DESAFÍOS ÉTICOS EN LA APLICACIÓN DE LA INTELIGENCIA ARTIFICIAL A LOS SISTEMAS DE DEFENSA

ETHICAL CHALLENGES ON ARTIFICIAL INTELLIGENGE APPLIED TO DEFENSE SYSTEMS

Juan A. Moliner González
General de División del Ejército del Aire -R- y profesor del Instituto Universitario "General Gutiérrez Mellado"
juan_moliner@msn.com

Fecha recepción artículo: 03/11/2020 • Fecha aprobación artículo: 15/12/2020

RESUMEN:

La Inteligencia Artificial se vislumbra actualmente como el más disruptivo de los avances tecnológicos de la Revolución Digital en curso. En la intensa competición por obtener la supremacía en el tablero geoestratégico internacional y, entre muchas otras consideraciones, la ética está comenzando a desempeñar un papel cada vez más relevante.

En el ámbito de la Defensa, el patrón siempre ha sido, y lo sigue siendo, la incorporación de nuevos desarrollos científicos y tecnológicos en las capacidades militares, tratando de mantener la superioridad sobre rivales y competidores. Algunos de esos progresos, como los Sistemas de Armas Letales Autónomos y otras capacidades militares, que utilizan la Inteligencia Artificial como guía directora, presentan problemas éticos que deben ser considerados no solo por los Estados y sus Fuerzas Armadas, sino también por todos aquellos implicados en su investigación y desarrollo, como los ingenieros y fabricantes.

Entre esos problemas el que se considera en este trabajo como más relevante es el del «control humano significativo», exigido para evitar que la guerra abandone su naturaleza de fenómeno humano y social bajo control del ser humano y se transforme en un juego dirigido por máquinas y su Inteligencia Artificial.

Palabras clave: Inteligencia Artificial, Geoestrategia, ética, Control humano significativo, Sistemas de armas letales autónomos.

ABSTRACT:

Artificial Intelligence is currently foreseen as the most disruptive technological advance of the ongoing Digital Revolution. There is a strong competition for gaining supremacy in the international geostrategic landscape and, among many other elements, ethical considerations are starting to strongly show up.

In the Defense field, the pattern is and has always been to incorporate new tecnological and scientific developments to military capabilities, aiming to gain an advantage over rivals and competitors. Some of these breakthroughts, such as Lethal Autonomous Weapons Systems and other military capabilities using Artificial Intelligence as their central driver, present ethical concerns which must be considered not only by States and their Armed Forces, but also by those involved in their development like engineers and manufacturers.

Among these concerns the one considered most relevant in this work is the «meaningful human control», demanded to prevent warfare from abandoning its nature of human and social phenomenon under human control and becoming a machine and artificial intelligence driven game.

Keywords: Artificial Intelligence, Geostrategy, Ethics, Meaningful human control, Lethal autonomous Weapons systems.

Juan A. Moliner González. *Aunque sus aportaciones iniciales se han centrado en la Estrategia y la Política de Defensa, desde 2007 ha publicado diversos artículos y capítulos en libros centrados en la ética y moral militar, prestando especial atención a las implicaciones de las nuevas tecnologías en el devenir de guerras y conflictos y en los principios y valores de la profesión militar. Ha participado en diversos grupos de trabajo relacionados con la Ética y el Liderazgo militares, y ha colaborado con el Ministerio de Educación en la elaboración de unidades formativas para la promoción de la cultura de Defensa.*

1. INTRODUCCIÓN

La Inteligencia Artificial (IA) es una de las tecnologías, posiblemente la más decisiva, en la transformación social que está produciendo la revolución digital. Si se pretende que sea beneficiosa para las personas y sociedades, la ética, posiblemente la parte de mayor aplicación práctica de la filosofía, debe desempeñar un papel decisivo para que los progresos que se produzcan mejoren las sociedades y promuevan el avance de los Derechos Humanos.

En nuestra época los retos a la Seguridad son multidimensionales y aunque no sean específicos de la Defensa y sus Fuerzas Armadas (FAS), éstas y sus instrumentos se ven implicados a la hora de darles respuesta. La pandemia del coronavirus, con la reciente «operación Balmis» y otras tareas asignadas a las FAS dan muestra de ello.

Esto ocurre también con la IA, que en este mundo de transformación digital que vivimos, asociada a otras tecnologías como la robótica, cibernética, neurociencias, tecnologías de la Información y las Comunicaciones y la supercomputación, por citar solo algunas, ya ha empezado a ser utilizada en el ámbito de la Defensa para desarrollar capacidades militares.

Parece fuera de toda duda que la tecnología de la IA evoluciona a una gran velocidad y se ha convertido en una poderosa herramienta en las esferas política, económica y militar. Su auge se ha trasformado en un factor estratégico de las Relaciones Internacionales y para muchos está llamada a cambiar de forma significativa el orden internacional.

En consecuencia, a los que tratamos sobre la Seguridad y la Defensa nos interesan las condiciones y consecuencias de una IA que se aplica de forma creciente, y creemos se va a aplicar aún más, en la Defensa y en sus sistemas de armas. Esta realidad parece que no se puede cuestionar.

Desde siempre las tecnologías que se han ido perfeccionando por el ser humano se han empleado en guerras y conflictos bélicos. En ocasiones, una vez desarrollados y probados como inventos científicos del mundo civil (bolígrafo, maquinilla de afeitar, conservas); en otras, como inventos específicamente diseñados para su uso militar (satélites, GPS, microondas), que luego se han reintroducido en el mundo civil para intentar satisfacer necesidades o hacer más cómoda la vida de los ciudadanos.

Cada vez que un nuevo desarrollo científico y tecnológico se ha incorporado a la Defensa y, en consecuencia, al conflicto y a la guerra, han aparecido discusiones y planteamientos diversos sobre su idoneidad política, utilidad militar y conformidad ética, pues como dice Ortega Klein: «Esta cuestión de la ética en la IA se engloba dentro de una temática más amplia sobre el control de la tecnología por los propios humanos» (Ortega Klein, 2020, p. 3).

Es la reflexión ética desde la que consideramos a la Inteligencia Artificial como la aplicación a sistemas de armas, muchos de ellos letales, de una capacidad para el análisis y la toma de decisiones que hacen que la máquina, el arma, funcione o pueda funcionar independientemente del ser humano, pero actuando sin sus limitaciones morales y con una inteligencia que supera o puede superar a la humana.

Las consideraciones éticas, que junto a las geoestratégicas son las que centran este análisis, se refieren a las capacidades militares en su conexión con la IA. Esas capacidades tienen un amplio espectro[1], tanto de sistemas no letales como letales y a todos se les pueden aplicar esas reflexiones, pero que sin duda tienen su consideración más profunda y grave en los sistemas que tienen la capacidad de producir muerte y destrucción, como los sistemas de armas letales autónomos (SALAS).

Las cuestiones sobre la ética y la IA se están debatiendo en nuestros días por multitud de interesados y afectados, gobiernos, organismos gubernamentales y no gubernamentales, empresas y, por supuesto, personas individuales. El resultado, como siempre que se habla de ética, es la existencia de una amplia diversidad de conceptos y posiciones, que, de momento, no se es capaz de consensuar en normas y regulaciones.

Se anticipa que, a menudo, se plantearán problemas y cuestiones más que soluciones, en la esperanza de que puedan aportar algo al debate en marcha y que en España tiene un importante punto de inflexión en la próxima aprobación de la Estrategia Española en Inteligencia Artificial[2].

[1] A este respecto, más información puede consultarse en el Documento de Trabajo 04/2019 del Instituto español de Estudios Estratégicos (IEEE) titulado: Usos militares de la inteligencia artificial, la automatización y la robótica (IAA&R) y el Documento de Trabajo 6/2018: La inteligencia artificial aplicada a la Defensa, del mismo IEEE.

[2] La Estrategia Nacional de Inteligencia Artificial se presentó por el presidente del Gobierno español, Pedro Sánchez, el 2 de diciembre de 2020, después de la recepción de este artículo (nota del editor). https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/ENIA2B.pdf

2. ESTRATEGIA Y ÉTICA DE LA INTELIGENCIA ARTIFICIAL

El mencionado Ortega Klein hace unas consideraciones generales sobre la geopolítica de la ética en la IA, en la que muestra las aproximaciones y consideraciones éticas que Estados, organizaciones, empresas y, también, los ingenieros e investigadores de la IA tienen al respecto.

Las estrategias y documentos similares, aprobados o en vías de promulgación, sobre la IA, también tienen consecuencias éticas y deben responder a un riguroso escrutinio moral en la que están implicados no solo valores morales sino la responsabilidad social y política como sociedades e incluso como especie humana.

La IA está actualmente en el centro de las preocupaciones geopolíticas de los Estados, sean grandes potencias o emergentes, así como de las organizaciones supranacionales que aspiran a intervenir en las relaciones internacionales con autonomía estratégica. Las posiciones de los países más significativos en relación con el desarrollo de la IA desde posturas éticas se reflejan a continuación.

Así, en los Estados Unidos, el presidente Trump firmó el 11 de febrero de 2019 una Orden Ejecutiva[3] para mantener el liderazgo norteamericano en IA que refleja la posición de este país. En su Sección 1 se recoge que «el continuo liderazgo americano en IA es de primordial importancia para mantener la seguridad económica y nacional de los Estados Unidos y para dar forma a la evolución global de la IA de forma consistente con los valores, políticas y prioridades de nuestra Nación».

Quizá por esto, en el ámbito de la Defensa, el Centro Conjunto de IA del Pentágono informó el pasado 24 de febrero que se habían adaptado los principios para un diseño, desarrollo, despliegue y usos éticos de capacidades militares habilitadas por la IA en todo el departamento de Defensa[4]. Estos principios son los de Responsabilidad, Equidad, Identificabilidad, Fiabilidad y Gobernabilidad. El teniente general Shanahan, jefe del Centro Conjunto, en relación con esos principios reafirmó dos ideas principales:

- EE. UU. está comprometido con la ética y jugará un papel de liderazgo para asegurar que las democracias adopten las tecnologías emergentes de forma responsable.
- Constató que tienen mucho en común con los de los aliados y socios de EE. UU. y contrastan con los de China y Rusia que plantean serias preocupaciones sobre Derechos Humanos, ética y normas internacionales.

China, aunque ha tardado en considerar los principios éticos, a mediados de 2019 publicó un documento oficial: «Principios de Pekín de IA», elaborados por la Academia de IA[5]. Distingue entre principios aplicables a la investigación y aquellos aplicables al uso de la IA.

Aunque algunos de ellos muestran una asombrosa voluntad de aproximarse a Occidente en estas cuestiones las dudas persisten, pues formalmente el documento de Pekín dice: *Human privacy, dignity, freedom, autonomy, and rights should be sufficiently respected*. La palabra *sufficiently* introduce incertidumbre sobre asuntos como la privacidad o si las libertades serán respetadas en un país que, hasta ahora, no ha protegido los derechos individuales, algo que constituye un imperativo moral para Occidente.

[3] https://www.presidency.ucsb.edu/documents/executive-order-13859-maintaining-american-leadership-artificial-intelligence
[4] https://spacenews.com/dod-adopts-new-ethical-principles-for-the-use-of-artificial-intelligence
[5] https://www.baai.ac.cn/blog/beijing-ai-principles

Para China la realidad en el uso de sistemas de IA (reconocimiento facial, vigilancia, control social, etc.), no solo es una herramienta de represión, sino una forma alternativa de gobierno muy alejada del respeto a la voluntad de ciudadanos libres, pero que intenta legitimar como respuesta a las necesidades de la gente.

Rusia, aunque tiene una «Estrategia Nacional para el desarrollo de la IA en el periodo 2020-2030», no aborda en ninguno de sus diez puntos cuestiones éticas, aunque sí se «determina la creación de un sistema integrado para regular la acción de la IA, que formule reglas éticas para la interacción humana con la Inteligencia artificial» (Ortega K., 2020, p. 17).

Por su parte, la Unión Europea (UE) ha producido diversos documentos que inciden en los principios éticos en el diseño, desarrollo y utilización de la IA. Destacamos:

- El documento *Ethics Guidelines for Trustworthy Artificial Intelligence*, preparado por un Grupo de Expertos de Alto Nivel y publicada el 8 de abril de 2019[6]. Sus 7 principios son: agencia y supervisión humana; robustez técnica y seguridad; privacidad y gobernanza de datos; transparencia; diversidad, no discriminación y equidad; bienestar social y ambiental; y responsabilidad.

- El *Report on liability for Artificial Intelligence and other emerging technologies*, preparado por un Grupo de Expertos en Responsabilidad y Nuevas tecnologías, de 21 de noviembre de 2019 (Peets et al., 2019).

En relación con las aplicaciones militares de la IA, una Resolución del Parlamento Europeo de 12 de septiembre de 2018[7], ha solicitado a los Estados miembro que elaboren una posición común sobre los sistemas de armas autónomos letales que garantice un «control humano significativo» de los mismos. El significado de este concepto, aunque aún discutido y no plasmado en una posición común internacional, se refiere a la necesidad de que se mantenga un control humano sobre todos los sistemas de armas, lo que significa la atribución de responsabilidad en cualquier ocasión a una persona y la verificabilidad de sus decisiones y consecuencias en la eventual utilización de sistemas de armas autónomos letales y también no letales.

La UE, en su desarrollo de la IA plantea un modelo industrial ético, considerado por algunos como una isla en medio de los planteamientos globales y en el que algunos han visto en este proyecto la última señal de división entre Europa y EE. UU. y China sobre la ética de la IA.

Por su parte, los países de la OCDE firmaron en mayo de 2019 sus «Principios sobre IA»[8]. También el G-20 en la reunión ministerial de junio de 2019 aprobó unos «Principios para la IA», que han sido firmados por todos los participantes, incluyendo China, Rusia y Arabia Saudí[9].

[6] https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines

[7] https://www.europarl.europa.eu/doceo/document/TA-8-2018-0341_ES.html

[8] https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 Los principios son: la IA debe beneficiar a los países y al planeta: los sistemas de IA deben diseñarse respetando el Estado de Derecho, los Derechos Humanos, los valores democráticos y la diversidad: debe haber trasparencia y divulgación responsable: los sistemas de IA deben diseñarse de manera robusta y segura: las organizaciones e individuos que desarrollan, implementan u operan sistemas de IA deben ser responsables de su correcto funcionamiento.

[9] OECD (2019), «Recommendation of the Council on Artificial Intelligence», https://legalinstruments.oecd.org/api/print?ids=648&lang=en. Son: crecimiento inclusivo; desarrollo sostenible y bienestar; valores centrados en el ser humano y equidad; trasparencia y explicabilidad; robustez, seguridad y responsabilidad.

No solo Gobiernos u organismos internacionales se hacen eco de consideraciones éticas en el desarrollo e implementación de la IA, también grandes empresas como Microsoft, Google, IBM, o en España Telefónica[10], plantean y centran sus preocupaciones éticas en línea con los principios y áreas anteriormente mencionados. En general, indica Ortega: «Hay una gran resistencia por parte de algunas empresas a la regulación. Las empresas tienden más bien a propugnar una autorregulación, que, como se está comprobando en la práctica, no es suficiente» (Ortega K., 2020, p. 22). Pero no todas. Por ejemplo, Microsoft ha realizado declaraciones solicitando una regulación general y aceptada[11].

También los individuos tienen un importante papel en la ética de la IA, particularmente los científicos que en los diferentes saberes están implicados en su desarrollo. De momento da la sensación de que los problemas y los posibles códigos éticos a establecer no les han afectado mucho, como tampoco parece que la formación que tienen no está, hasta ahora, interesada en cubrir este campo.

Ortega cita a Robert McGinn que, en su libro *The Ethical Engineer. Contemporary Concept and Cases (2018)*, indica que los ingenieros tienen que asumir cuatro responsabilidades fundamentales, que resumidas son:

- No causar daño o no crear un irrazonable riesgo de daño a los demás con su trabajo.
- Intentar prevenir el daño anterior que puede ser causado por su propio trabajo, el de otros ingenieros si él está involucrado o el de un trabajo que él conozca.
- Intentar alertar e informar sobre el riesgo de daño en los casos anteriores.
- Trabajar lo mejor posible para atender los intereses legítimos de su empleador o cliente.

Haciendo una recapitulación de los principios éticos recogidos por los diferentes actores mencionados, vemos que algunos se repiten sistemáticamente y se consideran los más comunes. Serían:

- La IA debe estar centrada en los seres humanos.
- El uso y despliegue de aplicaciones de IA debe ser confiable.
- La IA tiene que respetar la autonomía humana.
- Hay que prevenir la posibilidad de hacer daño.
- Es necesario mantener la equidad y que nadie se quede atrás.
- Los resultados de los desarrollos de la IA deben ser explicables.

Lo que parece más evidente es que el riesgo está en la aplicación, no en la tecnología en sí misma, pues resulta difícil trasladar automáticamente esos principios a la práctica, tanto por los diversos enfoques entre organizaciones, países, empresas e individuos, como por la dificultad de alcanzar una regulación que debe ser consensuada internacionalmente, dado el alcance global y estratégico (desde el punto de vista militar) de la IA.

[10] En su «Manifiesto por un Nuevo Pacto Digital», Telefónica establece como principios a seguir el establecimiento de un marco ético sobre los datos, en el que exista trasparencia y las personas deciden cómo y cuándo se utilizan los datos y disfruten su valor, además de que exista responsabilidad; y un desarrollo de la IA que se centre en las personas teniendo en cuenta consideraciones éticas y valores establecidos.

[11] https://www.geekwire.com/2020/microsoft-president-brad-smith-calls-ai-regulation-davos/

Desafíos éticos en la aplicación de la inteligencia artificial a los sistemas de defensa. Juan Moliner
Revista DIECISIETE. 2021. Nº 4. Madrid. ISSN 2695-4427 Versión digital (pp. 111-126) DOI: 10.36852/2695-4427_2021_04.06

3. ESPAÑA Y LA ESTRATEGIA NACIONAL DE INTELIGENCIA ARTIFICIAL

Cuanto mayor sea el control de la máquina en la toma de decisión, mayores serán las implicaciones en los aspectos de Seguridad. En el Informe de la Fundación ESYS, resultado del seminario celebrado en septiembre de 2020[12], y en el que el autor de este artículo tuvo la oportunidad de participar, se considera que la IA, en cuanto elemento clave en la automatización de la toma de decisiones, tendrá diferentes aplicaciones en el ámbito de la Seguridad (que también incluye a la Defensa). Se destacan las siguientes perspectivas:

- Inteligencia Artificial para la Seguridad. Utilización de la Inteligencia Artificial como una tecnología que mejora las prestaciones de las herramientas y sistemas de ciberseguridad.
- Inteligencia Artificial como una nueva amenaza de Seguridad. Al igual que la IA puede ayudar a potenciar la ciberseguridad, igualmente también puede utilizarse como una nueva amenaza. La IA ha pasado a incorporarse en las tecnologías utilizadas en los ciber-ataques, en una gama que abarca desde las ciber-guerras, el ciber-espionaje, el ciber-terrorismo, el ciber-crimen, a la más simple ciber-malicia.
- Seguridad para la Inteligencia Artificial. La IA, como nueva tecnología, y dado su carácter transversal, amplía la superficie de ataque, y por tanto requiere medidas específicas para garantizar la robustez, resiliencia y seguridad de los sistemas que la incorporan. Son muchas las tácticas de ataque que se está utilizando contra sistemas que incorporan inteligencia artificial, que abarcan desde al acceso a los datos, hasta la contaminación o manipulación de estos.

El Ministerio de Ciencia, Innovación y Universidades, promulgó en marzo de 2019 la «Estrategia Española de I+D+i en IA», que establecía la elaboración de una «Estrategia Nacional para la IA» en cuyo trabajo un grupo de expertos tanto públicos como privados están elaborando[13]. Para llevarla a cabo, en borradores previos a los que se ha tenido acceso, se han establecido unas prioridades, siendo la sexta la que exige tener en cuenta consideraciones con implicaciones organizativas y éticas como:

- Los desarrollos de la tecnología de la IA deberán evitar el sesgo negativo y los prejuicios de género u otras formas de discriminación.
- El Comité Español de Ética debe liderar las actividades de análisis y valoración de los aspectos éticos del uso e implantación de la IA en las actividades desarrolladas en los Planes Estatales de I+D+i.
- Desde una visión multidisciplinar, el diseño general de los sistemas de IA se hará alineando los aspectos éticos, legales y sociales.
- Se contribuirá a la redacción de un Código Ético de la IA codesarrollado a nivel interministerial.

[12] Fundación ESYS. https://www.fundacionesys.com/en/noticias
[13] Véase nota 2

Con el compromiso de ser aprobada este año 2020, la participación y contribución del Ministerio de Defensa en la elaboración de la Estrategia Nacional de IA[14] se ha llevado a cabo a través de la Dirección General de Armamento y Material. Esta implicación del Ministerio de Defensa ha sido intensa y muy activa. Así, se considera que:

- La financiación debe alinearse con los Programas Europeos que disponen de financiación orientada a la IA como, por ejemplo, el Programa Marco de I+D+I H2020 o el futuro Horizonte Europa; el Programas *Life*; el Programa Europa Digital (DEP); el Programa Europeo de Desarrollo Industrial en materia de Defensa (EDIDP); o el futuro Fondo Europeo de Defensa (EDF).

- Hay que promover la explotación las sinergias entre la investigación civil y la investigación en defensa, aprovechando el Protocolo General de Actuación vigente entre el MINISDEF y MICIU, CDTI y AEI.

- Debe abordarse la utilización de la IA, en su dimensión de FAS, incorporando la IA en sensores, plataformas, sistemas de mando y control, etc., utilizados en las misiones asignadas. Esta segunda dimensión es la más compleja técnicamente y la que exige un mayor esfuerzo inversor y tiene un carácter estratégico para España, tanto por las mejoras que puede proporcionar a la operatividad de nuestras FAS, redundando en una mayor seguridad para el país, como por ser la dimensión que permite la capacitación del tejido tecnológico nacional (centros tecnológicos y empresas).

En esta reflexión desde la ética militar, la Estrategia Nacional de IA que se está planteando debería tener en cuanta algunos aspectos que se consideran relevantes en relación con la Seguridad y la Defensa (aunque su alcance es obviamente general), y que serían:

- Consideración de la reflexión ética, entre otros aspectos, para que se haga un uso de la IA que respete los principios y valores constitucionales e identifique aspectos normativos clave para la implantación de la IA en España de forma confiable y segura al servicio del desarrollo económico y social

- Ante la posible creación de un «Comité Nacional de Ética de IA», en el mismo se deberían analizar también las cuestiones relacionadas con las capacidades militares, sin obviar las referidas a los sistemas de armas autónomos letales. Campo sobre el que se estima hay una limitada capacitación en España.

4. DESAFÍOS Y RETOS ÉTICOS EN LA UTILIZACIÓN DE SISTEMAS DE ARMAS REGIDOS POR LA IA

Tras las consideraciones sobre la situación política y estratégica de la ética y la IA, entremos ahora en el campo concreto de la Defensa, con la determinación de los problemas éticos que en el futuro plantea la utilización de sistemas militares regidos por esa IA y sus algoritmos diseñados para la guerra.

Aunque la evolución en el desarrollo de guerras y conflictos ha sido constante, lo disruptivo de la trasformación que se produce en nuestros días con las nuevas tecnologías es que alcanza a los principios éticos y valores morales que constituyen el sustrato de nuestras sociedades democráticas.

[14] Ya se ha mencionado la presentación de la Estrategia Nacional de Inteligencia Artificial el pasado 2 de diciembre, pocas semanas después de la redacción de este artículo.

Aceptamos el derecho de nuestras sociedades a la defensa militar, sustentada en todos los supuestos de legalidad y legitimidad, como un servicio público que contribuye a la salvaguardia de los valores e intereses de un sistema político que promueve la libertad, la dignidad del ser humano (con su sentido moral) y los principios democráticos. En la búsqueda de la disuasión y superioridad para la defensa militar, esta recibe el apoyo de la tecnología, lo que significa que la misma se debe poner, en nuestras sociedades, al servicio de consideraciones éticas.

«El avance científico y tecnológico es una producción humana, y además de ser en sí mismo posible causa de conflicto, está haciendo aflorar cuestiones éticas que afectan al desarrollo, empleo y control de nuevas armas y sistemas incorporados al conjunto de las capacidades militares» (Moliner, 2018, p.6).

Tecnología y nuevos sistemas de armas van suplantando progresivamente al ser humano, que está dotado de razones y emociones, tanto para el control de la ejecución, como en la cadena de mando que toma la decisión final del empleo de la fuerza letal contra el enemigo.

En nuestro país, el documento del Estado Mayor de la Defensa «Entorno operativo 2035»[15] refleja la necesidad de prestar atención a la operación de sistemas autónomos regidos por la IA en el campo de batalla y al grado de autonomía y seguridad de esos sistemas.

Especialmente importante es que la IA contribuye de forma notable a la toma de decisiones, de forma que integrada en sistemas cada vez con mayor autonomía, estos puedan llegar a tomar sus propias decisiones. Decisiones que dependerán de los algoritmos codificados y que habrán de ser preparados por un ser humano. De aquí la trascendencia de que en esas líneas de código los dictados éticos sean incluidos de forma apropiada.

Decisiones, por otro lado, que no solo serán de naturaleza táctica u operacional, sino que también pueden contribuir al asesoramiento en los más altos niveles estratégicos y políticos, lo que refuerza la necesidad de los planteamientos éticos en sus análisis, algo que debiera tener muy en cuenta las limitaciones que al respecto presenta la Inteligencia Artificial.

En el campo operativo hay que recordar que las acciones militares, por muy bajo que sea el escalón que las lleva a cabo, en los conflictos asimétricos e irregulares de este siglo XXI, tienen una gran implicación estratégico-política.

Dada la obligación ética de todo comandante militar de reducir al máximo las bajas entre los combatientes a sus órdenes y los civiles de ambos bandos, el empleo de Sistemas de Armas Autónomos apoyados en la IA contribuye a ese objetivo. Los Sistemas de Armas Autónomos (SAA) son descritos como aquellos sistemas que gracias a la IA «una vez activados, son capaces de seleccionar y atacar objetos sin una intervención adicional de un operador humano» (López-Sánchez, 2017, p. 12).

Las misiones que pudieran ser desempeñadas por sistemas autónomos con mayor eficacia en entornos peligrosos, desde la inteligencia, vigilancia y reconocimiento del campo de batalla en todos los dominios hasta el desminado y el combate en áreas urbanas, la lucha antisubmarina o las misiones de interdicción aérea, evitan la sobreexposición a riesgos de soldados mucho más vulnerables físicamente y con menos capacidades funcionales, además de aumentar la rapidez de reacción y respuesta. También proporcionan otras ventajas en los ámbitos de la fiabilidad, el coste y la multiplicación de fuerzas.

[15] https://publicaciones.defensa.gob.es/entorno-operativo-2035-libros-papel.html

También se argumenta en sentido favorable a los sistemas autónomos y robóticos, que además de evitar las bajas militares propias causadas por el conflicto, las máquinas no estarían sujetas al influjo de emociones humanas como la ira o el temor, por lo que sería mucho más difícil que cometieran violaciones o actos punibles en conflictos y guerras. Por ejemplo, disparando a un combatiente ya fuera de combate, dejándose llevar por la colera y el arrebato de ver a compañeros caídos en el fragor de la lucha.

La consecuencia negativa de estas últimas ideas es que, al ser sustituidos los combatientes por robots dirigidos por la IA, la limitación en el número de pérdidas humanas propias y ajenas, que tantas restricciones induce a los decisores políticos para emplear la fuerza, hace que estas desaparezcan pudiendo llevar a una multiplicación de los conflictos. Como siempre ocurre ante las grandes cuestiones que plantean dilemas éticos y morales, también se puede plantear que la IA ayudaría a los responsables de esas decisiones a tomarlas con mayor eficacia y efectividad.

Volviendo al problema ético que la IA plantea en su aplicación a los sistemas de armas, lo esbozamos así: ¿podemos dejar a una IA que es, que va a ser, el centro, la entidad que dirige el empleo de los sistemas de armas, particularmente los Sistemas de Armas Autónomos Letales, que tome, de forma totalmente autónoma, decisiones sobre el uso de la fuerza letal en el complejo campo de batalla del presente y del futuro, independientemente del ser humano? Aunque compleja, la cuestión se plantea de forma más sencilla en términos éticos: ¿dejamos la responsabilidad de decisiones sobre la vida y la muerte en la guerra a la IA y a las máquinas, desentendiéndonos los humanos?

Habitualmente el Derecho, incluyendo el Internacional, se plasma en normas legales inspiradas en principios éticos y en los Derechos Humanos y por esto es tan importante que todos los aspectos referidos al empleo de la IA en la Defensa sean objeto:

- De la reflexión ética sobre las razones para su empleo y las consecuencias de este, y esto desde el momento inicial de su concepción, diseño y desarrollo por científicos e investigadores.
- De la búsqueda de normas legales consensuadas para su regulación.

Aquí entra en acción y tiene pleno sentido que en las estrategias que se elaboren sobre la IA se tenga en cuenta su utilización en la Defensa y especialmente en los sistemas de armas.

En relación con el importante principio ético y legal de la discriminación tiene enorme importancia la relativa a delegar las decisiones de elegir y atacar objetivos militares, aquellos cuya destrucción parcial o total supone una ventaja militar definitiva en el desarrollo de las operaciones. Pero esa decisión es una obligación para seres humanos bajo el Derecho Internacional Humanitario y dada la dificultad de interpretación de este en la confusión del conflicto (la denominada «niebla de la guerra»), desde la perspectiva de la IA, aparece el problema ético de la «delegación de funciones a un algoritmo».

Los riesgos de esta «delegación» afectan a aspectos éticos, además de jurídicos. Por ejemplo, en selección y ataque a objetivos con sistemas autónomos se argumenta que no se puede dejar la responsabilidad de esa decisión en máquinas y robots por la trascendencia de la evitación de víctimas inocentes, la diferenciación entre civiles y combatientes, la responsabilidad del que decide su utilización y su rendición de cuentas, además de la falta de empatía de los sistemas autónomos si llegan a tener «la capacidad de seleccionar a los objetivos y atacar a estos por su cuenta en un conflicto» (Travieso, 2015, p. 2).

«Lo militar se ha digitalizado y se está robotizando de forma acelerada. La inteligencia artificial es ya un componente indispensable de las Fuerzas Armadas, y de las de seguridad en sentido amplio, con el

Desafíos éticos en la aplicación de la inteligencia artificial a los sistemas de defensa. Juan Moliner
Revista DIECISIETE. 2021. Nº 4. Madrid. ISSN 2695-4427 Versión digital (pp. 111-126) DOI: 10.36852/2695-4427_2021_04.06

riesgo de perder el control» (Ortega K., 2020, p. 198). Esta idea nos lleva a que el empleo de sistemas de armas regidos por la IA, desde una perspectiva ética, plantea ventajas e inconvenientes. Entre las primeras se pueden citar:

- El comandante militar tiene que seguir un principio ético que es el de evitar a sus subordinados el riesgo innecesario para evitar pérdidas humanas propias.
- El uso de la IA aumenta o puede aumentar la eficacia y la disminución de errores, pues las máquinas ni se cansan ni se dejan llevar por las emociones.
- Hay una clara disminución de costes.
- Se podrían reducir los daños colaterales.

Como inconvenientes se encuentran:

- El escalamiento del conflicto puede ser mucho más rápido, al utilizar las máquinas una escala de tiempos varios ordenes de magnitud inferior a los tiempos humanos de reacción.
- Se produce una disminución o perdida de la responsabilidad humana, tanto ética como legal, responsabilidad que no tiene, de momento, la IA.
- Se deja la decisión sobre la vida y la muerte -eso es el combate militar- a las máquinas regidas por la IA y el humano se desentiende una vez activadas.
- En la complejidad enorme de la guerra, es imposible planificar y programar todas las situaciones que se pueden dar y en muchas de ellas los humanos aplicamos nuestra conciencia, el sentido común y las emociones. De momento la IA no tiene ninguna de estas capacidades. Pero ¿podrá llegar a tenerlas y desarrollar todas las exigencias que impone el Derecho Internacional Humanitario? Nos referimos fundamentalmente a la discriminación y respeto a los no combatientes y la proporcionalidad en la violencia letal y destrucción a causar por los sistemas de armas que se empleen.

Las posturas ante estos dilemas morales pueden ir desde la impugnación absoluta hasta la defensa a ultranza de los sistemas de armas letales autónomos dirigidos por la IA. De esta forma, nos encontramos con posiciones de rechazo como las siguientes:

- El movimiento *Stop Killer Robots*. Movimiento de NGO e intelectuales de muy amplia difusión[16].
- La ONU, con su «Convención sobre Ciertas Armas Convencionales» y el Grupo de Expertos Gubernamentales[17], donde muchos abogan por la prohibición a través de un tratado internacional.
- El Comité Internacional de la Cruz Roja, que defiende la necesidad de una norma restrictiva legal y global[18].

[16] https://www.stopkillerrobots.org/learn/?lang=es

[17] Organización de Naciones Unidas (2019). Informe del Grupo CCW/GGE.1/2019/3, https://undocs.org/pdf?symbol=es/CCW/GGE.1/2019/3

[18] https://www.icrc.org/spa/resources/documents/faq/autonomous-weapons-20

En su defensa aparecen:

- Algunos científicos implicados en el desarrollo de la IA que propugnan la posibilidad de desarrollar una ética robótica. Peter Singer defiende que las máquinas dirigidas por la IA y gracias a esta pueden llegar a ser más éticas que los humanos en el uso de la fuerza letal (Singer, 2009).

- El Principio de prevención, que exige a los científicos de la IA que desde el comienzo de sus diseños e investigaciones configuren sus algoritmos de forma que se prevengan los efectos perniciosos que pueden producir las máquinas.

Parece claro que las capacidades militares que utilicen la IA, en particular los sistemas de armas letales autónomos, deben tener en cuenta el concepto clave de control humano significativo. A este respecto hay que diferenciar:

- Sistemas militares *human in the loop*: sistemas semiautónomos en los que el hombre decide qué objetivos se van a seleccionar y atacar y el sistema ejecuta la acción con completa autonomía.

- Sistemas militares *human on the loop*: sistemas en los que el hombre no decide los objetivos a seleccionar y enfrentar, tarea que lleva a cabo el sistema de forma independiente, pero aquel puede intervenir en la máquina y modificar su funcionamiento o pararla completamente en cualquier momento que observe un fallo o disfunción.

- Sistemas militares *human out of the loop*: sistemas capaces de operar sin intervención de un operador. El hombre no decide los objetivos a seleccionar y enfrentar y el sistema lleva a cabo con plena autonomía esas funciones sin que aquel pueda intervenir en ningún momento, aunque lo considere necesario.

Para avanzar en la búsqueda de consenso internacional, que de momento parece lejano, la ONU ha celebrado en Ginebra desde el año 2014 y en el marco de la Convención de 1980 sobre las Prohibiciones o Restricciones en el uso de Ciertas Armas Convencionales (*Certain Conventional Weapons:* CCW, en inglés), diversas reuniones de un Grupo de Expertos Gubernamentales sobre las tecnologías emergentes en el Área de los Sistemas de Armas Autónomos Letales para lograr un marco regulador internacional, o incluso una prohibición total de los mismos.

Tras las sesiones celebradas en 2019 y debatirse cuestiones tecnológicas, militares, legales y éticas, dadas las diferentes posiciones de los Estados, el grupo continuará las discusiones, que han quedado abiertas y en las que se tratará de enmarcar definiciones, principios y otros conceptos que faciliten la posibilidad de llegar a algún tipo de consenso con alcance político internacional. Se trata de evitar que el no acuerdo pudiera llevar a una escalada

Las conclusiones más relevantes de esas sesiones, a las que se tuvo la oportunidad de asistir, fueron (véase nota a pie de página 17):

- Sobre las potenciales aplicaciones militares de esos sistemas:

  - En el diseño, desarrollo, pruebas y despliegue de los SALAS se deben tomar todas las precauciones para evitar bajas y daños a civiles y sus propiedades, así como los riesgos de ataques no deseados, pérdida de control sobre los sistemas, la posibilidad de proliferación de estos y el hackeo o la adquisición por grupos terroristas.

  - Los interfaces hombre-máquina deben ser comprensibles, el entrenamiento del personal adecuado y el establecimiento de doctrinas, procedimientos y eventuales reglas de enfrentamiento adecuados.

Desafíos éticos en la aplicación de la inteligencia artificial a los sistemas de defensa. Juan Moliner
Revista DIECISIETE. 2021. Nº 4. Madrid. ISSN 2695-4427 Versión digital (pp. 111-126) DOI: 10.36852/2695-4427_2021_04.06

- Sobre la interacción hombre-máquina:
  - » Es preciso asegurar que su uso sea de conformidad con el Derecho Internacional Humanitario, teniendo en cuenta el contexto operativo y las capacidades del sistema y, en particular, que cumplan los requisitos de distinción, proporcionalidad y precaución en el ataque.
  - » Esos principios deben aplicarse a través de un sistema de mando y control responsable.
  - » Los Estados, las partes y los individuos que tomen parte en un conflicto en el que se emplean estos sistemas mantienen la responsabilidad de sus acciones y los primeros deben asegurar la apropiada rendición de cuentas individual.
- Sobre el juicio humano:
  - » Es esencial para cumplir con las obligaciones del Derecho Internacional Humanitario en el uso de estos sistemas, apoyado en la información disponible en el momento.
  - » En cualquier caso, los civiles y los no combatientes están protegidos por los principios derivados de la costumbre, el principio de humanidad y los dictados de la conciencia pública.
  - » Algunos Estados defienden que estos sistemas reducen el error humano y aumentan la precisión de los ataques por lo que su utilización es compatible con el Derecho Internacional Humanitario, mientras que otros demandan una utilización basada siempre en el juicio humano.

Ante todos estos desarrollos futuros que la IA podría introducir en robots y sistemas de armas a los que se pretende dotar de restricciones éticas en su arquitectura algorítmica, las objeciones que se presentan se expresan a continuación:

- La objeción epistemológica. Implementar un «software ético» en los robots implica una reducción de la ética militar, a menudo compleja reflexión y decisión, a procedimientos algorítmicos que al tener que basarse en normas concretas y prefijadas implican la elección de un bien, de una conducta, en detrimento de otras. Pero la conciencia ética, el respeto al principio de humanidad en el combate, no puede ser trasladada, al menos de momento, en algoritmos de computación.
- La objeción antropológica. El uso de robots dirigidos por la IA lleva a la deshumanización del conflicto bélico, al introducir el enorme riesgo de que los humanos se liberen consciente o inconscientemente de su responsabilidad. Al quedar exonerados de responsabilidad podría resultar más cómodo y fácil consentir que sean los robots autónomos armados los que se impliquen en el combate, tolerando que sean ellos los que tomen las decisiones y pretendiendo olvidar la responsabilidad humana esencial en la utilización de la fuerza letal.
- La objeción sobre la legitimidad de la causa. En esta línea, en muchas operaciones modernas un objetivo esencial es ganar «la mente y los corazones» de las poblaciones locales en las que despliegan fuerzas, pues es la única forma de que el combate tenga legitimidad, y ello suele implicar el aumento del riesgo en las tropas propias. Que sean robots, posiblemente, arruinaría la consecución de ese objetivo.

5. CONCLUSIONES

Considero que la necesidad de un «control humano significativo», tal y como se ha definido, sobre cualquier sistema de armas autónomo guiado y regido por la IA, resulta un imperativo desde posiciones éticas, especialmente en sociedades que, envueltas en una auténtica revolución digital, se asientan en la primacía de los Derechos Humanos y la dignidad del ser humano.

Si como postulan algunos las dos grandes crisis del futuro serán el «calentamiento global y la inteligencia artificial» (Lezama, 2020), debemos prepararnos para hacerles frente en todos los campos, y en esa tarea el papel de la ética en cuanto reflexión humana sobre lo correcto y lo incorrecto en el comportamiento de individuos y grupos sociales, incluyendo los implicados en la Seguridad y la Defensa, es muy relevante.

Por esto, las consideraciones éticas en el uso de la IA en capacidades militares y sistemas de armas y las regulaciones que en torno a la misma se produzcan deberán tener en cuenta las consideraciones presentadas y otras que exceden este trabajo, como la decisión ética cuando se trata de definir los umbrales o niveles de confianza de una cualificación del objetivo, o la regulación que se establezca en foros internacionales sobre el desarrollo y utilización de sistemas de armas autónomos letales.

Para lograr que se consideren y adopten posiciones que tengan en cuenta los principios éticos en las futuras capacidades militares generadas entorno a la IA es necesario su alineación con las posturas defendidas por nuestro país en los foros internacionales, en que los aspectos éticos de los sistemas de armas, entre otros, son tratados y considerados. En este sentido, España no ha definido todavía una posición individual, manteniendo las defendidas dentro de la Unión Europea que tampoco son, de momento, concluyentes.

La guerra continúa teniendo la misma naturaleza y sus efectos son destrucción y muerte, por lo que el uso de la fuerza en las sociedades moralmente avanzadas, como la nuestra, y dotadas de la más moderna tecnología exige que responda a criterios de legalidad (Derecho) y de legitimidad (ética).

Ante la pregunta clave que hemos planteado más arriba sobre si dejamos a la IA y a las máquinas la responsabilidad exclusiva de tomar decisiones sobre vida y muerte, la opinión de quién esto escribe es que siempre debe haber un «control humano significativo» en el uso de sistemas de armas autónomos letales Es decir, siempre debe haber una trazabilidad y capacidad de atribución a un ser humano en las decisiones regidas por la IA que impliquen el uso de la fuerza letal, utilización que deberá inspirarse en los Derechos Humanos y el principio de humanidad.

Por otra parte, considero absolutamente necesario llegar a un acuerdo internacional para evitar, en caso negativo, una escalada en el desarrollo e implementación de los sistemas de armas autónomos letales. Ese escalamiento podría llevarnos a una doctrina similar a la de la «destrucción mutua asegurada», inaceptable desde posiciones éticas a las que habitualmente las relaciones internacionales, basadas en el realismo político de los intereses, nunca han sido muy sensibles y que haría inviable el cumplimiento del Derecho Internacional Humanitario.

Si no es así, si se deja que sean los robots regidos por la IA quién tengan plena autonomía en las decisiones sobre el uso de la fuerza sin tener en cuenta las consideraciones éticas, quizás podría cambiar la naturaleza de la guerra y ya no sería un fenómeno social humano, sino un fenómeno entre máquinas, escapando a cualquier control del ser humano y poniendo en peligro, a lo peor, a nuestra propia especie.

Desafíos éticos en la aplicación de la inteligencia artificial a los sistemas de defensa. Juan Moliner
Revista DIECISIETE. 2021. Nº 4. Madrid. ISSN 2695-4427 Versión digital (pp. 111-126) DOI: 10.36852/2695-4427_2021_04.06

REFERENCIAS BIBLIOGRÁFICAS

- Beijing AI Principles, 2019-5-28. https://www.baai.ac.cn/blog/beijing-ai-principles
- Comité Internacional de la Cruz Roja. https://www.icrc.org/spa/resources/documents/faq/autonomous-weapons...
- Estado Mayor de la Defensa (2019). Entorno operativo 2035. https://publicaciones.defensa.gob.es/entorno-operativo-2035-libros-papel.html
- European Commission. Ethics Guidelines for Thustworthy AI, https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines
- Fundación ESYS (2020). Seminario «Una Estrategia Nacional de Inteligencia Artificial que incorpore elementos claves de Seguridad», Madrid, 24 de septiembre. https://www.fundacionesys.com/en/noticias
- Geneva Internet Platform. Russian president approves national AI strategy, 10 Oct 2019, https://dig.watch/updates/russian-president-approves-national-ai-strategy
- Lezama, E. (2020). El mundo posible: los cambios que traerá la pandemia. Claves de Razón Práctica, núm. 272, septiembre-octubre 2020.
- López-Sánchez, M. (2017). *Some insights in Artificial Intelligence Autonomy in Military Technology*, https://ttac21.net/2017/11/10/autonomy-in-future-military-and-security-technologie
- Moliner González, J.A. (2018). Algunos problemas éticos de las tecnologías militares emergentes, Documento de opinión 16/2018, 19 de febrero de 2018, Madrid: Instituto Español de Estudios Estratégicos.
- Naciones Unidas, Convención sobre Ciertas Armas Convencionales https://undocs.org/pdf?symbol=es/CCW/GGE.1/2019/3
- Ortega Klein, A. (2020). Geopolítica de la ética en Inteligencia Artificial, Documento de trabajo 1/2020, 9 de enero de 2020, Madrid: Real Instituto Elcano.
- Peets, H., Hansem, M. y Maynard, P. (2019). Commisssion Expert Group on Liability for European Digital Technology, December, 2019, https://www.insideprivacy.com/artificial-intelligence/commission-expert-group-report
- Singer, P.W. (2009). *Wired for war. The robotics revolution and conflict in the 21st century.* Nueva York: The Penguin Press.
- Stop killer Robots. https://www.stopkillerrobots.org/learn/?lang=es
- Telefónica. Manifiesto por un Nuevo Pacto Digital, https://www.telefonica.com/manifiesto-digital/
- The White House, Artificial Intelligence for the American People. https://www.whitehouse.gov/ai/
- Travieso, J. (2015). Las consecuencias de mandar a la guerra a ´robots asesinos`, *ELDIARIO.es*, 17 de abril. https://www.eldiario.es/.../debate-torno-robots-asesinos_0_378312866.html
- U.S. Department of Defense. DOD Adopts Ethical Principles for Artificial Intelligence. https://www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adop
- Varios autores (2019). Usos militares de la inteligencia artificial, la automatización y la robótica (IAA&R). Documento de Trabajo 04/2019, Madrid: IEEE.
- Varios autores (2018). La inteligencia artificial aplicada a la Defensa. Documento de Trabajo 6/2018, Madrid: IEEE.