Criterios éticos y de derecho internacional humanitario en el uso de sistemas militares dotados de inteligencia artificial

Cómo citar este artículo [Chicago]: Cotino Hueso, Lorenzo y Gómez de Ágreda, Ángel. "Criterios éticos de derecho internacional humanitario en el uso de sistemas militares dotados de inteligencia artificial". *Novum Jus* 18, núm. 1 (2024): 249-283. https://doi.org/10.14718/NovumJus.2024.18.1.9

Lorenzo Cotino Hueso
Ángel Gómez de Ágreda

Código: 1355657756 • Autor: istockphoto.com

https://doi.org/10.14718/NovumJus.2024.18.1.9

Criterios éticos y de derecho internacional humanitario en el uso de sistemas militares dotados de inteligencia artificial*

Lorenzo Cotino Hueso**
Universidad de Valencia (Valencia, España)

Ángel Gómez de Ágreda***
Universidad Politécnica de Madrid (Madrid, España)

Recibido: septiembre 3 de 2023 **| Evaluado:** octubre 25 de 2023 **| Aceptado:** octubre 28 de 2023

Resumen

Los usos de inteligencia artificial (IA) y sistemas autónomos en defensa no cuentan con una regulación internacional específica y en las propuestas de regulación de IA son habituales las exclusiones normativas para defensa. Pese a ello, los sistemas militares dotados de IA deben cumplir con el derecho internacional. Su diseño debe respetar sus principios y los de la ética de la IA. Para ello, hay que tener en cuenta los elementos diferenciadores del ámbito militar, como la eficacia, criticidad de los resultados, protección y calidad de la información, así como los datos, su complejidad y dinamismo, el carácter dual de las tecnologías, y el potencial empleo por grupos u organizaciones terroristas o escalabilidad del uso. A partir de la experiencia comparada, se formulan los principios éticos de la inteligencia artificial en defensa, similares a los generales, pero con las particularidades del contexto militar. Se subraya especialmente la necesidad de control humano (transferencia limitada de autonomía, control humano significativo, responsabilidad en todo el ciclo de vida), ausencia de sesgos y robustez, especialmente frente a los *unintended engagements*, así como los principios de la fiabilidad, transparencia, trazabilidad y seguridad de los sistemas IA militares. Se exponen las especiales dificultades que se dan en el sector por la responsabilidad individual que rige el derecho internacional humanitario (DIH), o la dificultad de proyectar la "doctrina del doble efecto" a sistemas autónomos por la imprevisibilidad de estos.

Palabras clave: inteligencia artificial militar, sistemas letales de armas autónomas, defensa, derecho internacional humanitario

* El presente estudio es resultado de investigación de los siguientes proyectos: MICINN Proyecto "Derechos y garantías públicas frente a las decisiones automatizadas y el sesgo y discriminación algorítmicas" 2023-2025 (PID2022-136439OB-I00), financiado por MCIN/AEI/10.13039/501100011033/; Proyecto "Algoritmic law" (Prometeo/2021/009, 2021-24 Generalitat Valenciana); Convenio de Derechos Digitales-SEDIA Ámbito 5 (2023/C046/00228673) y Ámbito 6. (2023/C046/00229475); "Algorithmic Decisions and the Law: Opening the Black Box" (TED2021-131472A-I00). Estancia Generalitat Valenciana CIAEST/2022/1 y "Transición digital de las Administraciones públicas e inteligencia artificial" (TED2021-132191B-I00) del Plan de Recuperación, Transformación y Resiliencia; Cátedra ENIA "Derecho y regulación de la Inteligencia Artificial" U. Valencia-Cuatrecasas, 2024-2027.

** Doctor en derecho, Catedrático de Derecho Constitucional de la Universidad de Valencia (España). Valencian Graduate School and Research Network of Artificial Intelligence (Valgrai). Observatorio del Impacto Social y Ético de la Inteligencia Artificial (OdiseIA). ORCID: 0000-0003-2661-0010 http://www.researcherid.com/rid/H-3256-2015 scopus id 58308834300. Correo electrónico: cotino@uv.es

*** Ministerio de Defensa de España. Doctor en Ingeniería de Organización por la Universidad Politécnica de Madrid. Observatorio del Impacto Social y Ético de la Inteligencia Artificial (OdiseIA). ORCID: 0000-0003-1036-6324. Correo electrónico: a.gdeagreda@alumnos.upm.es y agdeagreda@gmail.com.

https://doi.org/10.14718/NovumJus.2024.18.1.9

Ethical and International Humanitarian Law Criteria in the Use of Artificial Intelligence-Powered Military Systems

Lorenzo Cotino Hueso
Universidad de Valencia (Valencia, España)

Ángel Gómez de Ágreda
Universidad Politécnica de Madrid (Madrid, España)

Received: September 03, 2023 **| Evaluated:** October 25, 2023 **| Accepted:** October 28, 2023

Abstract

The uses of artificial intelligence (AI) and lethal autonomous weapon systems (LAWS) have no specific international regulation and proposals for general AI regulation usually decline to engage in discussing them. Nevertheless, military AI systems must comply and integrate into their design the applicable International Humanitarian Law (IHL) and the principles of AI ethics. This should consider military distinctive elements such as effectiveness, criticality of the results, protection and quality of information and data, complexity and dynamism, dual nature of technologies, potential use by terrorist groups or organizations or scalability of use. Based on comparative experience, the authors formulate ethical principles of artificial intelligence in defense, in line with general ones, but with due regard for the particularities of the military context. Several topics are particularly emphasized, such as the need for human control (limited transfer of autonomy, meaningful human control, accountability throughout the life cycle), absence of bias and robustness, especially against unintended engagements, as well as the principles of reliability, transparency, traceability, and security of military AI systems. Specific aspects concerning the sector are discussed such as the individual responsibility that governs IHL, the difficulty of projecting the "doctrine of double effect" to autonomous systems or the unpredictability of these systems.

Keywords: military artificial intelligence, lethal autonomous weapon systems, defense, international humanitarian law.

Los usos de IA y sistemas autónomos en defensa

Un informe del Consejo de Seguridad de la ONU describe el primer uso registrado de Sistemas Letales de Armas Autónomas (en adelante LAWS, por sus siglas en inglés) en un teatro de combate en Libia en la primavera de 2020 y en 2021[1]. El sistema tenía capacidades de reconocimiento automático de objetivos que, según la información del momento, el fabricante estaba ampliando para incluir el reconocimiento facial. Posteriormente, se informó ampliamente sobre el uso de municiones merodeadoras en Ucrania desde 2022.

La inteligencia artificial (IA) es una disciplina relativamente joven, todavía en evolución y solo parcialmente comprendida por los legisladores[2]. El potencial de la IA en defensa va mucho más allá de los LAWS y, entre otros, incluye los sistemas diseñados para apoyo a la toma de decisiones. Sin embargo, los LAWS atraen especialmente la atención por su letalidad y la emotividad que generan[3]. Incluso a bordo de las plataformas militares, las aplicaciones de la IA son muy heterogéneas, y con una tendencia a permear todas las funciones. Como ejemplo de usos, que puede ser generalizable más allá del aire, Airbus plantea una serie de funciones de combate que desarrollará el futuro sistema de combate aéreo (FCAS, por sus siglas en inglés, *Future Combat Air System*)[4] producido conjuntamente por Francia, Alemania y España:

- planificación y ejecución de misiones militares;
- detección, reconocimiento e identificación de objetivos potenciales;
- conocimiento de la situación táctica para la toma de decisiones;
- guiado, navegación y control de vuelo;
- redes neuronales para la evaluación de amenazas y análisis de puntería;
- técnicas de IA para detectar anomalías y actividades adversas en el ciberespacio;

---

1 "Letter dated 8 March 2021 from the Panel of Experts on Libya Established pursuant to Resolution 1973 (2011) addressed to the President of the Security Council", *Naciones Unidas, Biblioteca digital*. https://digitallibrary.un.org/record/3905159?ln=es

2 Naciones Unidas, Institute for Disarmament Research, "The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches, a primer", *UNIDIR Resources*, núm. 6 (2017).

3 Michael C. Horowitz, "Public opinion and the politics of the killer robots debate", *Research and Politics* 3, núm. 1 (2016).

4 "Airbus Defense White paper: The Responsible Use of Artificial Intelligence in FCAS – An Initial Assessment", *Airbus, Fraunhofer FKIE*, 2020. https://www.fcas-forum.eu/articles/responsible-use-of-artificial-intelligence-in-fcas

- formación de operadores de sistemas y
- análisis de *big data* en producción, mantenimiento y logística para la reducción del coste del ciclo de vida, desde su fase de diseño hasta su baja en servicio.

Los sistemas IA se utilizan para el mantenimiento y el abastecimiento logísticos, la ayuda y la asistencia médica en el campo de batalla, la inteligencia, la vigilancia y el reconocimiento, y para las operaciones humanitarias y de socorro. Se puede proponer una taxonomía de las aplicaciones de la IA en el ámbito de la defensa en analogía con las funciones de un Estado Mayor:

1. Gestión del personal: desde el ámbito logístico, la gestión de carrera, la atención médica y psicológica, el apoyo a familiares y veteranos, etc.
2. Gestión de inteligencia: acceso, procesamiento y diseminación de inteligencia, fusión de análisis, etc.
3. Operaciones: consciencia situacional del mando y las unidades, gestión de fuegos, sincronización de acciones, etc.
4. Logística: gestión del abastecimiento, manuales inteligentes de mantenimiento, diseño y fabricación aditiva de piezas, etc.
5. Planeamiento: simulación y emulación de escenarios y plataformas, apoyo a la coordinación y gestión de situaciones, etc.
6. Comunicaciones: con numerosas aplicaciones ya en uso para la gestión de comunicaciones seguras, ciberseguridad, etc.
7. Instrucción y adiestramiento: con tutores digitales y simuladores.
8. Comunicación pública: con *bots*, detección de patrones y desinformación, etc.

Pues bien, cabe señalar que, más allá de los LAWS, estos otros usos de IA obviamente tienen que cumplir con los principios específicos del derecho aplicables exclusivamente a la guerra. No obstante, solo son más discutibles que otros usos de IA por la criticidad del contexto bélico en que se emplean.

El Comité Internacional de la Cruz Roja (ICRC, por sus siglas en inglés) se ha pronunciado repetidamente sobre el impacto del uso de la IA en la guerra[5] y afirma que

[5] Así, en International Committee of the Red Cross (ICRC), "Views of the ICRC on autonomous weapon systems", *ICRC*, 11 abril 2016. Son muchos los documentos clave desde el ICRC: Vicent Boulanin Neil

> el uso de AWS introduce un aumento significativo del riesgo para las personas afectadas por un conflicto armado, al socavar la protección de los civiles, desafiar el estado de derecho y generar inquietudes en virtud de los principios de humanidad.[6]

Del mismo sentir son numerosos Estados[7] y académicos[8]. En los mismos foros internacionales se ha reclamado la prohibición absoluta de los "robots asesinos"[9] junto con otras prohibiciones y pronunciamientos a favor de regulaciones más específicas[10]. La academia[11] y la misma industria vienen reclamando una normativa clara que les permita incorporar principios y directrices éticos a los sistemas[12], o ha desarrollado sus propios códigos éticos como guía de actuación y como herramienta de comunicación corporativa[13]. Diversos autores, en la línea de propuestas de este estudio, consideran que la prohibición del uso de IA en los sistemas de armas no es realista dado el carácter dual de estas tecnologías. Ahora bien, es preciso alcanzar un acuerdo de mínimos sobre su empleo ético y sobre la subordinación de su uso a los preceptos del Derecho Internacional Humanitario (DIH)[14].

Davison, Netta Goussac y Moa Peldán Carlsson, *Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control* (Ginebra: ICRC y Stockholm International Peace Research Institute, 2020), 21-25; ICRC, "Statement of the ICRC to the UN CCW GGE on Lethal Autonomous Weapons Systems", 21-25 de septiembre de 2020, Ginebra.

6 ICRC, "ICRC Position on Autonomous Weapon Systems & Background Paper", *ICRC*, 12 de mayo de 2021.

7 Así se recuerda desde diversos países como Australia por ejemplo, en Kate Devitt, Michael Gan, Jason Scholz y Robert Bolia, *A Method for Ethical AI in Defence*. Australia: Departamento de defensa, 2020. Publicación número DSTG-TR-3786; o desde EEUU, en Defense Innovation Board, "AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense", *Department of Defense USA (DoD), Supporting Document*, octubre 31 de 2019.

8 Además de los que aquí se siguen, son muchos los estudios que por límites de extensión no es posible citar ahora, como los de Keith Abney, Daniel Araya, Peter Asaro, Alexander Blanchard, Thompson Chengeta, Timothy J. Demy, Jai Galliott, Maziar Homayounnejad, Meg King, Alfonso López-Casamayor, Thomas W. Simpson o Mariarosaria Taddeo, entre otros.

9 "Campaign to Stop Killer Robots", *Stop Killer Robots*, 2018. https://www.stopkillerrobots.org/

10 Además de los documentos ICRC, resulta básico el de Human Rights Watch: "An Agenda for Action Alternative Processes for Negotiating a Killer Robots Treaty", *Human Rights Watch*, 19 de noviembre de 2022.

11 Responsible AI in the Military domain Summit (REAIM), "REAIM Call to Action", *Gobierno de Países Bajos*, 16 de febrero de 2023.

12 Airbus, "Airbus Defense White paper".

13 DeepMind, "Deepmind ethics and society principles", 2017, https://ai.google/responsibility/principles/; International Business Machines (IBM), "Foundation models: Opportunities, risks and mitigations", julio de 2023, https://www.ibm.com/downloads/cas/E5KE5KRZ; Sundar Pichai, "AI, Google: our principles" (Blog), 7 de junio de 2018, https://blog.google/technology/ai/ai-principles/, y Microsoft. "AI Principles", 2019, https://www.microsoft.com/en-us/ai/principles-and-approach, directamente implicados por su modelos de negocio, estuvieron entre los primeros en hacerlo.

14 Jorge Ulloa Plaza y María A. Benavides Casals, "Moralidad, guerra y derecho internacional. Tres cuerdas para un mismo trompo: la humanidad". *Novum Jus* 17, núm. 1 (2023): 259-282.

Sin perjuicio de los impactos más negativos que imponen una regulación, aunque sea de mínimos, quienes suscriben también consideramos que los sistemas IA y autónomos militares pueden ser beneficiosos para un mejor cumplimiento ético y normativo y un mecanismo para cumplir "la obligación de tomar precauciones viables para reducir el riesgo de daños a la población civil y a otras personas u objetos protegidos"[15]. Estos sistemas inteligentes pueden reducir el riesgo para los combatientes, podrían reducir bajas civiles si tienen capacidad de diferenciar, reducen también los errores humanos debidos a fatiga, falta de comunicaciones, o la emociones como el miedo o la venganza. Es por ello que cabe plantearse el deber, y no sólo la posibilidad, de desarrollar y utilizar estos sistemas. Así, si un Estado puede minimizar los daños propios o ajenos con el empleo de sistemas inteligentes, ¿tiene la obligación de desarrollarlos y emplearlos?[16]

Especialmente en el actual contexto geopolítico, es evidente la relevancia de definir la aplicabilidad y especificidad de los principios éticos aplicables a la IA general al entorno de los conflictos bélicos. Es también urgente, porque estos sistemas ya están siendo empleados en el campo de batalla y han de ser acordes a los principios éticos cásicos al tiempo de adecuarse al derecho, en especial, al derecho humanitario.

El presente artículo pretende contribuir al debate sobre el uso ético de los sistemas dotados de IA en la guerra. Se plantean cuestiones vinculadas entre sí: ¿son aplicables los principios generales de la ética de la IA a los sistemas militares? ¿Cómo debe integrarse el derecho humanitario actual en los sistemas militares? ¿Qué especificidades hay que tener en cuenta en dicha aplicación e integración?

Para ello, se parte de un estudio previo de la multiplicidad de códigos éticos generados desde 2016, así como de la regulación jurídica existente. Se efectúa un estudio analítico y sintético de los mismos en relación con las funciones que tienen que cumplir estos sistemas en el campo de batalla. La mayoría de los códigos éticos generados ignoran las aplicaciones militares y ello obliga a decantar los elementos comunes a todos ellos en el contexto de la defensa. Por cuanto al derecho humanitario, además de analizar los intentos de una nueva regulación

[15] Estados Unidos, Departamento de Defensa, "AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense Supporting Document Defense Innovation Board", 54. https://media.defense.gov/2019/Oct/31/2002204459/-1/-1/0/DIB_AI_PRINCIPLES_SUPPORTING_DOCUMENT.PDF.

[16] Andrew Williams y Paul D. Scharre, *Autonomous Systems: Issues for Defence Policymakers* (Norfolk: Nato Communications and Information Agency, 2015).

para sistemas autónomos, se apuesta por la integración del derecho existente en el diseño de los sistemas militares.

En el estudio se pretende evitar una aproximación estrictamente restrictiva, para incorporar también aspectos en los que la IA puede contribuir a aliviar el sufrimiento humano asociado a cualquier situación bélica. De igual modo, dadas las limitaciones del derecho para el marco de la defensa, se considera especialmente adecuada tanto la metodología jurídico-normativa como la de la ética aplicada.

Elementos diferenciadores del ámbito militar a tener en cuenta respecto del uso de inteligencia artificial

La guerra es un acto político y mantiene lo que se denomina un "continuo de las operaciones" que impide establecer límites claros con los periodos de paz. La escalada suele ser un proceso progresivo que transita por la llamada "zona gris"[17]. En ella las acciones ofensivas se intensifican, pero se intentan mantener por debajo del umbral de fuerza que provoque una respuesta armada. El establecimiento de estándares éticos en el ámbito militar conlleva una serie de beneficios, tanto a nivel interno de las fuerzas armadas, como para el conjunto de la sociedad y de la tecnología:

- Imprime *legitimidad* al empleo de sistemas dotados de IA y fomenta su uso responsable,
- los estándares aplicables son fácilmente extrapolables a *otras tecnologías* militares emergentes,
- genera *confianza*, tanto en la población propia como en los países afines,
- contribuye a la *interoperabilidad* de los sistemas, basados en principios de empleo comunes,
- *agiliza el proceso de adquisición y favorece la* I+D, proporcionando sistemas más efectivos.

Existen una serie de elementos propios del ámbito militar que lo diferencian del contexto civil y deben ser tenidos también en cuenta respecto del uso de los sistemas dotados de IA en el ámbito militar, que se enuncian a continuación.

[17] Michael Schmitt, "Grey Zones in the International Law of Cyberspace", *The Yale Journal of International Law Online* 42, núm. 2 (2016): 1-21.

Orientación a la eficacia y la misión

La eficacia en el cumplimiento de la misión se impone por criterios de seguridad nacional. La victoria es irrenunciable y se alcanza solo por la consecución de los objetivos estratégicos. Las consideraciones de eficiencia y rentabilidad del entorno civil ocupan un lugar muy marginal en el entorno de la guerra.

Criticidad de los resultados

Esta primacía en el cumplimiento de la misión responde a las consecuencias que se derivan del fracaso. Las operaciones militares pueden implicar la pérdida de vidas humanas, incluso cuando no involucran acciones cinéticas. El estrés derivado del alto *tempo* de las operaciones puede conducir a una excesiva confianza en los sistemas de armas disponibles. La intoxicación de los datos sobre los que operan dichos sistemas, o los propios sesgos de su programación, tiene el potencial de generar resultados no deseados.

Protección y calidad de la información y los datos

La información y la confianza son vitales en las operaciones militares. Durante estas se recopilan datos críticos sobre el entorno y las personas (del bando propio, a fin de generar confianza mediante las habilitaciones personales de seguridad, y del contrario). De la veracidad y oportunidad de los datos depende el éxito de las operaciones. Sin embargo, su retención, empleo o difusión fuera de estas puede resultar en un grave daño para sus propietarios. La protección de dichos datos frente al enemigo es, además, fundamental para la seguridad de las propias fuerzas y requiere de un proceso de concienciación que no está presente en la vida civil[18]. Este riesgo es tanto mayor cuanto más integradas se encuentren la inteligencia humana y la artificial (i.e., en casos de inteligencia aumentada) y más datos compartan ambas.

Complejidad y dinamismo

La *niebla de la guerra* o la *fricción* de Clausewitz hacen referencia a la complejidad e incertidumbre de los escenarios y situaciones de la guerra. A la propia complejidad de una actividad política en la que se ponen en juego todos los recursos del Estado,

[18] Dimitris Potoglou et ál., "The value of personal information online: Results from three stated preference discrete choice experiments in the UK" (conferencia presentada en el 21st European Conference for Information Systems ORCA, Utrech, 5-8 de junio de 2013).

hay que añadir la acción adversaria del enemigo que pretende desbaratar nuestros medios y planes, y un normalmente elevado *tempo* de las operaciones necesario para obtener la ventaja estratégica.

Los escenarios no se limitan tampoco a entornos predeterminados y predecibles (como carreteras), sino que se amplían a terrenos y circunstancias no preparados para alcanzar la sorpresa. La guerra contemporánea ha ampliado estos horizontes para convertirse en *multidominio*. En ella, además de los tradicionales entornos terrestre, naval y aéreo, es preciso tomar en consideración otros como el electromagnético, el cibernético y el cognitivo.

Afectación asimétrica de la tecnología

La complejidad de los escenarios lleva asociado que los efectos de la tecnología se apliquen trasversalmente sobre civiles y militares y, a lo largo del tiempo, sobre militares en activo y en situación de reserva. Ello plantea el desafío de discriminar el uso y conservación de los datos de unos y otros en función de los derechos que les amparan. El principio de distinción adquiere, en el ámbito digital, una dimensión temporal evolutiva añadida.

Carácter dual de las tecnologías

Las aplicaciones dotadas de IA rara vez son desarrolladas completamente *ex novo*, al incluir los datos sobre los que son entrenadas u operadas. Esto genera un doble riesgo:

- la incorporación de componentes de uso general con bajos estándares de seguridad, y
- la filtración de componentes militares susceptibles de ser incorporados a aplicaciones civiles, dotadas así de capacidades de grado militar.

En el primer caso, el empleo de capas adicionales de seguridad será preceptivo para garantizar el funcionamiento no comprometido del *software*.

Potencial empleo por grupos u organizaciones terroristas

El umbral tecnológico y económico de acceso a muchas aplicaciones que emplean IA es relativamente bajo. La filtración de conocimientos o procedimientos a estos

grupos, o su desarrollo autóctono pueden poner en sus manos herramientas con una alta capacidad de generar efectos militares. Es preciso garantizar la protección de estos sistemas y, con independencia de que no se desarrollen para uso propio, mantener un alto grado de preparación frente a su potencial uso por otros actores.

Grado de autonomía permisible

El grado de autonomía que se puede permitir a un sistema algorítmico variará en función de su capacidad para "entender" el contexto y aplicar los preceptos de DIH y del derecho de los conflictos armados. El grado de autonomía será variable, por tanto, en función de lo evolucionado que esté el sistema y de lo complejo que sea el escenario en que desarrolla su actividad.

Escalabilidad del uso

El control político sobre el nivel de fuerza a emplear se materializa a través de las reglas de enfrentamiento (RoE, del inglés *Rules of Engagement*) que, en función de la situación, lo amplían o limitan. El potencial de los sistemas autónomos para provocar escaladas de violencia no intencionadas requiere de una estructura de mando y control capaz de garantizar que prevalezca la visión del comandante.

Asimetría en la concepción de legalidad

El diálogo permanente entre todos los actores implicados (diseñadores, desarrolladores, comercializadores y usuarios) favorecerá una comprensión común del modo de aplicar el DIH a los sistemas dotados de IA. Esto, a su vez, sentará las bases para una referencia común para su cumplimiento.

La falta de una regulación específica, pese a los intentos, y las exclusiones normativas para defensa

Ni los sistemas de armas autónomos, ni los sistemas IA utilizados en defensa están específicamente regulados. Quizá lo más cercano hoy por hoy sean las "recomendaciones" o "principios" en los EEUU[19]. Desde la UE se va a dar la primera

[19] Estados Unidos, Departamento de Defensa. "The Department of Defense AI Ethical Principles, The Joint Artificial Intelligence Center", 24 de febrero de 2020, https://www.ai.mil/blog_02_24_20-dod-ai_principles.html

regulación general en el mundo con el futuro Reglamento (*Ley de IA*)[20]. Ahí se impondrán muy importantes garantías y requisitos para los llamados sistemas de "alto riesgo", como podrían serlo muchos de los sistemas de IA en defensa. Sin embargo, la futura norma "no se aplicará a los sistemas de IA desarrollados o utilizados exclusivamente con fines militares" o, en la versión del Consejo de la UE en 2022, no se aplicará "en cualquier caso, a actividades militares, de defensa o de seguridad nacional, con independencia del tipo de entidad que lleve a cabo dichas actividades" (art. 2.3º). La normativa de protección de datos, que muchas veces se aplica al ámbito de la IA, también está excluida de la aplicación para el ámbito de defensa y militar[21]. El borrador de la Convención sobre IA y derechos humanos del Consejo de Europa (julio de 2023), aunque no excluye de su aplicación, sí hace referencia a las "restricciones, derogaciones o excepciones (…) en relación con la protección de la seguridad nacional, la defensa, la seguridad pública" (Capítulo III)[22]. Tampoco hay regulación específica de la IA y de LAWS en los tratados de DIH. El ICRC propone normas jurídicamente vinculantes: restringir su uso a objetivos militares por naturaleza, o a situaciones en las que no estén presentes personas civiles u objetos de carácter civil; establecer obligaciones tanto respecto de la duración, el alcance geográfico y la escala de uso, así como deberes de evaluación en relación con un ataque específico, y garantizar una supervisión humana eficaz[23]. La *Convention on Conventional Weapons* (CCW), entre los años 2014-2019, aunque no llegó a consensuar ningún documento ejecutivo, dio lugar a la consolidación de diez principios éticos de la IA en defensa que luego se mencionarán. Ha habido avances significativos en la CCW de 2022[24]. Y en 2023 muchos países, incluso de los involucrados en estos desarrollos tecnológicos y militares, coinciden en

[20] Inicialmente, Comisión Europea, "Propuesta de Reglamento por el que se establecen normas armonizadas en materia de inteligencia artificial", *EUR-Lex Access to European Union Law*, 21 de abril de 2021. Esta propuesta tiene probable aprobación para inicios de 2024.

[21] Cabe seguir el artículo 2a del Reglamento de protección de datos de la UE (RGPD) que excluye de su aplicación "el ejercicio de una actividad no comprendida en el ámbito de aplicación del Derecho de la Unión" y, en este sentido, el Considerando 4 que afirma como ejemplo de ello "las actividades relativas a la seguridad nacional. Tampoco […] la política exterior y de seguridad común de la Unión". La normativa nacional conlleva exclusiones similares.

[22] Consejo de Europa, Comité sobre inteligencia artificial (CAI), "Borrador de trabajo consolidado del convenio marco sobre inteligencia artificial, derechos humanos, democracia y Estado de derecho", *Council of Europe (COE)*, Estrasburgo, 7 de julio de 2023.

[23] Especialmente cabe tener en cuenta Boulanin et ál., *Limits on Autonomy in Weapon Systems*.

[24] Naciones Unidas, "Reunión de las Altas Partes Contratantes de la Convención sobre prohibiciones o restricciones del uso de ciertas armas convencionales que puede considerarse excesivamente perjudicial o tener efectos indiscriminados", Reporte Final, UN Document CCW/MSP/2022/7, Ginebra, 16-18 de noviembre de 2022.

garantizar el control humano con obligaciones concretas [25] y prohibir los sistemas que no permitan un control humano suficiente. Llama la atención la propuesta desde EEUU (con Corea, Australia, etc.)[26] que afirma la necesidad de limitar los tipos de objetivos que los sistemas pueden atacar, así como la "duración, el alcance geográfico y la escala de la operación" de un sistema de armas, y tomar medidas para garantizar la supervisión por los operadores.

No obstante, se discute qué sistemas y usos ya están prohibidos *de facto*. Y, sobre todo, las diferencias parecen insalvables por la forma del acuerdo y su obligatoriedad. Estos desacuerdos obligan más si cabe que las obligaciones se deduzcan a partir del derecho existente ante la improbabilidad de un tratado o acuerdo vinculante.

Ochenta y seis países se han pronunciado a favor del desarrollo de un instrumento legalmente vinculante para regular los sistemas de armas autónomos[27]. Sin embargo, once Estados se han pronunciado en contra de una respuesta legal internacional, así como treinta y seis no han adoptado una posición clara[28]. EEUU, Reino Unido y otros han rechazado explícitamente negociar un instrumento legalmente vinculante. Francia, Alemania y otros[29] sugieren que los estándares deben ser implementados por los Estados a nivel nacional, pero sin tratado internacional. Otros países como India, Israel y Rusia rechazan el desarrollo de respuestas legales o de cualquier otro tipo. Desde la campaña "Multinational Capability Development Campaign" (MCDC), del Mando Aliado para la Transformación de la OTAN (NATO ACT, esta última sigla del inglés *Allied Command Transformation*), se afirmaba que no es necesario un tratado específico sobre sistemas con capacidades autónomas, pero que puede serlo en el futuro en función de su desarrollo.

[25] Sobre el tema, de especial interés, Elizabeth Minor, "Laws for LAWS. Towards a treaty to regulate lethal, autonomous weapons", *Friedrich Ebert Stiftung New York Analysis*, Febrero de 2023. https://library.fes.de/pdf-files/international/20013.pdf , así como ICRC, "ICRC Position on Autonomous".

[26] Australia, Canadá, Japón, República de Corea, Reino Unido y Estados Unidos, "Principles and Good Practices on Emerging Technologies in the Area of LAWS", *reachingcriticalwill.org*, 7 de marzo de 2022.

[27] Por ejemplo, Argentina, Costa Rica, Guatemala, Kazakhstan, Nigeria, Panamá, Filipinas, Sierra Leona, Palestina y Uruguay, "Proposal: Roadmap Towards New Protocol on Autonomous Weapons Systems", *reachingcriticalwill.org*, 13 de marzo de 2022.

[28] "State positions", *Automated Decision Research*, https://automatedresearch.org/state-positions/

[29] Finlandia, Francia, Alemania, Países Bajos, Noruega, España y Suecia, "Working paper submitted to the 2022 Chair of GGE on LAWS", 13 de julio de 2022, https://documents.unoda.org/wp-content/uploads/2022/07/WP-LAWS_DE-ES-FI-FR-NL-NO-SE.pdf

El derecho humanitario actual debe integrarse en los sistemas IA y autónomos

Pese a la falta de regulación específica, los sistemas IA en defensa habrán de cumplir el derecho ya existente que sea aplicable. Cabe recordar[30] que el artículo 36 sobre "Armas nuevas" dispone que:

> cuando una Alta Parte contratante estudie, desarrolle, adquiera o adopte una nueva arma, o nuevos medios o métodos de guerra, tendrá la obligación de determinar si su empleo, en ciertas condiciones o en todas las circunstancias, estaría prohibido por el presente Protocolo o por cualquier otra norma de derecho internacional aplicable a esa Alta Parte contratante[31].

Así, las normas del Derecho de Guerra pueden "regular los usos de la IA en los conflictos armados" y "pueden aplicarse cuando las nuevas tecnologías, como la IA, se utilizan en conflictos armados"[32].

Según se ha expuesto, pese a que no haya regulación específica debe cumplirse el derecho vigente. Es por ello por lo que, a continuación, se hace una proyección para el uso de IA y sistemas autónomos de las 161 normas o reglas del DIH consuetudinario identificadas desde el ICRC[33]. Así, se presta atención esencialmente al "principio de distinción" entre civiles y combatientes, la prohibición de "ataques indiscriminados, la proporcionalidad y sus derivaciones", las reglas relativas a "personas y objetos específicamente protegidos", a "métodos específicos de guerra" y "armas" y, finalmente, al "trato debido a las personas civiles o fuera de combate".

Sin duda, en todo el ciclo de vida del sistema de IA debe proyectarse "el principio de distinción" (1ª Sección): "distinguir en todo momento entre civiles y combatientes. Los ataques sólo pueden dirigirse contra los combatientes" (norma 1), con todas sus derivaciones, concreciones y precauciones en los capítulos 1 al 6. El sistema debe integrar la definición de objetivos militares de la regla 8. No obstante, el sistema IA

30 Williams y Scharre, *Autonomous Systems*.

31 International Committee of the Red Cross (ICRC), "Protocolo I adicional a los Convenios de Ginebra de 1949 relativo a la protección de las víctimas de los conflictos armados internacionales", 18 de junio de 1977.

32 Estados Unidos, Departamento de Defensa, *AI Principles*, Anexo III.

33 Las normas citadas en los siguientes párrafos provienen de la "Base de datos sobre DIH consuetudinario", consolidada por International Committee of the Red Cross (ICRC).

podrá diseñarse para supuestos en los que los bienes civiles se utilizan para fines militares y, por ello, pierdan su protección (norma 10).

También hay que proyectar a los sistemas IA la prohibición de "ataques indiscriminados" (capítulo 3, norma 11). No será sencillo determinar cuándo "emplean un método o medio de combate que no puede ser dirigido a un objetivo militar específico" (b) o "cuyos efectos no pueden ser limitados como lo requiere el derecho internacional humanitario" (c). Habrá que integrar en el sistema la regla 13 sobre "bombardeos de zona", en los que los objetivos militares se mezclan con la presencia de civiles o bienes de carácter civil. Cada objetivo militar dentro de esa área debe ser identificado y atacado de manera separada, con todas las precauciones necesarias para minimizar el daño a los civiles y a los bienes civiles (norma 13).

La proporcionalidad es un principio esencial que hay integrar en el sistema inteligente (capítulo 4): "los daños y sufrimientos causados a la población civil y a los bienes de carácter civil no han de ser excesivos en relación con la ventaja militar concreta y directa que se espera obtener" (en especial, norma 14). Son muchas las proyecciones de la proporcionalidad respecto de las "precauciones en el ataque" (regla 15); verificación de los objetivos (norma 16); elección de los métodos y medios de guerra (norma 17); evaluación de los efectos de los ataques (norma 18); control durante los ataques (norma 19); aviso con debida antelación (norma 20); elección de los objetivos (norma 21). La conexión de la proporcionalidad con el principio de distinción es clara. No se puede ignorar la especial dificultad de que el sistema IA integre la doctrina del doble efecto y valore si las ventajas militares son excesivas o no respecto de la muerte incidental de civiles (norma 14 o 18)[34]. El sistema IA debe permitir que en cualquier caso se pueda suspender o anular un ataque, ello es obligatorio en cuanto se advierta que el objetivo no es militar o que se causará desproporcionalidad (norma 19)[35]. La proporcionalidad también se proyecta en la cuarta sección, titulada "Armas" (capítulo 20): "queda prohibido el empleo de medios y métodos de guerra de tal índole que causen males superfluos o sufrimientos innecesarios" (norma 70) o armas cuyos "efectos sean indiscriminados" (norma 71).

[34] Es obligatorio "evaluar si el ataque causará incidentalmente muertos o heridos entre la población civil, daños a bienes de carácter civil o ambas cosas, que sean excesivos en relación con la ventaja militar concreta y directa prevista". ICRC, "Base de datos", Normas.

[35] "Hacer todo lo que sea factible para suspender o anular un ataque si se advierte que el objetivo no es militar o si es de prever que el ataque cause incidentalmente muertos o heridos entre la población civil, daños a bienes de carácter civil o ambas cosas, que sean excesivos en relación con la ventaja militar concreta directa y prevista". ICRC, "Base de datos", Normas.

El sistema IA militar habrá de diseñarse para cumplir con la sección segunda sobre "personas y objetos específicamente protegidos" y supervisar la protección específica del personal y objetos médicos y religiosos, de ayuda humanitaria o del personal que participa en misiones de mantenimiento de la paz. También se regula la protección de los bienes culturales, las obras que contienen fuerzas peligrosas, las zonas protegidas o la protección del entorno natural. Las estratagemas de guerra quedan permitidas (norma 57), lo cual puede incluir tácticas de desinformación. Ello puede tener particular proyección para el uso de IA en defensa y militar.

Obviamente, habrán de cumplirse las normas sobre "trato debido a las personas civiles o fuera de combate" (5ª sección). Entre ellas, la prohibición de discriminación (norma 88), de actos de tortura, tratos crueles, inhumanos o degradantes (norma 90). Los sistemas IA pueden ser útiles para la detección y localización de personas heridas y para el cumplimiento de las normas del capítulo 34 sobre "heridos, enfermos y náufragos" (ver normas 109-111).

Unos principios éticos de la inteligencia artificial para su uso militar y de defensa

Referentes éticos de la IA general, de los sistemas autónomos y de la IA militar

Entre decenas de declaraciones y documentos[36], la Declaración del Parlamento UE de 2017[37] sobre robótica es buena expresión de lo que constituyen los principios éticos esenciales de la IA[38]. Destacan sin duda las directrices éticas para una IA fiable desde la UE[39], la "Recomendación del Consejo sobre Inteligencia Artificial de la OCDE" de 22 de mayo de 2019[40] y la "Recomendación sobre la ética de la inteligencia

[36] Lorenzo Cotino Hueso, "Ética en el diseño para el desarrollo de una inteligencia artificial, robótica y *big data* confiables y su utilidad desde el derecho", *Revista Catalana de Derecho Público*, núm. 58 (2019).

[37] Unión Europea, Parlamento Europeo, "Normas de Derecho civil sobre robótica", Resolución del 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL)).

[38] Swaroop Chakraborty, "Inteligencia artificial y derechos humanos: ¿son convergentes o paralelos entre sí?", *Novum Jus* 12, núm. 2 (2018): 13-38.

[39] Primera versión en 2018, Comisión Europea, Dirección General de Redes de Comunicación, Contenido y Tecnologías, "Directrices éticas para una IA fiable", Oficina de Publicaciones, 8 de abril de 2019 https://data.europa.eu/doi/10.2759/14078.

[40] Organización para la Cooperación y el Desarrollo Económicos (OCDE), "Recommendation of the Council on Artificial Intelligence", OECD/LEGAL/0449, OECD Legal Instruments, 2023.

artificial" de noviembre de 2021 de la UNESCO[41]. Ya en 2018, el proyecto *AI4People* contabilizó 47 principios éticos proclamados internacionalmente y los destiló en cinco puntos: beneficencia ("hacer el bien"), no maleficencia ("no hacer daño"), autonomía o acción humana ("human agency", "respeto por la autodeterminación y elección de los individuos") y justicia ("trato justo y equitativo para todos"). Estos cuatro principios básicos provienen del ámbito de la biomedicina desde los años 2001 y esencialmente se le ha añadido el principio de rendición de cuentas, explicabilidad y transparencia[42].

Desde Harvard[43] se analizaron más de treinta de las principales declaraciones internacionales y corporativas de ética de la IA y se sintetizaron en los siguientes temas clave o categorías a cumplir por la IA: *privacidad*, tanto en el uso como en la capacidad de decidir sobre sus datos; mecanismos para garantizar la *rendición de cuentas* por el uso de IA; necesidad de que los sistemas de IA sean seguros; *transparencia y explicabilidad* de los sistemas para permitir la supervisión y el control; maximización de la *equidad*, *no discriminación* y promover la inclusión; *control humano* de la tecnología, de modo que las decisiones importantes estén sujetas a revisión humana; *responsabilidad profesional* e integridad de las personas que participan en el desarrollo del sistema, y finalmente, que los *valores humanos* sirvan para orientar los fines a los que se dedica la IA, así como sus límites.

Pues bien, como se analiza, muchos de estos principios se proyectan con sus especialidades para el concreto ámbito de los LAWS y los sistemas IA en defensa. La CCW, entre los años 2014-2019, alumbró diez principios éticos para los LAWS que pueden considerarse como punto de partida de otros que, a nivel nacional o multinacional, se han publicado desde entonces.

41 Gabriela Ramos, "Ética de la inteligencia artificial", *Unesco. Inteligencia artificial*. https://www.unesco.org/es/artificial-intelligence/recommendation-ethics

42 Comisión Europea, "Directrices éticas", 10; Luciano Floridi et ál., "AI4People —An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations", *Minds and Machines* 28, núm 4 (2018): 699-700.

43 Jessica Fjeld et ál., "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI", Berkman Klein Center for Internet & Society Research at Harvard University, 2020.

1.	El DIH sigue siendo de aplicación a todos los sistemas de armas, incluyendo su desarrollo potencial y el uso de sistemas de armas autónomos letales.
2.	Se debe retener la responsabilidad humana por las decisiones en el uso de sistemas de armas ya que la accountability no puede transferirse a las máquinas. Este principio debería tenerse en cuenta a lo largo de todo el ciclo de vida del sistema de armas.
3.	Se debe garantizar la accountability por el desarrollo, despliegue y uso de cualquier sistema de armas emergente en el marco de la Convención para ciertas armas (CCW) de acuerdo con el Derecho Internacional aplicable, incluso cuando la operación de dichos sistemas de armas se produzca en el marco de una cadena de mando y control humana responsable.
4.	De acuerdo con las obligaciones de los estados bajo el Derecho Internacional, en el estudio, desarrollo, adquisición o adopción de un arma, medio o método de guerra nuevos, se debe determinar si su empleo estaría, en alguna o en todas las circunstancias, prohibido por el Derecho Internacional.
5.	Cuando se desarrollen o adquieran nuevos sistemas de armas o tecnologías emergentes en el área de los sistemas de armas autónomos letales, es preciso tomar en consideración la seguridad física, las salvaguardas no físicas (incluida la ciberseguridad frente a hackeos o suplantación de identidad), el riesgo de su adquisición por parte de grupos terroristas y el riesgo de proliferación.
6.	La evaluación de los riesgos y las medidas de mitigación deberían ser parte integral del ciclo de diseño, desarrollo, prueba y despliegue de tecnologías emergentes en cualquier sistema de armas.
7.	Se debe prestar la debida consideración al uso de tecnologías emergentes en el área de los sistemas de armas autónomos letales en cumplimiento de las obligaciones contempladas en el DIH y otras normas legales de aplicación.
8.	En la redacción de potenciales medidas de política relativas a los sistemas de armas autónomos letales, estos no deben dotarse de apariencia humana.
9.	Las discusiones sobre cualquier medida potencial adoptada en el contexto del CCW no debería limitar el progreso o el acceso a usos pacíficos de las tecnologías autónomas inteligentes.
10.	El CCW ofrece un marco adecuado para tratar el asunto de las tecnologías emergentes en el área de los sistemas de armas autónomos letales en el contexto de los objetivos y propósitos de la Convención, que persigue alcanzar el equilibrio entre el principio de necesidad militar y las consideraciones humanitarias

Figura 1. Principios sugeridos por el Grupo de Expertos Gubernamentales en Sistemas de Armas Autónomos Letales[44]. Traducción de los autores

[44] Naciones Unidas, Convention on Conventional Weapons (CCW), "Report of the 2018 Group of Governmental Experts on Lethal Autonomous Weapons Systems", 31 de agosto de 2018.

La legitimidad del documento se sustenta en la amplia base de participación y aportaciones que presenta. No solamente está avalado por Naciones Unidas, sino que incluye entre sus ponentes –además de a más de 90 naciones– actores clave en el DIH, como el ICRC y grupos de sociedad civil. Su foco se centra en dos criterios:

- la aplicabilidad de los principios de DIH (puntos 1, 4, 7 y 10), tanto para los desarrollos actuales como para los futuros, y
- el control humano sobre la parte algorítmica de estos sistemas (puntos 2 y 3), incluso en caso de fallo o de intrusión externa (punto 6).

También incorpora dos principios mucho más concretos:

- la prohibición de dotar de apariencia humana a los sistemas de armas autónomos (punto 8), y
- la necesidad de evitar que la regulación del armamento autónomo pueda ralentizar el desarrollo de aplicaciones de carácter civil o el acceso de la población a este (punto 9).

El carácter dual de muchas tecnologías de uso militar (o, como avisa el punto 5, el uso paramilitar) y su procedencia de desarrollos civiles convierten esta labor en algo mucho más complejo. Este problema se extiende a numerosas tecnologías, además de la IA[45].

Desde la primera formulación por Nadella en 2016[46] y la posterior redacción de los principios del CCW se ha producido una eclosión de códigos éticos sobre el uso de la IA[47], tanto a nivel gubernamental, como desde los ámbitos de la empresa, la academia o la sociedad civil. También se han elaborado algunos para tratar aspectos concretos de las aplicaciones de la IA a lo militar (como el caso de Francia en relación con el aumento artificial de capacidades humanas en combate)[48]. Un análisis de estos documentos refleja un alto grado de homogeneidad, con algunas diferencias que son fruto, principalmente, de la distinta nomenclatura empleada y del grado de especificidad. En la Tabla 1 se ha unificado la denominación de

[45] Tania Scalia et ál, "Final technical report. Study on the dual-use potential of Key Enabling Technologies (KETs)", *Comisión Europea, Agencia Ejecutiva para las Pequeñas y Medianas Empresas*, 13 de enero de 2017.

[46] Satya Nadella, "The Partnership of the Future. Microsoft's CEO explores how humans and A.I. can work together to solve society's greatest challenges", *Slate*, 28 de junio de 2016.

[47] Ángel Gómez de Ágreda, "Ethics of autonomous weapons systems and its applicability to any AI systems", *Telecommunications Policy* 44, núm. 6 (2020): 1-15.

[48] Francia, Defence Ethics Committee, "Opinion on the Augmented soldier", *Ministère des Armées*, 18 de septiembre de 2020.

aquellos principios que expresaban valores equivalentes para mostrar el grado de comunalidad entre todos los códigos.

Tabla 1. Principios éticos recogidos en los códigos de los principales países y organizaciones

Estados Unidos	Canadá	Francia	Australia	Reino Unido	Países Bajos	OTAN
Responsabilidad	Responsabilidad	Control y responsabilidad humanos	Responsabilidad	Responsabilidad	Control	Responsabilidad y *accountability*
Equidad	Equidad					
Trazabilidad	Privacidad, confidencialidad y seguridad	Documentación y trazabilidad	Trazabilidad	Comprensibilidad	Sobredependencia	Explicabilidad y trazabilidad
Fiabilidad	Fiabilidad	Fiabilidad		Fiabilidad	Error	Fiabilidad
Gobernabilidad		Juicio del operador	Gobernanza			Gobernabilidad
		Robustez y seguridad	Confianza	Mitigación de sesgos y daños	Confianza	Mitigación de riesgos
		Análisis de riesgos				
		Estandarización y certificación				
				Centralidad humana		
			Legalidad			Legalidad

Fuente: elaboración propia, a partir de los códigos respectivos.

La equidad, que puede resultar menos crítica en un ambiente bélico, solo se menciona en los códigos norteamericanos. La escasa mención de la sujeción a principios del DIH solo puede entenderse como una normatividad sobreentendida como parte de la actividad militar. La responsabilidad y la trazabilidad (o la explicabilidad, concepto más ambicioso) aparecen recogidos en todos los códigos, casi igual que la robustez y la fiabilidad. En su conjunto, los principios recogidos reflejan la necesidad de que los algoritmos actúen con coherencia, que su lógica del funcionamiento sea comprendida, y que se preserve la "agencia", la capacidad de decisión, en el componente humano del sistema. El componente humano en su conjunto, no necesariamente el operador final, ha de ser siempre el responsable del resultado alcanzado[49]. Estos principios no entran en las peculiaridades de sistemas concretos

[49] Ilse Verdiesen, Filippo Santoni de Sio y Virginia Dignum, "Accountability and Control Over Autonomous Weapon Systems: A Framework for Comprehensive Human Oversight", *Minds and Machines* 31, (2021): 137-163.

recogidas en otros códigos más específicos. Permiten, eso sí, establecer un consenso sobre temas urgentes y llevar a cabo una aproximación incremental al problema según se desarrollen nuevos sistemas y aplicaciones.

El ICRC recomienda "descartar el uso de sistemas de armas autónomos para atacar a seres humanos". Ello es así porque "la determinación de si una persona está protegida contra un ataque o si es un objetivo legal es muy contextual y no se presta a ser estandarizada en un perfil de objetivo", "estas caracterizaciones legales pueden cambiar rápidamente"[50].

Principios éticos de la inteligencia artificial en el contexto militar

La dualidad de uso de los sistemas dotados de IA y el carácter definitorio de la autonomía (frente a la letalidad, que aparece como propio de la función guerrera) permiten asumir un alto grado de coincidencia de los principios aplicables a los LAWS, y de otros sistemas de uso militar, con sus equivalentes civiles.

El carácter diferencial, como se ha visto, es el de la especial criticidad de las decisiones que se adoptan en una situación bélica o que afectan a la seguridad nacional. Por la misma razón, tampoco parece aplicable el principio de beneficencia cuando estamos tratando de la guerra. Por lo tanto, dos riesgos aparecen como evidentes, la pérdida de control del armamento por parte del operador humano y el funcionamiento anómalo de aquel. De ello se deriva que los principios éticos básicos de la IA en el contexto militar deben apuntar a retener el control del elemento humano sobre el sistema y a garantizar la robustez de este (ver Figura 2).

[50] ICRC, "ICRC Position on AWS".

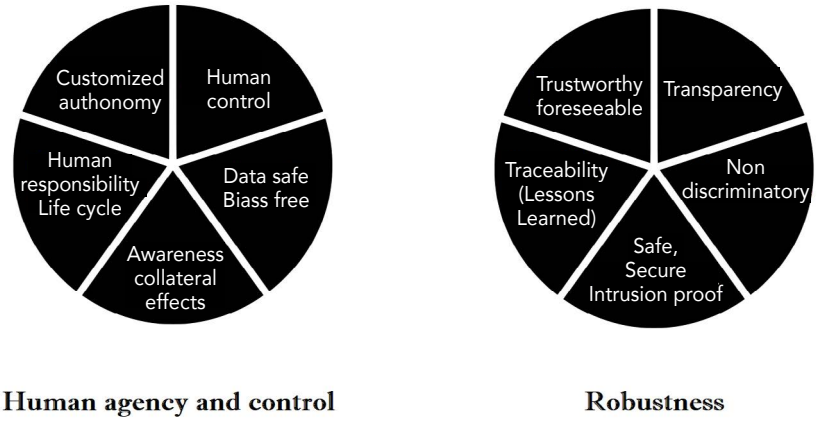

Figura 2. Propuesta de principios éticos para el uso de la IA en el ámbito militar

Fuente: elaboración propia.

a) Control humano

La responsabilidad sobre las decisiones del sistema recae siempre sobre el elemento humano del mismo. Este comprende a los responsables del diseño, desarrollo, adquisición y operación, y no solo al operador. El control implica la capacidad para decidir el resultado final sin sesgos inducidos. Igual que en el combate convencional, el ritmo de batalla marcará la granularidad de la información sobre la que se actúa.

Transferencia limitada de control

Algunos sistemas, especialmente los vinculados a actividades en las que el *tempo* es muy alto (como la defensa aérea), pueden requerir una amplia autonomía de actuación de las máquinas. En este caso, la cesión del control debe efectuarse con razonables garantías de que el resultado seguirá cumpliendo con los principios éticos y legales aplicables. Se deberán establecer límites a la autonomía de las máquinas, que podrán ser:

- Geográficos: delimitando el alcance de los efectos (zonas deshabitadas, fondos marinos, espacio exterior, …)
- Temporales: delimitando la duración de su autonomía a periodos concretos[51]

[51] Este principio se basa en los estudios sobre minas antipersonal y pretende limitar en el tiempo los efectos no controlables de los sistemas de armas.

- Funcionales: permitiendo solo el acometimiento de objetivos incompatibles con errores que puedan dar lugar a una brecha en la legislación aplicable (cuando este objetivo, por ejemplo, reúna características de velocidad, aceleración, temperatura u otros que garanticen que se trata de un vector o un vehículo no tripulado).

Obviamente, el DIH "debe ser cumplido por los usuarios de un AWS, no por el arma en sí"[52]. Existen armas autónomas desde hace mucho tiempo, como las minas, y nunca se ha exigido que dichas armas determinen por sí mismas el cumplimiento de los requisitos legales. Las obligaciones son para las personas y pasan esencialmente por la supervisión[53], un control humano "significativo" o "efectivo", o "niveles apropiados de juicio humano" sobre los sistemas de armas y el uso de la fuerza.

Se requiere la intervención humana en las diferentes fases de desarrollo y ensayo, la decisión del comandante u operador de activar el sistema de armas (fase de activación) y el funcionamiento del sistema de armas autónomo durante el cual selecciona y ataca objetivos de forma independiente (fase de funcionamiento). En muchos casos será más admisible un sistema hombre-máquina que el desarrollo de máquinas completamente autónomas (*brittleness of machine intelligence*)[54].

Control humano significativo

Más allá de definiciones como "human in the loop", "human on the loop", "human out of the loop" o "human beyond the loop", la definición de "control humano significativo" sigue siendo vaga e imprecisa[55]. En todo caso, lo relevante es que el proceso de toma de decisión se realice sin la presencia de sesgos algorítmicos que conviertan al componente humano en mero ejecutor de acciones decididas por la máquina. También debe implicar la capacidad para retener el control sobre el resultado, incluso en caso de injerencia externa en el sistema o fallo de este.

Responsabilidad en todo el ciclo de vida

La responsabilidad del elemento humano por actuaciones no deseadas del sistema puede recaer en el operador o en cualquiera de los que intervengan en las fases

52 ICRC, "ICRC Position on AWS".
53 Estados Unidos, Departamento de Defensa, *AI Principles*, Anexo III.
54 Williams y Scharre, *Autonomous Systems.*
55 M. L. Cummings, "Lethal Autonomous Weapons: Meaningful Human Control or Meaningful Human Certification?" [Opinion], en *IEEE Technology and Society Magazine* 38, núm. 4, (2019): 20-26.

de diseño, desarrollo, adquisición o puesta en operación del sistema[56]. Deberá igualmente tenerse en cuenta el grado de madurez de la tecnología para determinar el grado de autonomía que puede asumir sin desviarse del propósito del operador.

Ausencia de sesgos

Los sistemas algorítmicos "ven" el mundo a través de los datos, las decisiones del elemento humano que se basen en el resultado de sus evaluaciones pueden crear sesgos inducidos adicionales a los propios de la persona. La protección de los datos y su calidad se extiende desde la fase de entrenamiento de los algoritmos hasta la disposición segura de estos una vez ha finalizado la función para la que se recogieron. Durante el entrenamiento hay que evitar la introducción de sesgos indeseados por envenenamiento de las bases de datos[57]. En la fase de operación, la introducción de datos falseados altera el resultado de la toma de decisiones.

b) Robustez y los "unintended engagements"

Es la capacidad del sistema para actuar en consonancia con sus especificaciones. Por lo tanto, de ser coherente y estar dotado de protección frente a manipulaciones de terceros y frente a errores o circunstancias medioambientales cambiantes en la medida en que estas puedan ser previstas. No se puede pedir que un sistema militar sea previsible en la medida en que eso podría ir en detrimento de su utilidad militar, pero sí que mantenga coherencia con las intenciones de su operador. La Directiva de EEUU tiene el objetivo de lograr sistemas "suficientemente robustos" para minimizar los "unintended engagements", esto es, "el uso de la fuerza contra personas u objetos que los comandantes u operadores no tenían la intención de ser objetivos de las operaciones militares estadounidenses"[58].

[56] Lorna McGregor, Daragh Murray y Vivian Ng, "International human rights law as a framework for algorithmic accountability", *International and Comparative Law Quarterly* 68, núm. 2 (2019): 309-343.

[57] Como ya recoge el informe para la Comisión Europea de 2020. Ver Ibán García del Blanco, "REPORT with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies", *Parlamento Europeo*, 8 de octubre de 2020, Report - A9-0186/2020.

[58] Estos fallos pueden darse por "errores humanos, fallas defectuosas en la interacción hombre-máquina, mal funcionamiento, degradación de las comunicaciones, errores de codificación de software, ciberataques enemigos", o "la infiltración en la industria, cadena de suministro, interferencias, suplantación de identidad, señuelos, otras contramedidas o acciones enemigas, o situaciones imprevistas en el campo de batalla", Defense Innovation Board, "AI Principles", 21.

Fiabilidad

El sistema tiene que alcanzar u ofrecer el resultado deseado (no necesariamente el esperado) de forma consistente. La falta de fiabilidad repercute en la confianza en el sistema y, por consiguiente, en su usabilidad.

Transparencia

También vinculada a la confianza, debe permitir la trazabilidad y la explicabilidad del proceso. DARPA XAI[59] es un proyecto del Departamento de Defensa estadounidense en este sentido[60]. Se ha subrayado que si el sistema es opaco y falto de transparencia no será posible garantizar el cumplimiento del DIH, ni por los responsables del sistema autónomo, ni por las personas responsables del cumplimiento del DIH[61].

Trazabilidad

Es la capacidad para realizar auditorías del proceso de decisión o de entrenamiento algorítmico que permite identificar los errores en el mismo y, en su caso, las opciones de mejora. Habilita un proceso de identificación de lecciones aprendidas. Tiene que poder aplicarse tanto a los datos, como a los cómputos, a los procesos ejecutados por la máquina y a las decisiones incorporadas por los diseñadores, desarrolladores y operadores.

Libertad frente a sesgos

La discriminación tiene que producirse en función de los criterios más objetivos posibles. Debe estar libre de sesgos de datos, entrenamiento o inducidos por el programador.

Seguridad

El sistema tiene que ser resiliente frente a fallos en sus componentes y frente a la acción de los elementos o del medio ambiente. En ningún caso debe continuar operando en condiciones de falta de fiabilidad o bajo la sospecha de hallarse en ellas. Finalmente, la permanencia de los datos en el sistema supone una brecha

[59] DARPA es la sigla para *Defense Advanced Research Projects Agency*. XAI es el acrónimo de inteligencia artificial explicable.

[60] David Gunning, "Explainable Artificial Intelligence (XAI). The Need for Explainable AI", *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* 58, núm. 4 (2017).

[61] ICRC, "International Committee of the Red Cross (ICRC) position on autonomous weapon systems: ICRC position and background paper", *IRRC*, núm 915, enero de 2022.

en la privacidad y seguridad de sus propietarios. La confidencialidad y la seguridad tienen que estar garantizadas en todo momento, incluso tras la disposición del sistema, durante el periodo que se fije reglamentariamente.

Las especiales dificultades para el cumplimiento ético y normativo del derecho humanitario por los sistemas IA y especialmente los sistemas autónomos

Responsabilidad individual

Un pilar esencial del DIH es el de la responsabilidad penal individual[62]: se permite enjuiciar, declarar plenamente responsables y sancionar a personas específicas por violaciones graves de las normas y principios del DIH durante situaciones de conflicto armado. Todo ello independientemente de su posición o rango. Este principio promueve la rendición de cuentas, contribuye a prevenir futuras violaciones y, en su caso, ayuda a establecer una base para la justicia y la reconciliación. Este principio ha ido reconociéndose históricamente con expresiones como que "toda persona que cometa un acto que constituya un delito de derecho internacional es responsable del mismo y está sujeta a sanción" (Principio I), incluso si el Derecho interno no considera que dicho acto constituya un delito (Principio II). No se exime de responsabilidad al individuo "si efectivamente ha tenido la posibilidad moral de opción" (Principio IV)[63].

Sin embargo, la proliferación de sistemas autónomos y de IA puede poner en jaque este principio. Pese a que se mantenga en todos sus términos, la responsabilidad puede difuminarse entre diseñadores, supervisores y usuarios del sistema y hacerse poco operativa en la práctica.

62 Sobre el tema: Edoardo Greppi, "Evolución de la responsabilidad penal individual bajo el derecho internacional", *Revista Internacional de la Cruz Roja (RICR)*, núm. 835 (30 de septiembre de 1999): 531-554; Antonio Cruz Mate, Israel, "Responsabilidad internacional penal del individuo por violaciones de normas de derecho internacional humanitario relativas a la protección de las personas civiles y la población civil en los conflictos armados internos" (tesis doctoral, Universidad Complutense, Madrid, 2015), 244 y ss.

63 "Principios de Derecho Internacional reconocidos por el Estatuto y por las sentencias del Tribunal de Nuremberg de 1950", en *The Laws of Armed Conflicts: A Collection of Conventions, Resolutions and other Documents*, editado por Dietrich Schindler y Jiri Toman Martinus Nijhoff (Ginebra: Instituto Henry Dunant, 1988).

La imprevisibilidad

Los sistemas de inteligencia artificial que fijan sus propios objetivos "aprenden" y adaptan su funcionamiento, por ello pueden considerarse intrínsecamente imprevisibles, especialmente cuando se combinan con un entorno a menudo imprevisible y hostil[64]. El DIH prohíbe las armas que son indiscriminadas por naturaleza y un sistema autónomo "implica un riesgo significativo de que los civiles protegidos y los combatientes fuera de combate puedan desencadenar un ataque de AWS". Es por ello por lo que el ICRC recomienda "descartar sistemas de armas autónomos impredecibles"[65]. Cuanto mayor sea la incertidumbre y la imprevisibilidad, mayor será el riesgo de violar el DIH.

Se debe establecer un umbral de riesgo que determine cuál es la probabilidad aceptable de que los sistemas letales autónomos no actúen como se espera[66]. Sin embargo, estos sistemas son también más imprevisibles si realizan múltiples tareas o adaptan su funcionamiento contra distintos tipos de objetivos en un entorno complejo, en una zona amplia o durante mucho tiempo, también debido a que están más o menos supervisados.

Los particulares problemas de la "doctrina doble efecto"

El uso militar de sistemas autónomos y de IA genera auténticos *nudos gordianos,* como sucede con la Doctrina del Doble Efecto[67]. En esencia, es moralmente permisible la muerte no intencional pero previsible de no combatientes, siempre y cuando se cumpla la proporcionalidad. El "cortocircuito" se da porque los sistemas autónomos pueden seleccionar y atacar objetivos de no combatientes, sin intervención humana directa, como objetivo estadísticamente inevitable, lo que constituye una muerte que no es ni colateral ni intencionada. Lorenzo, Floridi y Taddeo concluyen que no es admisible que un sistema autónomo seleccione como objetivo a no combatientes.

64 Neil Davison, "A legal perspective: Autonomous weapon systems under international humanitarian law", *UNODA Occasional Papers*, núm 30 (2017) 5-18; Airbus, "Airbus Defense White paper".

65 ICRC, "ICRC Position on AWS".

66 Alexander Blanchard, Luciano Floridi y Mariarosaria Taddeo, "The Doctrine of Double Effect & Lethal Autonomous Weapon Systems", *SSRN*, 27 de diciembre de 2022.

67 Waltzer resume las condiciones que permiten la posibilidad de muertes de no combatientes: "el acto es bueno en sí mismo o al menos indiferente, lo que significa... que es un acto de guerra legítimo"; además, que el efecto directo sea moralmente aceptable, como la destrucción de suministros militares o la muerte de soldados enemigos; que la intención del actor es buena, solo persigue el efecto aceptable y que el efecto sea suficientemente bueno (proporcionalidad). Michael Walzer, *Just and Unjust Wars: A Moral Argument with Historical Illustrations* (New York: Basic Books, 1977), 153. Blanchard, "The Doctrine of Double Effect".

Conclusiones

La irrupción de los sistemas dotados de IA, tengan o no capacidad letal, es un hecho que con seguridad expandirá sus propios límites actuales. La regulación general de IA, como la de la UE que se va desarrollando, difícilmente es exigible por las habituales exclusiones y excepciones de la defensa. Pese a que no haya normas específicas sobre el uso de IA, las normas internacionales generales de aplicación a la actividad bélica deben interpretarse de forma expansiva a los usos de IA. No en vano, es la actividad y el entorno lo que determina la pertinencia de la norma, nunca el instrumento empleado para la acción.

La proyección del DIH o los principios y regulaciones generales de IA al ámbito de la defensa exige un importante esfuerzo. Ello es así por los elementos diferenciadores del ámbito militar que se han analizado, así como por la criticidad de las actividades que se llevan a cabo, y por los motivos expuestos: los desafíos adicionales que supone el uso de sistemas dotados de distintos grados de autonomía o la integración de estos con los propios humanos, la responsabilidad individual que rige el DIH, la dificultad de proyectar la doctrina del doble efecto a sistemas autónomos o la imprevisibilidad de éstos.

La ética de la IA ha tenido un enorme predicamento en general. Precisamente en el ámbito de defensa la ética adquiere una mayor relevancia por esta ausencia de normativa específica y por las dificultades de proyección de las obligaciones normativas generales existentes. Es preciso garantizar el cumplimiento de las normas internacionales como un compromiso ético de mínimos que no comprometa los avances de estas tecnologías en otros campos y que no menoscabe la confianza del público en sus productos. Asimismo, en la medida de lo posible, cabe considerar como un compromiso ético la asunción voluntaria de los Estados de las declaraciones y recomendaciones, así como del derecho de la IA que se vaya generando, como el futuro Reglamento de la UE.

En cualquier caso, la aplicación de los principios éticos y las normas jurídicas sigue recayendo siempre sobre el componente humano del sistema inteligente. Para que la responsabilidad de su cumplimiento pueda ejercerse con suficiente criterio es preciso que el componente algorítmico sea robusto y consistente, y que el control por parte del humano pueda ejercerse sin sesgos inducidos hasta el momento en el que el resultado de la acción sea inequívoco. Esto último es particularmente difícil en la complejidad del entorno bélico que, por lo tanto, supone un escenario adecuado para observar los límites de la autonomía de los sistemas dotados de IA.

Referencias

"Airbus Defense White paper: The Responsible Use of Artificial Intelligence in FCAS – An Initial Assessment", *Airbus, Fraunhofer FKIE*, 2020. https://www.fcas-forum.eu/articles/responsible-use-of-artificial-intelligence-in-fcas

"An Agenda for Action Alternative Processes for Negotiating a Killer Robots Treaty", *Human Rights Watch*, 19 de noviembre de 2022. https://www.hrw.org/report/2022/11/10/agenda-action/alternative-processes-negotiating-killer-robots-treaty

Argentina, Costa Rica, Guatemala, Kazakhstan, Nigeria, Panamá, Filipinas, Sierra Leona, Palestina y Uruguay. "Proposal: Roadmap Towards New Protocol on Autonomous Weapons Systems", *reachingcriticalwill.org*, 13 de marzo de 2022. https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2022/gge/documents/G13_March2022.pdf

Australia, Canadá, Japón, República de Corea, Reino Unido y Estados Unidos. "Principles and Good Practices on Emerging Technologies in the Area of LAWS", *reachingcriticalwill.org*, 7 de marzo de 2022. https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2022/gge/documents/USgroup_March2022.pdf

Automated Decision Research, *State positions*. https://automatedresearch.org/state-positions/

Blanchard Alexander, Luciano Floridi y Mariarosaria Taddeo. "The Doctrine of Double Effect & Lethal Autonomous Weapon Systems", *SSRN*, 27 de diciembre de 2022 http://dx.doi.org/10.2139/ssrn.4308862

Boulanin, Vicent, Neil Davison, Netta Goussac y Moa Peldán Carlsson, *Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control* (Ginebra: ICRC y Stockholm International Peace Research Institute, 2020), 21-25 https://www.icrc.org/en/download/file/121024/icrc_sipri_limits_on_autonomy_june_2020.pdf

"Campaign to Stop Killer Robots", *Stop Killer Robots*, 2018. https://www.stopkillerrobots.org/

Chakraborty, Swaroop. "Inteligencia artificial y derechos humanos: ¿son convergentes o paralelos entre sí?". *Novum Jus* 12, núm. 2 (2018): 13-38. https://doi.org/10.14718/NovumJus.2018.12.2.2

Comisión Europea, Dirección General de Redes de Comunicación, Contenido y Tecnologías. "Directrices éticas para una IA fiable". Oficina de Publicaciones, 8 de abril de 2019. https://data.europa.eu/doi/10.2759/14078

Comisión Europea, "Propuesta de Reglamento por el que se establecen normas armonizadas en materia de inteligencia artificial", *EUR-Lex Access to European Union Law*, 21 de abril de 2021. https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex:52021PC0206

Consejo de Europa, Comité sobre inteligencia artificial (CAI). "Borrador de trabajo consolidado del convenio marco sobre inteligencia artificial, derechos humanos, democracia y Estado

de derecho". *Council of Europe (COE)*, Estrasburgo, 7 de julio de 2023. https://rm.coe.int/cai-2023-18-consolidated-working-draft-framework-convention/1680abde66

Cotino Hueso, Lorenzo. "Ética en el diseño para el desarrollo de una inteligencia artificial, robótica y *big data* confiables y su utilidad desde el derecho". *Revista Catalana de Derecho Público*, núm. 58 (2019): 29-48. http://dx.doi.org/10.2436/rcdp.i58.2019.3303

Cruz Mate, Antonio, "Responsabilidad internacional penal del individuo por violaciones de normas de derecho internacional humanitario relativas a la protección de las personas civiles y la población civil en los conflictos armados internos". Tesis doctoral, Universidad Complutense, Madrid, 2015.

Cummings, M. L. "Lethal Autonomous Weapons: Meaningful Human Control or Meaningful Human Certification?" [Opinion], en *IEEE Technology and Society Magazine* 38, núm. 4, (2019): 20-26. https://doi.org/10.1109/MTS.2019.2948438

Davison, Neil. "A legal perspective: Autonomous weapon systems under international humanitarian law". *UNODA Occasional Papers*, núm 30 (2017) 5-18. https://doi.org/10.18356/29a571ba-en

Devitt, Kate, Michael Gan, Jason Scholz y Robert Bolia. *A Method for Ethical AI in Defence*. Australia: Departamento de defensa, 2020. Publicación número DSTG-TR-3786. https://www.dst.defence.gov.au/publication/ethical-ai

Dietrich Schindler y Jiri Toman. *The Laws of Armed Conflicts: A Collection of Conventions, Resolutions and other Documents*. Ginebra: Martinus Nijhoff/Instituto Henry Dunant, 1988. https://doi.org/10.1163/9789047405238

Estados Unidos, Departamento de Defensa, Defense Innovation Board. "AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense", *Department of Defense USA (DoD), Supporting Document*, octubre 31 de 2019. https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF

Estados Unidos, Departamento de Defensa. "The Department of Defense AI Ethical Principles", The Joint Artificial Intelligence Center, 24 de febrero de 2020. https://www.ai.mil/blog_02_24_20-dod-ai_principles.html

Estados Unidos, Departamento de Defensa, *DoD Law of War Manual*. Washington: Consejo General del Departamento de Defensa, 2023. https://media.defense.gov/2023/Jul/31/2003271432/-1/-1/0/DOD-LAW-OF-WAR-MANUAL-JUNE-2015-UPDATED-JULY%202023.PDF

Finlandia, Francia, Alemania, Países Bajos, Noruega, España y Suecia, "Working paper submitted to the 2022 Chair of GGE on LAWS", 13 de julio de 2022, https://documents.unoda.org/wp-content/uploads/2022/07/WP-LAWS_DE-ES-FI-FR-NL-NO-SE.pdf

Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar. "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based

Approaches to Principles for AI", Berkman Klein Center for Internet & Society Research at Harvard University, 2020. https://doi.org/10.2139/ssrn.3518482 http://nrs.harvard.edu/urn-3:HUL.InstRepos:42160420

Floridi, Luciano, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, Peggy Valcke y Effy Vayena. , "AI4People —An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations", *Minds and Machines* 28, núm 4 (2018): 689-707 https://doi.org/10.1007/s11023-018-9482-5

Francia, Defence Ethics Committee. "Opinion on the Augmented soldier", *Ministère des Armées*, 18 de septiembre de 2020. https://n9.cl/6qw9u

García del Blanco, Ibán. "REPORT with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies", *Parlamento Europeo*, 8 de octubre de 2020, Report - A9-0186/2020. https://www.europarl.europa.eu/doceo/document/A-9-2020-0186_EN.html

Gómez de Ágreda, Ángel. "Ethics of autonomous weapons systems and its applicability to any AI systems", *Telecommunications Policy* 44, núm. 6 (2020): 1-15. https://doi.org/10.1016/j.telpol.2020.101953

Greppi, Edoardo. "Evolución de la responsabilidad penal individual bajo el derecho internacional". *Revista Internacional de la Cruz Roja (RICR)*, núm. 835 (30 de septiembre de 1999): 531-554. https://www.icrc.org/es/doc/resources/documents/misc/5tdnnf.htm

Gunning, David. "Explainable Artificial Intelligence (XAI). The Need for Explainable AI". *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* 58, núm. 4 (2017). https://doi.org/10.1111/fct.12208

International Business Machines (IBM). "Foundation models: Opportunities, risks and mitigations", julio de 2023. https://www.ibm.com/downloads/cas/E5KE5KRZ

International Committee of the Red Cross (ICRC). "Base de datos sobre DIH consuetudinario", Normas. https://ihl-databases.icrc.org/es/customary-ihl/v1

International Committee of the Red Cross (ICRC). "ICRC Position on Autonomous Weapon Systems & Background Paper". *ICRC*, 12 de mayo de 2021. https://www.icrc.org/en/download/file/166330/icrc_position_on_aws_and_background_paper.pdf

International Committee of the Red Cross (ICRC), "International Committee of the Red Cross (ICRC) position on autonomous weapon systems: ICRC position and background paper", *IRRC*, núm 915, enero de 2022. https://international-review.icrc.org/articles/icrc-position-on-autonomous-weapon-systems-icrc-position-and-background-paper-915

International Committee of the Red Cross (ICRC). "Protocolo I adicional a los Convenios de Ginebra de 1949 relativo a la protección de las víctimas de los conflictos armados internacionales". 18 de junio de 1977. https://www.icrc.org/es/document/

protocolo-i-adicional-convenios-ginebra-1949-proteccion-victimas-conflictos-armados-internacionales-1977

International Committee of the Red Cross (ICRC). "Statement of the ICRC to the UN CCW GGE on Lethal Autonomous Weapons Systems". 21-25 de septiembre de 2020, Ginebra.

International Committee of the Red Cross (ICRC). "Views of the ICRC on autonomous weapon systems". *ICRC*, 11 abril 2016. https://www.icrc.org/en/document/views-icrc-autonomous-weapon-system

"Letter dated 8 March 2021 from the Panel of Experts on Libya Established pursuant to Resolution 1973 (2011) addressed to the President of the Security Council", *Naciones Unidas, Biblioteca digital*. https://digitallibrary.un.org/record/3905159?ln=es

McGregor, Lorna, Daragh Murray y Vivian Ng, "International human rights law as a framework for algorithmic accountability". *International and Comparative Law Quarterly* 68, núm. 2 (2019): 309-343. https://doi.org/10.1017/S0020589319000046

Michael C. Horowitz, "Public opinion and the politics of the killer robots debate", *Research and Politics* 3, núm. 1 (2016). https://doi.org/10.1177/2053168015627183

Microsoft. "AI Principles", 2019. https://www.microsoft.com/en-us/ai/principles-and-approach

Minor, E. "Laws for LAWS. Towards a treaty to regulate lethal, autonomous weapons". *Friedrich Ebert Stiftung New York Analysis*, Febrero de 2023. https://library.fes.de/pdf-files/international/20013.pdf

Nadella, Satya. "The Partnership of the Future. Microsoft's CEO explores how humans and A.I. can work together to solve society's greatest challenges", *Slate*, 28 de junio de 2016. https://slate.com/technology/2016/06/microsoft-ceo-satya-nadella-humans-and-a-i-can-work-together-to-solve-societys-challenges.html

Naciones Unidas, Convention on Conventional Weapons (CCW), "Report of the 2018 Group of Governmental Experts on Lethal Autonomous Weapons Systems", 31 de agosto de 2018.

Naciones Unidas. "Reunión de las Altas Partes Contratantes de la Convención sobre prohibiciones o restricciones del uso de ciertas armas convencionales que puede considerarse excesivamente perjudicial o tener efectos indiscriminados". Reporte Final, Ginebra, 16-18 de noviembre de 2022. UN Document CCW/MSP/2022/7. https://unoda-documents-library.s3.amazonaws.com/Convention_on_Certain_Conventional_Weapons_-_Meeting_of_High_Contracting_Parties_(2022)/CCW-MSP-2022-7-Advance_version.pdf

Naciones Unidas, Institute for Disarmament Research. "The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches, a primer". UNIDIR Resources, núm. 6 (2017). https://unidir.org/publication/

the-weaponization-of-increasingly-autonomous-technologies-concerns-characteristics-and-definitional-approaches/

Organización para la Cooperación y el Desarrollo Económicos (OCDE). "Recommendation of the Council on Artificial Intelligence", OECD/LEGAL/0449, OECD Legal Instruments, 2023. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Pichai, Sundar. "AI, Google: our principles" (Blog). 7 de junio de 2018. https://blog.google/technology/ai/ai-principles/

Potoglou, Dimitris, Sunil Patil, Covadonga Gijon, Juan Palacios y Claudio Feijoo. "The value of personal information online: Results from three stated preference discrete choice experiments in the UK". Conferencia presentada en el 21st European Conference for Information Systems ORCA, Utrech, 5-8 de junio de 2013. http://orca.cf.ac.uk/51292/

Ramos, Gabriela. "Ética de la inteligencia artificial". *Unesco. Inteligencia artificial*. https://www.unesco.org/es/artificial-intelligence/recommendation-ethics

Responsible AI in the Military domain Summit (REAIM). "REAIM Call to Action", *Gobierno de Países Bajos*, 16 de febrero de 2023. https://www.government.nl/documents/publications/2023/02/16/reaim-2023-call-to-action

Scalia, Tania, Alessandro Di Mezza, Alessandra Masini, Sebastien Sylvestre, Robert Thomas, Jean-Louis Szabo, Marcel De Heide, Maurits Butter y David Parker. "Final technical report. Study on the dual-use potential of Key Enabling Technologies (KETs)". *Comisión Europea, Agencia Ejecutiva para las Pequeñas y Medianas Empresas*, 13 de enero de 2017. https://doi.org/10.2826/12343

Schmitt, Michael. "Grey Zones in the International Law of Cyberspace". *The Yale Journal of International Law Online* 42, núm. 2 (2016): 1-21. https://bpb-us-w2.wpmucdn.com/campuspress.yale.edu/dist/8/1581/files/2017/08/Schmitt_Grey-Areas-in-the-International-Law-of-Cyberspace-1cab8kj.pdf

Unión Europea, Parlamento Europeo. "Normas de Derecho civil sobre robótica". Resolución del 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL)). http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2017-0051+0+DOC+XML+V0//ES

Ulloa Plaza, Jorge y Maria A. Benavides Casals. "Moralidad, guerra y derecho internacional. Tres cuerdas para un mismo trompo: la humanidad". *Novum Jus* 17, núm. 1 (2023): 259-282 https://doi.org/10.14718/NovumJus.2023.17.1.11

Verdiesen, Ilse, Filippo Santoni de Sio y Virginia Dignum. "Accountability and Control Over Autonomous Weapon Systems: A Framework for Comprehensive Human Oversight". *Minds and Machines* 31, (2021): 137-163. https://doi.org/10.1007/s11023-020-09532-9

Walzer, Michael. *Just and Unjust Wars: A Moral Argument with Historical Illustrations*. New York: Basic Books, 1977.

Williams, Andrew y Paul D. Scharre, *Autonomous Systems: Issues for Defence Policymakers*. Norfolk: Nato Communications and Information Agency, 2015. https://apps.dtic.mil/sti/pdfs/AD1010077.pdf