Resumen

La aplicación de la Inteligencia Artificial (IA) al ámbito militar suele vincularse con los drones autónomos o robots asesinos. Su carácter letal les da una mayor visibilidad que ha propiciado intentos de regulación ética y jurídica. Sin embargo, la verdadera clave se encuentra en la autonomía en la toma de decisiones de los comandantes en una actividad que es eminentemente política y social: la guerra. Sistemas de selección de blancos o de targeting deben contar con una supervisión significativa por parte de sus operadores. Sin embargo, las presiones de tiempo y la cada vez mayor transparencia de las interfaces humano-máquina (HMI) lo hacen cada vez más difícil.

La capacidad de la IA para generar realidades sintéticas resulta en un riesgo añadido para la toma de decisiones, tanto por la acción del adversario como a través de las propias acciones.

Unos criterios universales sobre los límites de la preservación de la agencia humana en el ámbito militar podrían trasladarse más adelante al entorno civil.

1 Coronel del Ejército del Aire y del Espacio de España en situación de reserva. Doctor en Ingeniería de Organización Industrial por la Universidad Politécnica de Madrid. Máster en Terrorismo, Máster en Política Europea de Seguridad Común. Diplomado de Estado Mayor. Ha sido profesor de Estrategia y Relaciones Internacionales en el Centro Superior de Estudios de la Defensa (CESEDEN) de España, jefe de Cooperación en el Mando Conjunto de Ciberdefensa, jefe del Área de Análisis Geopolítico en el Ministerio de Defensa y Agregado de Defensa en las embajadas de España en Japón y en la República de Corea. Piloto militar y paracaidista, se desempeña también como profesor universitario, conferenciante. Es autor de numerosas publicaciones, entre las que destaca su libro "Mundo Orwell", que obtuvo el Premio Knowsquare al Mejor Libro de Empresa 2019.



Palabras clave

Sistemas autónomos letales
Toma de decisión
Ética
Realidad sintética
Interfaces hombre-máquina

Keywords

Lethal autonomous systems
Decision-making
Ethics
Synthetic reality
Human-machine interfaces



Abstract

The application of artificial intelligence (AI) in the military domain is often associated with autonomous drones or killer robots. Their lethal nature grants them greater visibility, which has led to attempts at ethical and legal regulation. However, the true key lies in the autonomy of decision-making by commanders in an activity that is inherently political and social: war. Target selection systems must have significant oversight by their operators. Nonetheless, time pressures and the increasing transparency of human-machine interfaces (HMI) make this increasingly difficult.

The ability of AI to generate synthetic realities introduces an additional risk to decision-making, both due to the actions of adversaries and through one's own actions.

Universal criteria regarding the limits of preserving human agency in the military sphere could later be transferred to the civilian environment.

Introducción

Cuando se habla de Inteligencia Artificial (IA) en el ámbito de la defensa, se piensa indefectiblemente en robots asesinos.² Los Sistemas de Armas Autónomos Letales (SALAS) cautivan la imaginación en base al relato procedente de las películas de Hollywood.

No obstante, a pesar de ser el empleo más espectacular de la IA, su vistosidad y el morbo asociado a la misma reside en su letalidad, no en la autonomía. La imagen icónica de Arnold Scharzenegger con sus irrompibles gafas de sol prestadas desvía la atención de lo verdaderamente relevante, la toma de conciencia de sí mismo del sistema Skynet que lo crea.

Es más, dejando de un lado inciertas prospectivas de un futuro más o menos lejano, la ética asociada a las máquinas también aparta el foco de la ética de los humanos asociados a esas máquinas y la relación entre ambos.

A cualquier soldado le resulta evidente la relación inequívoca entre guerra y letalidad. No es la letalidad, por lo tanto, lo que supone un carácter distintivo de los sistemas autónomos letales, sino el grado de autonomía que permanece en el ser humano (incluso más que el nivel de autonomía que se confiere a las máquinas).

Lo relevante es, en suma, el grado de control humano sobre las decisiones que se adoptan.³ Para ello resulta evidente que se precisa también un alto nivel de conocimiento del sistema informático con el que se está trabajando. Conocer para controlar podría ser la premisa básica del uso de la IA en la guerra.

No es suficiente que el humano se reserve la responsabilidad de ejecutar una acción o de apretar un gatillo. La decisión que toma tiene que venir precedida de una suficiente comprensión del entorno.

2 STOP Killer Robots, 2018.

3 BOULANIN, Vincet., DAVISON, Neil, GOUSSAC, Netta, PELDÁN, Moa, & Sipri, C. 2020. Limits on autonomy in Identifying Practical Elements of Human Control



La IA es una herramienta extraordinaria para mejorar esa conciencia situacional. Pero también tiene la capacidad para crear mundos ficticios perfectamente verosímiles que alimenten con sensaciones erróneas a su socio humano.

Sistemas de Armas Autónomos Letales

Existen numerosos equívocos relativos a los Sistemas de Armas Autónomos Letales (SALAS, o LAWS, en inglés).

La proliferación de drones y aparatos pilotados de forma remota tiende a confundirse con las plataformas dotadas de IA y capacidad de aprendizaje o decisión propia.

Esta confusión se ve tanto más agravada cuanto los grados de autonomía y automatismo de los sistemas de armas varían en incrementos diferenciales que hacen muy difícil establecer dónde está realmente la frontera entre unos y otros.⁴

Para empezar, cabe distinguir entre sistemas pilotados de forma remota, aquellos con sistemas de guía o de disparo automatizados y aquellos dotados de algoritmos de IA capaces de actuar de forma independiente de su operador y variando determinados parámetros de su programación.

De hecho, la regulación de los SALAS establece un paralelismo remoto con la aplicable a las minas antipersonal. Se trata de dispositivos que, una vez instalados, actúan de forma independiente

del operador. Si bien –en la mayor parte de los casos– no puede hablarse de sistemas autónomos, sino automáticos (responden siempre igual ante determinados estímulos), de los problemas que ha generado su uso se pueden extraer enseñanzas muy útiles para los SALAS.

Se habla, en este caso, de la necesidad de establecer límites geográficos, temporales u operativos a la actuación de estos sistemas.⁵ Es decir, hay que evitar que campos de minas (en este caso, que sistemas de armas autónomos) puedan seguir activos y operando más allá del plazo temporal del conflicto o de la función que debían cumplir.

Las limitaciones en cuanto al alcance o a la duración de estos sistemas deben estar presentes en las fases de diseño y desarrollo de forma necesaria. De otro modo, serían susceptibles de contravenir los preceptos más básicos del Derecho Internacional Humanitario.⁶

No solo la legitimidad de estos sistemas quedaría en entredicho, sino también la dignidad humana de los combatientes o de los civiles afectados por estos sistemas.

Los incrementos diferenciales de los que se habla más arriba tienen su reflejo en el empleo, desde hace ya décadas, de misiles del tipo fire-forget dotados de capacidad para gestionar su trayectoria hasta el blanco de forma autónoma.

4 DWORKIN, Gerald. the Concept of Autonomy. *Grazer Philosophische Studien*, 12(1), 2016, pp. 203-213. [en línea]. Disponible en: <https://doi.org/10.1163/18756735-90000122>

5 GÓMEZ de ÁGREGA, Ángel, Ethics of autonomous weapons systems and its applicability to any AI systems. *Telecommunications Policy*, 44(6), 2020, pp. 1-15. [en línea]. Disponible en: <https://doi.org/10.1016/j.telpol.2020.101953>

6 International Committee of the Red Cross (ICRC), *Convention Conventional Weapons (CCW): Statement of the International Committee of the Red Cross (ICRC)*.2018.



Sin embargo, estos misiles –aunque puedan alterar su trayectoria hacia el blanco– carecen de capacidad para variar su objetivo o de abortar la misión en el caso de que “perciban” algún problema. La principal diferencia, por lo tanto, no está en la autonomía de la plataforma o el vector para gestionar su misión, sino en la agencia remanente en el operador y en el momento en que se ejerce.

La IA supone un salto incremental respecto de estos sistemas del siglo pasado y, como tal, se espera de ella una mayor eficiencia y un mayor control.

Cabe recordar también que, aunque los sistemas aéreos de combate sean los más reconocidos, la autonomía se ha trasladado también a los ámbitos terrestre, naval y submarino, e incluso al espacial. Cada uno de ellos presenta retos particulares desde el punto de vista de su empleo ético y legal.

En todas las instancias históricas en que un nuevo sistema de armas ha alterado la dinámica regular del campo de batalla se han producido reacciones en contra e intentos de limitar su presencia y actividad.

Ocurrió así con la llegada de la ballesta, por ejemplo. El 2º Concilio de Trento pretendió excluir su uso “contra cristianos” al mitigar la ventaja que tenían los caballeros montados y con armadura respecto de las huestes de arqueros y ballesteros. La amenaza de excomunión no disuadió, obviamente, a los ejércitos del uso del arma.

Así, frente a los intentos contemporáneos de

limitar el desarrollo de los SALAS, los primeros borradores de regulación o de códigos éticos internacionales apuntan a la necesidad de evitar la ralentización del desarrollo tecnológico que pueda ser de aplicación en ámbitos civiles beneficiosos para la población en general.

El uso dual de casi todas estas tecnologías hace, sin embargo, que el compromiso entre el avance tecnológico y la protección de la dignidad humana sea una línea muy delgada.⁷

Entre estos códigos éticos –formales o informales–, quizás el que más cerca ha estado de un consenso internacional fue el redactado en 2018 por la Convención para Ciertas Armas Convencionales (CCW) de Naciones Unidas, que tiene su sede en Ginebra, Suiza.⁸

El decálogo extractado indica la aplicabilidad del Derecho Internacional Humanitario al empleo de las armas dotadas de IA, requiere la no antropomorfización de los robots (aspecto que podría y debería entenderse como aplicable también a bots no robóticos capaces de imitar la apariencia de la voz u otras características humanas) y la necesidad de no obstaculizar el desarrollo pacífico de la IA.

Más allá del avance ético y regulatorio que puedan suponer la legislación y los códigos éticos diseñados para los SALAS, es de esperar que buena parte de dichos preceptos sean la base para establecer también los límites de los sistemas duales cuyo empleo inicial reside en el ámbito civil y en usos beneficiosos.

7 JONATHON-Penney, J. P. Advancing Human-Rights-By-Design In The Dual-Use Technology Industry. *Journal of International Affairs*, April, 2018. [en línea]. Disponible en: <https://jia.sipa.columbia.edu/advancing-human-rights-design-dual-use-technology-industry>

8 Conventional Committee Weapons (CCW). Consideration of the human element in the development, deployment and use of emerging technologies in the area of lethal autonomous weapons systems. Vol. 15, Issue 29. 2018.



Este sería el punto de intersección entre las legislaciones de unos y otros sistemas dotados de IA toda vez que las normas dictadas hasta el momento para los sistemas civiles excluyen en todos los casos el ámbito militar y, en muchos, el de la seguridad.

Sería así en la legislación europea (IA Act), pero también en diversas normas legales nacionales.

La facilidad de desarrollo y comercio de este tipo de herramientas, que no requieren de una gran infraestructura física, hacen muy difícil la prevención del empleo de sistemas militares fuera del ámbito para el que fueron diseñados. Igual que en el caso de las herramientas cibernéticas, el riesgo de empleo de aplicaciones de “grado militar” en el mundo civil es muy alto.

Sistemas de ayuda a la toma de decisión

La noticia de que Israel estaba utilizando un sistema de selección de blancos apoyado en algoritmos con IA en la guerra que sostiene en Gaza supuso una de las más notorias apariciones públicas de estos sistemas.⁹

Lavender, el sistema utilizado por Israel ejemplifica las posibilidades y los riesgos que tienen este tipo de ayudas a la toma de decisión. Inicialmente, hay que decir que el algoritmo no es el que decide la ejecución de un ataque, ni el blanco, ni el que discrimina el tipo de arma a utilizar ni el momento de hacerlo. Todas estas decisiones son adoptadas por un operador humano.

No obstante, la máquina efectúa de forma previa todos los análisis necesarios para sugerir todos esos aspectos. En numerosas ocasiones, la presión de tiempo –y de exigencia en el número de operaciones a ejecutar– dejan apenas unos segundos al operador para confirmar –o no– el ataque.

En este caso concreto, y siguiendo lo publicado y reconocido en la prensa, el algoritmo ha concluido que el mejor momento para efectuar el ataque es cuando el objetivo se encuentra en su hogar, junto con su familia. De este modo, se minimizan las probabilidades de un error de identificación. Sin embargo, también aumentan significativamente las de causar daños y víctimas colaterales en las personas de los familiares del objetivo.

Así, mientras que los sistemas automatizados de guiado de los vectores hace tiempo que están en servicio y resultan aceptables (pensemos en los misiles fire-forget o en algunos sistemas de defensa aérea), los algoritmos de apoyo a la toma de decisiones apenas si están alcanzando ahora una madurez tecnológica comparable.

La diferencia es, además, importante desde el punto de vista de las capacidades aumentadas o sustituidas. Mientras que en un caso se trata de la habilidad manual para guiar un vector hasta su objetivo, en el otro hablamos de una habilidad cognitiva con un fuerte componente ético y legal.

El ámbito militar no es el primero en el que se ponen en marcha este tipo de sistemas. Previamente ya se han utilizado, por ejemplo, en los sistemas judiciales de algunos países para determinar la

9 McKERNAN, Bethan y DAVIES, Harry. “The machine did it coldly: Israel uses AI to identify 37,000 Hamas targets”. The Guardian, 3 de abril de 2024. [en línea]. Disponible en: <https://www.theguardian.com/world/2024/apr/03/israel-gaza-ai-database-hamas-airstrikes>



idoneidad de un recluso para optar, por ejemplo, a la libertad provisional.¹⁰

La capacidad de los sistemas informáticos para examinar un gran número de sentencias similares, todos los condicionantes concretos del caso y la jurisprudencia al respecto otorga una garantía al juez de no pasar por alto datos cruciales. Del mismo modo –y por el mismo motivo– supone una gran presión para él o ella si su criterio discrepa del consejo del algoritmo.

El carácter maquinal y supuestamente imparcial de las máquinas hace que sean vistas como sistemas casi infalibles. Sin embargo, como cualquier otro sistema, están muy influenciadas por la calidad de los datos sobre los que han sido entrenados (lo que sería un equivalente a su “educación”), la de los que se utilizan en su evaluación del caso y por los criterios que han sido instruidos en seguir.

Los algoritmos son meros extractores de patrones; máquinas estadísticas que regurgitan como resultado el que sea más probable de forma estadística. Si bien eso arroja –estadísticamente– un alto porcentaje de acierto, la decisión no está basada en un razonamiento lógico ni, mucho menos, en principios o valores éticos.

La sensibilidad de las máquinas al “envenenamiento” de los datos sobre lo que trabajan es también muy alta¹¹ Al trabajar sobre realidades sintéticas generadas en un entorno digital, la posibilidad

de sucumbir a un ataque informático que afecte a sus datos es algo que tiene que mantenerse siempre presente.

Por ello resulta imprescindible que la decisión humana tenga lugar en el momento éticamente significativo. Y que lo haga basada en unos criterios no sesgados por el mismo ambiente sintético de los algoritmos. Finalmente, es necesario que el operador disponga del tiempo suficiente para evaluar dicha decisión.¹²

Estas circunstancias son, a menudo, difíciles de encontrar en el ámbito de la guerra. Mientras que un juez puede sentir una cierta presión temporal para adoptar una decisión en un cierto margen de horas o días, el operador militar suele disponer de unos segundos o unos minutos para hacerlo. El caso de Lavender lo ilustra perfectamente. En los sistemas ofensivos, por la pérdida de la oportunidad. En los defensivos, por la premura en la reacción ante vectores cada vez más rápidos, discretos y precisos.

El momento ético adecuado para la intervención humana no tiene que coincidir necesariamente con el del disparo del vector. En muchos casos, la decisión deberá estar tomada bastante antes de accionar el “gatillo” y, a partir de ese momento, se convierte en un sistema fire-forget en el que, idealmente, tiene que estar prevista la capacidad para que, tanto por parte de la máquina como de su operador, se pueda

10 MORLEY, Jessica, FLORIDI, Luciano, KINSEY, Libby, & ELHALAL, Anat. From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Science and Engineering Ethics*, 26 (4), 2020, pp. 2141–2168. [en línea]. Disponible en: <https://doi.org/10.1007/s11948-019-00165-5>

11 Ong, Y.-S., & Gupta, A. AIR 5 : Five Pillars of Artificial Intelligence Research. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 3(5), 2019, pp. 411- 415. [en línea]. Disponible en: <https://doi.org/10.1109/tetci.2019.2928344>

12 EKELHOF, Merel. Autonomous weapons: Operationalizing meaningful human control. *Humanitarian Law and Policy*. 2018. [en línea]. Disponible en: <https://blogs.icrc.org/law-and-policy/2018/08/15/autonomous-weapons-operationalizing-meaningful-human-control/>



cancelar la operación en caso de que varíen las circunstancias.

Por lo tanto, el criterio tradicional de buscar un man-in-the-loop, o un man-on-the-loop que pueda intervenir en cualquier momento o en un momento dado viene subordinado a que esa intervención sea, por un lado, significativa y, por el otro, plenamente informada y comprendida.

Significativa en tanto que se produzca en el momento clave del proceso. Informada y comprendida en el sentido de que lo sea libre de los sesgos inducidos por el mismo sistema (siempre existirán sesgos en el operador, pero no es aceptable que a ellos se añadan los propios de sistema digital).

La creciente importancia de las realidades generadas algorítmicamente (incluso en nuestra interacción cotidiana con los teléfonos celulares) y su superposición con la física ha alterado la dinámica racional de los humanos.

En los sistemas de apoyo a la decisión descritos hasta ahora, la máquina es una herramienta separada del humano y a la que se accede mediante terminales. Estas interfaces permiten mantener un cierto grado de conciencia respecto de la separación entre lo real y lo sintético. Con la adopción de interfaces humano-máquina cada vez más transparentes esta distinción se volverá más sutil.¹³

Realidades sintéticas

La clave para una adecuada toma de decisiones por parte del operador humano reside en la objetividad y falta de sesgos de la realidad percibida. Sin embargo, la presencia de la IA en los procesos de adquisición, filtrado, consolidación y presentación de los datos introduce potencialmente distorsiones con alto grado de verosimilitud que pueden inducir a la toma de decisiones erróneas.¹⁴

El peligro, en este caso, es doble. Por un lado, por la posibilidad de que nuestros propios sistemas induzcan percepciones falaces debido a una incorrecta –o incompleta– selección de los datos o a una programación deficiente de los algoritmos de su procesado.

Por otro lado, también permite que agentes adversarios (no necesariamente enemigos, pero sí con intereses distintos a los nuestros) puedan presentar a nuestra consideración percepciones que generen emociones y motivaciones favorables a sus intereses y, por lo tanto, contrarias a los nuestros.

Existe un tercer riesgo en relación con la edición de los datos objetivos para la obtención de inteligencia. El grado de realismo que se alcanza con las tecnologías actuales (y hay que esperar que se siga superando a cada vez mayor velocidad) afecta directamente a los sentimientos, más allá de nuestras capacidades cognitivas racionales. La difusión de narrativas interesadas propias es

13 LIM, Yixiang, SAMRELOY, Therapong., CHANTARAVIWAT, Cholsanan., EZER, Neta, GARDI, Alessandro., & SABATINI, Roberto. Cognitive Human-Machine Interfaces and Interactions for Multi-UAV Operations. Australian International Aerospace Congress (AIAC18), 2019, pp. 1–7. [en línea]. Disponible en: https://www.academia.edu/39336849/Cognitive_Human-Machine_Interfaces_and_Interactions_for_Multi-UAV_Operations

14 VERDEOLIVA, Luisa. Media Forensics and DeepFakes: an overview. 2020, pp. 1-24. [en línea]. Disponible en: <http://arxiv.org/abs/2001.06564>



susceptible de crear un mundo ficticio que nos impida ver la realidad más allá de la ficción que hemos creado.

Estas cámaras de eco que repiten y refuerzan un relato dentro de un grupo afín desbordan la capacidad de control de su emisor. Una vez lanzado el mensaje, la complejidad del entorno digital en el que se mueve hace que se combine con otros relatos y emociones que lo refuercen más allá de lo reconocible.¹⁵

Es decir, uno de los principales riesgos que se corren en el ámbito de los relatos sintéticos es el de terminar por creer las propias mentiras una vez edulcoradas o reforzadas por agentes afines.

Un ejemplo temprano de este peligro es la simulación generada en la Segunda Guerra Mundial por los japoneses para conocer sus posibilidades en la guerra naval contra Estados Unidos. El *Soryokusen Kenkyujo* (Instituto de Investigación para la Guerra Total) generó este simulador que, como era apropiado al espíritu de la época, mostraba indefectiblemente la victoria japonesa en la confrontación. Para ello, refluotaba continuamente los acorazados japoneses hundidos. En este caso, no obstante, los resultados no fueron tenidos en cuenta, muy probablemente porque el proceso era lo suficientemente artesanal y manual como para que sus creadores fueran plenamente conscientes de la trampa.

Las realidades sintéticas se presentan en forma de realidad virtual (RV) y de realidad aumentada (RA), o de una mezcla de ambas. En el primer

caso, la percepción del “jugador” será de una realidad generada artificialmente sobre la base –o no– de una realidad subyacente. En el segundo, la IA añade un contexto a un fondo real que facilita incrementar la información disponible sobre los objetos y personas que aparecen en el campo de visión.

El grado de inmersión en estas realidades depende, en muy buena medida, de las interfaces que se empleen para su presentación. Esto es, cuanto más “transparente” y menos intrusivo sea el medio empleado para hacer llegar el contenido, mayor será el realismo percibido por el cerebro.

Si partimos de la experiencia acumulada con la utilización de terminales celulares, gafas de RV o equipos completos (normalmente incluyendo gafas, guantes y dispositivos de sonido), la capacidad para generar realidades paralelas en este momento es ya considerable.

Es previsible que las interfaces se mejoren sustancialmente en los próximos años haciendo la experiencia absolutamente realista para el operador.

Las consecuencias psico-sociológicas y operativas que puede tener la incorporación de estas tecnologías, así como la multitud de aplicaciones posibles, están ya siendo objeto de estudios académicos en todo el mundo.

Ciertamente, la introducción de sesgos en la presentación tiene un enorme potencial de alterar la toma de decisiones del usuario.

15 JAMIESON, Katheleen. H., & CAPELLA, Joseph. M. *Echo Chamber*. Oxford University Press. 2010. [en línea]. Disponible en: <https://global.oup.com/academic/product/echo-chamber-9780195398601?cc=es&lang=en&#>



Humanos aumentados

Las tendencias futuras de la aplicación de la IA apuntan a una hibridación de las inteligencias humana y artificial –la inteligencia del carbono y la inteligencia del silicio, como se han dado en llamar– en lo que se conoce como Humanos aumentados.

Aquí serán de particular importancia las interfaces transparentes que se han mencionado ya. El intercambio de información y de funciones entre ambos centros de procesamiento tiene que hacerse de modo que resulte natural y que se produzca en tiempo real.

Los cyborgs resultantes abarcarán versiones con porcentajes muy variables de participación artificial en la toma de decisiones. De nuevo, tampoco aquí hay una cifra que señale taxativamente el límite entre lo aceptable y lo inasumible. Será la retención del control, de la capacidad para realizar juicios no sesgados y la retención de la responsabilidad lo que realmente determine la admisibilidad del sistema.

Algunos estudios ya han comenzado a adelantar las consecuencias que se derivan del empleo –y de la interrupción de este– de estas ayudas cognitivas. Los primeros estudios tienen, precisamente, como referencia los usos militares.¹⁶

Entre los planteamientos estaría el grado de responsabilidad que se le puede atribuir a una persona que recibe información editada por un sistema informático y sobre la base de las cuales tienen que tomar sus decisiones. La falta de objetividad de las percepciones (sesgadas en

función del algoritmo que las presenta) muestra al combatiente con una mayor cantidad de información, pero con la incerteza sobre su veracidad.

La delgada línea entre la provisión de datos y la manipulación de la voluntad vuelve a recordar producciones cinematográficas distópicas.

Se plantea también la disyuntiva sobre si dotar a todos los combatientes de mejoras similares habida cuenta de la desventaja operativa en que quedarían aquellos que fuesen privados de ellas. El envío de tropas al frente con un fusil para cada varios infantes (que es recogido de las manos del caído por uno de los desarmados para continuar el asalto) fue una práctica real en el Ejército Rojo soviético o entre los batallones de “voluntarios” chinos en la Guerra de Corea, pero resulta poco aceptable en las Fuerzas Armadas de los países occidentales contemporáneos.

Humanos disminuidos

El siguiente problema que plantean los implantes e interfaces humano-máquina (HMI, en inglés) es su remoción. ¿En qué condiciones queda un humano después de ser privado de las ayudas cognitivas (o físicas) con las que ha entrenado y combatido una vez apartado del servicio activo? ¿Qué pérdida de capacidades naturales supone el uso continuado de sustitutos artificiales? ¿Se puede plantear mantener estas mejoras en la persona una vez que vuelve a la vida civil? ¿Con qué consecuencias para su equiparación con el resto de la población?

Este difícilmente es un problema nuevo que haya aparecido con la IA. Diversos estudios revelan

16 Defence Ethics Committee, 2020



pérdidas en las capacidades de cómputo o de lógica entre aquellas personas que mayor uso hacen de ayudas informáticas, aunque las interfaces sean tan burdas como las que proporciona un teléfono celular.

No estamos, en estos casos, hablando de terapias reparadoras que revierten pérdidas de capacidad naturales, sino de intervenciones encaminadas a mejorar dichas capacidades.

Conclusiones

La aplicación de sistemas de IA afecta, por primera vez en la historia, a las capacidades propias y distintivas de los humanos. Los verdaderos riesgos de su aplicación –tanto en el ámbito civil como en el militar– estriban en la pérdida de control; no sobre las decisiones de las máquinas, sino sobre las decisiones de los propios humanos.

Es preciso entender la IA como una herramienta capaz de facilitar el ejercicio de los rasgos humanos, no como un sustituto de los mismos.

Los límites son difusos y los conflictos han favorecido siempre opciones arriesgadas que pueden traspasarlos. Alejados del foco que otorga a los SALAS su carácter letal, los sistemas dotados de IA diseñados y empleados para apoyar la toma de decisiones tienen que dotarse de una regulación que evite que los operadores humanos se conviertan en meros ejecutores de sus decisiones.

La guerra es una confrontación entre voluntades humanas. Si estas quedan anuladas o disminuidas por sesgos inducidos por las máquinas, las acciones y sus consecuencias dejarán de poder denominarse guerra.

La clara atribución de la responsabilidad a todos y cada uno de los escalones del ciclo de vida del diseño, desarrollo y empleo de los sistemas dotados de IA debería mitigar la tentación de descargar en los algoritmos las funciones propias de los comandantes.

La hibridación de los sistemas humanos y algorítmicos complica todavía más la ecuación, difuminando la ya difusa relación entre ambos centros de proceso.

Es muy probable que el ámbito militar sea el catalizador de los siguientes avances en la aplicación de estos sistemas. Los límites que se alcancen en el campo de batalla fijarán, probablemente, el nivel ético de los desarrollos civiles.

Bibliografía

BOULANIN, Vincet., DAVISON, Neil., GOUSSAC, Netta., PELDÁN, Moa, & Sipri, C. Limits on Autonomy in Identifying Practical Elements of Human Control. 2020.

Conventional Committee Weapons (CCW). Consideration of the human element in the development, deployment and use of emerging technologies in the area of lethal autonomous weapons systems. Vol. 15, Issue 29. 2018.

Defence Ethics Committee. Opinion on the Augmented soldier. 2020. [en línea]. Disponible en: [https://www.defense.gouv.fr/sites/default/files/ministere-armees/20200921_Comited%27éthique de la défense - Avis soldat augmenté - version anglaise.pdf](https://www.defense.gouv.fr/sites/default/files/ministere-armees/20200921_Comited%27éthique%20de%20la%20d%C3%A9fense%20-%20Avis%20soldat%20augment%C3%A9%20-%20version%20anglaise.pdf)

DWORKIN, Gerald. the Concept of Autonomy. Grazer Philosophische Studien, 12(1), 2016, pp.



- 203-213. [en línea]. Disponible en: <https://doi.org/10.1163/18756735-90000122>
- EKELHOF, Merel. Autonomous weapons: Operationalizing meaningful human control. *Humanitarian Law and Policy*. 2018. [en línea]. Disponible en: <https://blogs.icrc.org/law-and-policy/2018/08/15/autonomous-weapons-operationalizing-meaningful-human-control/>
- GÓMEZ de ÁGREDA, Ángel. Ethics of autonomous weapons systems and its applicability to any AI systems. *Telecommunications Policy*, 44(6), 2020, pp. 1–15. [en línea]. Disponible en: <https://doi.org/10.1016/j.telpol.2020.101953>
- International Committee of the Red Cross (ICRC), Convention Conventional Weapons (CCW): Statement of the International Committee of the Red Cross (ICRC). 2018.
- JAMIESON, Katherine. H., & CAPELLA, Joseph. M. *Echo Chamber*. Oxford University Press. 2010. [en línea]. Disponible en: <https://global.oup.com/academic/product/echo-chamber-9780195398601?cc=es&lang=en&#>
- JONATHON-Penney, J. P. A. Advancing Human-Rights-By-Design In The Dual-Use Technology Industry. *Journal of International Affairs*, April, 2018. [en línea]. Disponible en: <https://jia.sipa.columbia.edu/advancing-human-rights-design-dual-use-technology-industry>
- KRAFFT, P. M., & DONOVAN, Joan. Disinformation by Design: The Use of Evidence Collages and Platform Filtering in a Media Manipulation Campaign. *Political Communication*, 37(2), 2020, pp. 194–214. [en línea]. Disponible en: <https://doi.org/10.1080/10584609.2019.1686094>
- LIM. Yixiang, SAMRELOY, Therapong., CHANTARAVIWAT, Cholsanan., EZER, Neta, GARDI, Alessandro, & SABATINI, Roberto. Cognitive Human-Machine Interfaces and Interactions for Multi-UAV Operations. *Australian International Aerospace Congress (AIAC18)*, 2019, pp. 1–7. [en línea]. Disponible en: https://www.academia.edu/39336849/Cognitive_Human-Machine_Interfaces_and_Interactions_for_Multi-UAV_Operations
- McKERNAN, Bethan y DAVIES, Harry. “The machine did it coldly: Israel uses AI to identify 37,000 Hamas targets”. *The Guardian*, 3 de abril de 2024. [en línea]. Disponible en: <https://www.theguardian.com/world/2024/apr/03/israel-gaza-ai-database-hamas-airstrikes>
- MORLEY, Jessica, FLORIDI, Luciano, KINSEY, Libby, & ELHALAL, Anat. From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Science and Engineering Ethics*, 26 (4), 2020, pp. 2141–2168. [en línea]. Disponible en: <https://doi.org/10.1007/s11948-019-00165-5>
- ONG, Y.-S. & GUPTA, A. AIR 5 : Five Pillars of Artificial Intelligence Research . *IEEE Transactions on Emerging Topics in Computational Intelligence*, 3(5), 2019, pp. 411- 415. [en línea]. Disponible en: <https://doi.org/10.1109/tetci.2019.2928344>
- STOP Killer Robots. Campaña para detener robots Asesinos. Retener el control humano sobre los sistemas de armas. 2018, pp. 1–5.
- VERDEOLIVA, Luisa. Media Forensics and DeepFakes: an overview. 2020, pp. 1-24. [en línea]. Disponible en: <http://arxiv.org/abs/2001.06564>